Department of Social Systems and Management

Discussion Paper Series

No.1287

Critical Comparisons between the Nash Noncooperative Theory and Rationalizability

by

Tai-wei HU and Mamoru KANEKO

January 2012

UNIVERSITY OF TSUKUBA
Tsukuba, Ibaraki 305-8573
JAPAN

Critical Comparisons between the Nash Noncooperative Theory and Rationalizability*

Tai-Wei Hu† and Mamoru Kaneko‡

08 January 2012

Abstract

The theories of Nash noncooperative solutions and of rationalizability intend to describe the same target problem of *ex ante* individual decision making, but they are distinctively different. We consider what their essential difference is by giving parallel derivations of their resulting outcomes. The derivations pinpoint that the difference is only in the use of quantifiers for each player's prediction about the other's possible decisions; the universal quantifier for the former and the existential quantifier for the latter. Using this difference, we argue that the former is compatible with the free-will postulate for game theory that each player has free will for his decision making, and that for the latter, the interpretation in terms of determinism would be more natural. In the present approach, however, the distinction between decisions and predictions still remains interpretational. For an explicit distinction, we undertake, in the companion paper, a study of those theories in a framework of common knowledge logic.
JEL Classification Numbers: B40, C70, C72
Key words: Nash equilibrium, Solvability, Rationalizability, Prediction/Decision Criterion, Infinite Regress, Simultaneous Equations

1. Introduction

We make critical comparisons between the theory of *Nash noncooperative solutions* due to Nash [17] and the theory of *rationalizable strategies* due to Bernheim [3] and Pearce [18]. Either is intended to be a theory about *ex ante* individual decision making in a game, i.e., decision making before the actual play of the game. The difference in their resulting outcomes has been well analyzed and known. However, their conceptual

*The authors are partially supported by Grant-in-Aids for Scientific Research No.21243016, Ministry of Education, Science and Culture.
†MEDS Department, Northwestern University, Evanston, IL, (t-hu@kellogg.northwestern.edu)
‡IPPS, University of Tsukuba, Ibaraki 305-8573, Japan (kaneko@sk.tsukuba.ac.jp)

difference has not been much discussed. In this paper, we evaluate these theories from the perspective of *ex ante* decision making and connect them to the basic postulates of game theory. We confine our scope of analysis to idealized decision making, partly because it is the focus of our targeted theories. We address the question of logical coherence, for the two theories, with conceptual bases of game theory.

First, we review the literature of these theories. It is well known that Nash [17] provides the concept of Nash equilibrium and proves its existence in mixed strategies. However, it is less known that the main focus of [17] is on *ex ante* individual decision making. He develops various other concepts such as interchangeability, solvability, subsolutions, symmetry, and values, which are ingredients of a theory of *ex ante* individual decision making, though the aim is not explicitly stated in [17]. This view is discussed only in a few papers such as Johansen [10] and Kaneko [11][1]. We call the entire argumentation the *Nash noncooperative theory*[2].

On the other hand, in the literature, the theory of rationalizability is typically regarded as a faithful description of *ex ante* individual decision making in games, and is interpreted as expressing the idea of the common knowledge of "rationality". According to Mas-Colell *et al.* [13], p.243, "*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 view is common in many standard game theory/micro-economics textbooks.

We find a puzzling feature of these two theories: Both theories target *ex ante* individual decision making, and are regarded as successful by some or many researchers. However, their formal definitions, predicted outcomes, and explanations differ considerably. This puzzling feature raises the following question: Are any components or basic postulates conceptually wrong in either (or both) of them? This paper attempts to answer this question.

We pinpoint the difference between the two theories; it emerges through formulating a new prediction/decision criterion for each theory. For the Nash theory, it is given as the following circular requirements:

N1^o: player 1 chooses his best strategy against <u>all</u> of his predictions
about player 2's choice based on N2^o;

N2^o: player 2 chooses his best strategy against <u>all</u> of his predictions
about player 1's choice based on N1^o.

---

[1] Millham [15] and Jansen [9] study the mathematical structure of the solution and subsolutions, but do not touch the view.

[2] The mathematical definition of Nash equilibrium allows different interpretations such as a steady state in a repeated situation. Some variant interpretations may sneak into our consideration of the Nash noncooperative theory, which prevents us from crystallizing the theory. See Johansen [10] and Kaneko [12] for those interpretations.

A possible decision for 1 is determined by $N1^o$ but requires a prediction about 2's possible decision which is determined by $N2^o$. The symmetric form $N2^o$ determines a possible decision for 2 with a prediction about 1's possible decision. These are regarded as a system of simultaneous equations with players' decisions/predictions as unknown. In Section 3, we show the theorem that $N1^o$ and $N2^o$ characterize the Nash noncooperative solution as the greatest set satisfying them if the game is solvable (the set of Nash equilibria is interchangeable); and if not, a maximal set satisfying them is a subsolution.

The rationalizable strategies are characterized by $R1^o$ and $R2^o$, which are obtained from $N1^o$-$N2^o$ simply by replacing the quantifier "for all" by "for some":

$R1^o$: player 1 chooses his best strategy against <u>some</u> of his predictions
about player 2's choice based on $R2^o$;

$R2^o$: player 2 chooses his best strategy against <u>some</u> of his predictions
about player 1's choice based on $R1^o$.

These requirements are closely related to the BP-property ("best-reponse property" in Bernheim [3] and Pearce [18]), and the characterization result is given in Section 3.

The characterization results unify the Nash noncooperative theory and rationalizability theory, and pinpoint their difference: It is the choice of the universal or existential quantifiers for predictions about the other player's possible decisions. A basic methodological postulate of game theory is that each player has free will, which is associated with decision making. The quantifier "for all" in $N1^o$-$N2^o$ can be understood as coherent in the application of this postulate between the players, but "for some" in $R1^o$-$R2^o$ is difficult to be reconciled with it.

In Section 4, we argue that the theory of rationalizability is better understood from the perspective of complete determinism. Indeed, the epistemic justification for rationalizability begins with a complete description of players' actions as well as mental states, and characterizes classes of those states by certain assumptions. On the other hand, the Nash noncooperative solutions correspond to predictions that result from players' active inferences based on certain axioms about their own and other players' decision-making. This insight has been emphasized by Johansen [10], and will be further discussed in the companion paper [8] of the present paper.

As a result, our problem is a choice between two methodological assumptions, the free-will postulate and complete determinism. This choice are discussed in Morgenstern [16] and Heyek [7] in the context of economics and/or social science in general. Based upon their arguments, we will conclude that the large part of social science is incompatible with complete determinism. From this perspective, the Nash noncooperative theory is preferable to rationalizability.

The Nash theory might be regarded as having a defect in that it does not generate definite predictions for unsolvable games. However, we argue that this is not a defect; rather, it points out that additional principles, other than the decision criteria given above, are needed for unsolvable games. The study of those additional principles is beyond this research project, but we remark that many applied works that use game theory appeal to principles such as symmetry (which is already discussed in Nash [17]) and Pareto optimality.

Related to this issue is the notion of *rationality* in game theory. In the theory of rationalizability, rationality is more or less equivalent to payoff maximization; here, we take a broader view of rationality, which includes, but not limited to, the decision/prediction criterion and logical abilities to understand their implications, while payoff maximization is only a component of rationality. With this broader view, one can incorporate additional principles or criteria such as symmetry or Pareto optimality and investigate whether those principles are consistent with more basic ones.

The paper is written as follows: Section 2 introduces the theories of Nash noncooperative solutions and rationalizable strategies; we restrict ourselves to finite 2-person games for simplicity. Section 3 formulates N1^o-N2^o and R1^o-R2^o, and gives two theorems characterizing the Nash noncooperative theory and rationalizability. In Section 4, we discuss implications from them considering foundational issues. Section 5 gives a summary and states continuation to the companion paper.

2. Preliminary Definitions

Here, we define basic concepts in a finite 2-person game. We do not allow mixed strategies in this paper, but the main results hold with mixed strategies. In Section 3.3, we discuss necessary changes in our formulation to accommodate mixed strategies.

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* against s_j, i.e.,

$$h_i(s_i; s_j) \geq h_i(s_i'; s_j) \text{ for all } s_i' \in S_i, \tag{2.1}$$

is denoted by $\mathrm{Best}(s_i; s_j)$. Since $S_1 \cap S_2 = \emptyset$, the expression $\mathrm{Best}(s_i; s_j)$ has no ambiguity. We say that (s_1, s_2) is a *Nash equilibrium* in G iff $\mathrm{Best}(s_i; s_j)$ holds for $i \in N$. We define $E(G)$ to be the set of all Nash equilibria in G. The set $E(G)$ may be empty.

Nash Noncooperative Solutions: Let E be a subset of $S_1 \times S_2$. We say that E is *interchangeable* iff

$$(s_1, s_2), (s_1', s_2') \in E \text{ imply } (s_1, s_2') \in E. \tag{2.2}$$

It is known that this is equivalent for E to have the product form. For completeness, we give it as a lemma. The only-if part is essential: If $(s_1, s_2) \in E_1 \times E_2$, then for some $s_1' \in S_1$ and $s_2' \in S_2$ we have $(s_1, s_2') \in E$ and $(s_1', s_2) \in E$, which, together with (2.2), implies $(s_1, s_2) \in E$.

Lemma 2.1. Let $E \subseteq S_1 \times S_2$ and let $E_i = \{s_i : (s_i; s_j) \in E$ for some $s_j \in S_j\}$ for $i = 1, 2$. Then, E satisfies (2.2) if and only if $E = E_1 \times E_2$.

Now, let $\mathbf{E} = \{E : E \subseteq E(G)$ and E satisfies (2.2)$\}$. We say that E is the *Nash solution* iff E is nonempty and is the greatest set in $\mathbf{E}$, i.e., $E' \subseteq E$ for any $E' \in \mathbf{E}$ and $E \neq \emptyset$. We say that E is a *Nash subsolution* iff E is a nonempty maximal set in $\mathbf{E}$, i.e., there is no $E' \in \mathbf{E}$ such that $E \subsetneq E'$. We call these the *Nash noncooperative solutions.*

Table 2.1

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$
$\mathbf{s}_{11}$	$(2,2)$	$(1,1)$
$\mathbf{s}_{12}$	$(1,1)$	$(0,0)$

Table 2.2

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$
$\mathbf{s}_{11}$	$(1,1)$	$(1,1)$
$\mathbf{s}_{12}$	$(1,1)$	$(0,0)$

When $E(G) \neq \emptyset$, $E(G)$ is the Nash solution if and only if $E(G)$ satisfies (2.2). When the Nash solution exists for game G, G is called *solvable.* The game of Table 2.1 is solvable. Thus, a game G is not solvable if and only if $E(G) = \emptyset$ or the nonempty greatest set does not exist. On the other hand, for a game G with $E(G) \neq \emptyset$, a subsolution exists always; specifically, for any $(s_1, s_2) \in E(G)$, there is a subsolution E^o with $(s_1, s_2) \in E^0$. This E^o may not be unique: The game of Table 2.2 is not solvable and has two subsolutions: $\{(\mathbf{s}_{11}, \mathbf{s}_{21}), (\mathbf{s}_{11}, \mathbf{s}_{22})\}$ and $\{(\mathbf{s}_{11}, \mathbf{s}_{21}), (\mathbf{s}_{12}, \mathbf{s}_{21})\}$, and both include $(\mathbf{s}_{11}, \mathbf{s}_{21})$.

In Section 3, we argue that the Nash solution can be regarded as describing *ex ante* individual decision making; here we give two comments about its interpretation. First, for a solvable game, each component of the solution consists of a pair of strategies, (s_1, s_2), rather than a single strategy. This means that from player 1's perspective, s_1 describes player 1's possible decision while s_2 is player 1's prediction of player 2's possible decisions. As shown later, a distinction between a decision and a prediction is crucial from the perspective of *ex ante* decision making in a game.

Second, the Nash theory does not provide a definite recommendation for possible decisions if the game is unsolvable and if a subsolution exists. Suppose that G has exactly two subsolutions, say, $F^1 = F_1^1 \times F_2^1$ and $F^2 = F_1^2 \times F_2^2$ with $F_i^1 \neq F_i^2$ for $i = 1, 2$. One may think that the Nash theory would recommend the set $E_i = F_i^1 \cup F_i^2$ for player i as the set of possible decisions to play G. However, we find neither E_1' or E_2' so that $E_1' \times (F_2^1 \cup F_2^2)$ or $(F_1^1 \cup F_1^2) \times E_2'$ satisfies interchangeability.

Rationalizable Strategies: Now, we turn to rationalizability. Although there are various definitions of rationalizability, we take the iterative one: A sequence of sets of strategies, $\{(R_1^\nu(G), R_2^\nu(G))\}_{\nu=0}^\infty$, is inductively defined as follows: for $i = 1, 2$, $R_i^0(G) = S_i$, and

$$R_i^\nu(G) = \{s_i : \text{Best}(s_i; s_j) \text{ holds for some } s_j \in R_j^{\nu-1}(G)\} \text{ for any } \nu \geq 1. \tag{2.3}$$

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 $i = 1, 2$; that is, we say a pure strategy $s_i \in S_i$ is *rationalizable* iff $s_i \in R_i(G)$. Note that $R_i^{\nu}(G)$ is nonempty for all ν and $i = 1, 2$, which is shown by induction over ν.

It is known that each $\{R_i^{\nu}(G)\}_{\nu}$ is monotonically decreasing. Because each $R_i^{\nu}(G)$ is finite and nonempty, $R_i^{\nu}(G)$ becomes constant after some $\overline{\nu}$; as a result, $R_i(G)$ is nonempty. In fact, this $\overline{\nu}$ can be taken as $2\min(|S_1|, |S_2|) + 1$. These facts are more or less known. However, they are significant for our purposes (in particular for the companion paper), and we give a proof for completeness.

Lemma 2.2.(1): $\{R_i^{\nu}(G)\}_{\nu}$ is a decreasing sequence, i.e., $R_i^{\nu}(G) \supseteq R_i^{\nu+1}(G)$ for all ν;

(2): $R_i^{\nu}(G) = R_i^{\nu+1}(G) \neq \emptyset$ for all $\nu \geq \overline{\nu} = 2\min(|S_1|, |S_2|) + 1$.

Proof (1): 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. Once this is shown, since S_i is finite, we have $R_i(G) = \bigcap_{\nu=0}^{\infty} R_i^{\nu}(G) \neq \emptyset$. For the base case of $\nu = 0$, we have $R_i^0(G) = S_i \supseteq R_i^1(G)$ for $i = 1, 2$. Now, suppose the hypothesis that this inclusion holds up to ν and $i = 1, 2$. Let $s_i \in R_i^{\nu+1}(G)$. By (2.3), $\text{Best}_i(s_i; s_j)$ holds for some $s_j \in R_j^{\nu}(G)$. Since $R_j^{\nu-1}(G) \supseteq R_j^{\nu}(G)$ by the supposition, $\text{Best}_i(s_i; s_j)$ holds for some $s_j \in R_j^{\nu-1}(G)$. This means $s_i \in R_i^{\nu}(G)$.

(2): Since each $R_i^{\nu}(G)$ is a nonempty finite set, we have, by (1), there is some $\overline{\nu}$ such that $R_i^{\nu}(G) = R_i^{\nu+1}(G) \neq \emptyset$ for all $\nu \geq \overline{\nu}$. Now we show that this $\overline{\nu}$ can be $2\min(|S_1|, |S_2|)+1$. Let $|S_1| \leq |S_2|$. The other case is symmetric. It suffices to show that $R_1^{\overline{\nu}}(G) = R_1^{\overline{\nu}+1}(G)$ and $R_2^{\overline{\nu}}(G) = R_2^{\overline{\nu}+1}(G)$, since these imply $R_1^{\overline{\nu}+k}(G) = R_1^{\overline{\nu}+k+1}(G)$ and $R_2^{\overline{\nu}+k}(G) = R_2^{\overline{\nu}+k}(G)$ for all $k \geq 0$.

Suppose $R_1^{\overline{\nu}}(G) \supsetneqq R_1^{\overline{\nu}+1}(G)$. This implies $R_2^{\overline{\nu}-1}(G) \supsetneqq R_2^{\overline{\nu}}(G)$, which further implies $R_1^{\overline{\nu}-2}(G) \supsetneqq R_1^{\overline{\nu}-1}(G)$. By induction, we can prove: for $k = 0, ..., |S_1|$,

$$R_1^{\overline{\nu}-2k}(G) \supsetneqq R_1^{\overline{\nu}-2k+1}(G) \text{ and } R_2^{\overline{\nu}-2k-1}(G) \supsetneqq R_2^{\overline{\nu}-2k}(G). \tag{2.4}$$

The cardinality of $R_1^{\overline{\nu}-2|S_1|}(G) = R_1^1(G)$ is $\left|R_1^1(G)\right| = |S_1| + \left|R_1^{\overline{\nu}+1}(G)\right| \geq |S_1| + 1$, which is impossible since $R_1^1(G) \subseteq S_1$. Thus, $R_1^{\overline{\nu}}(G) = R_1^{\overline{\nu}+1}(G)$. By the parallel argument, we have $R_2^{\overline{\nu}}(G) = R_2^{\overline{\nu}+1}(G)$. ■

Criterion for Decision/Prediction Making: Our discussion of *ex ante* decision making in games begin with a decision/prediction criterion. While our concern is about comparisons between the Nash theory and rationalizability, some simpler example of decision criteria may be helpful. A classical example of a decision criterion is the *maximin* criterion due to von Neumann-Morgenstern [20]: It recommends a player to choose a strategy maximizing the guarantee level (that is, the minimum payoff for a strategy). In $G = (N, \{S_i\}_{i \in N}, \{h_i\}_{i \in N})$, let E_i be a nonempty subset of $S_i, i = 1, 2$.

The set E_i is interpreted as the set of possible decisions for player i. The criterion is formulated as follows:

NM1: for each $s_1 \in E_1$, s_1 maximizes $\min_{s_2 \in S_2} h_1(s_1; s_2)$;

NM2: for each $s_2 \in E_2$, s_2 maximizes $\min_{s_1 \in S_1} h_2(s_2; s_1)$.

These are not interactive at all, since NMi, $i = 1, 2$, can recommend a decision without depending upon NMj and also player i needs to know only his's own payoff function. Thus, no prediction is involved for decision making with these criteria.

A more sophisticated criterion may allow a player to consider the other's criterion. One possibility is the following:

NM1: for each $s_1 \in E_1$, s_1 maximizes $\min_{s_2 \in S_2} h_1(s_1; s_2)$;

N2: for each $s_2 \in E_2$, Best(s_2, s_1) holds for all $s_1 \in E_1$.

The second requires player 2 to predict player 1's possible decisions and to choose his decision against that prediction, while player 1 still adopts the maximin criterion. In this sense, their interpersonal thought stops at depth 2. In the Nash theory and rationalizability theory, we would meet some circularity and their interpersonal thought goes beyond depth 2. Note that N2 is a mathematical formulation of N2^o and will be used in the characterization of the Nash theory.

We should comment on the choice of E_1 or E_2 when there are multiple candidates for them. Without other information than the criterion and components of the game, the outside observer cannot make a further choice of particular strategies. In the case of NM1-NM2, E_i should consist of all strategies maximizing $\min_{s_2 \in S_2} h_1(s_1; s_2)$; E_i is the greatest set satisfying NMi. In the case of NM1-N2, this should also be applied to player 2's predictions about 1's choice: E_1 in N2 should be the greatest set satisfying NM1. We will adopt this practice of taking the greatest set for E_i in Section 3. This is not a mere mathematical practice, but is very basic for the consideration of *ex ante* decision making: It is stated as Johansen's [10] postulate in Section 4.1.

3. Parallel Derivations of the Nash Noncooperative Solutions and Rationalizable Strategies

Our discussion of *ex ante* decision making in games begin with decision criteria. We give two parallel decision criteria, and derive the Nash noncooperative solutions and the rationalizable strategies from those criteria. Our characterization results pinpoint the difference between the two theories. This difference is used as the basis of our evaluation of these two theories of *ex ante* individual decision making in Section 4. We give remarks on the mixed strategy versions of those derivations in Section 3.3.

3.1. The Nash Noncooperative Solutions

The decision criterion for the Nash solution formalizes the statements $\mathrm{N1}^o$ and $\mathrm{N2}^o$ in Section 1. Let E_i be a subset of $S_i, i = 1, 2$, interpreted as the set of possible decisions: $\mathrm{N1}^o$ and $\mathrm{N2}^o$ are now formalized as:

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

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

These describe how each player chooses possible decisions; when one player's viewpoint is fixed, one of N1-N2 is interpreted as decision making, and the other is interpreted as prediction making. For example, from player 1's perspective, N1 describes his decision making, and N2 describes his prediction making.

Mathematically, N1 and N2 can be regarded as a system of simultaneous equations with unknown E_1 and E_2. First we give a lemma showing that (E_1, E_2) satisfies N1-N2 if and only if it consists only of Nash equilibria.

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 Nash equilibrium in G.

Proof. (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 a Nash equilibrium in G.
(If): Let $(s_1, s_2) \in E_1 \times E_2$ be a Nash equilibrium. Since $h_1(s_1, s_2) \geq h_1(s_1', s_2)$ for all $s_1' \in S_1$, we have N1. We have N2 similarly. ■

Regarding N1-N2 as a system of simultaneous equations with unknown E_1 and E_2, there may be multiple solutions; indeed, any pair of Nash equilibrium as a singleton set is a solution for N1-N2. However, the sets E_1 and E_2 should be based only on the information of the game structure G. This implies that we should look for the pair of greatest sets (E_1, E_2) satisfies N1-N2[3]. The following theorem characterizes conditions for the greatest pair to exist and and strategies in that pair in terms of Nash solutions. In the theorem, E is a subset of $S_1 \times S_2$ and $E_i = \{s_i : (s_i; s_j) \in E \text{ for some } s_j \in S_j\}$ for $i = 1, 2$.

Theorem 3.2 (The Nash Noncooperative Solutions): (0): G has a Nash equilibrium if and only if there is a nonempty pair (E_1, E_2) satisfying N1-N2.

(1): Suppose that G is solvable. Then the greatest pair (E_1, E_2) satisfying N1-N2 exists and $E = E_1 \times E_2$ is the Nash solution $E(G)$.

(2): Suppose that G has a Nash equilibrium but is unsolvable. Then E is a Nash subsolution if and only if (E_1, E_2) is a nonempty maximal pair satisfying N1-N2.

[3] If any additional information is available, then we extend N1-N2 to include it and should consider the pair of greatest sets satisfying the new requirements.

Proof. (0): If (s_1, s_2) is a Nash equilibrium of G, then $E_1 = \{s_1\}$ and $E_2 = \{s_2\}$ satisfy N1-N2. Conversely, a nonempty pair (E_1, E_2) satisfies N1-N2. By Lemma 3.1, any pair $(s_1, s_2) \in E_1 \times E_2$ is a Nash equilibrium of G.

(1):(If): Let (E_1, E_2) be the greatest pair satisfying N1-N2. It satisfies to show $E(G) = E_1 \times E_2$. By Lemma 3.1, any $(s_1, s_2) \in E_1 \times E_2$ is a Nash equilibrium. Conversely, let $(s_1', s_2') \in E(G)$ and $E_i' = \{s_i'\}$ for $i = 1, 2$. Since this pair (E_1', E_2') satisfies N1-N2, we have $(s_1', s_2') \in E_1' \times E_2' \subseteq E_1 \times E_2$. Hence, $E(G) = E_1 \times E_2$.

(Only-If): Since E is the Nash solution, it satisfies (2.2). Hence, E is expressed as $E = E_1 \times E_2$ by Lemma 2.1. Since it consists of Nash equilibria, (E_1, E_2) satisfies N1-N2 by Lemma 3.1. Since $E(G) = E = E_1 \times E_2$, (E_1, E_2) is the greatest pair having N1-N2.

(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 Nash equilibria. Let E' be a set of Nash equilibria satisfying (2.2) with $E_1 \times E_2 \subseteq E'$. Then, E' is also expressed as $E_1' \times E_2'$. Since $E_1' \times E_2'$ satisfies N1-N2 by Lemma 3.1, we have $E_i' \subseteq E_i$ for $i = 1, 2$ by maximality for (E_1, E_2). By the choice of E', we have $E_1 \times E_2 = E'$. Thus, E is a maximal set satisfying interchangeability(2.2).

(Only-If): Since E is a subsolution, it satisfies (2.2). Hence, E is expressed as $E = E_1 \times E_2$. Also, by Lemma 3.1, (E_1, E_2) satisfies N1-N2. Since $E = E_1 \times E_2$ is a subsolution, (E_1, E_2) is a maximal set satisfying N1-N2. ■

The pair (E_1, E_2) satisfying N1-N2 consists of the empty sets if there is no Nash equilibrium in G. When G has a Nash equilibrium but is unsolvable, there are multiple pairs of maximal sets (E_1, E_2) satisfying N1-N2. We do not have those problems in NM1-NM2 in Section 2.3, for which the greatest pair always exists and is nonempty. It may be the reason for this difference that N1-N2 are interactive but NM1-NM2 are not at all. In this respect, the theory of rationalizable strategies, to be discussed in Section 3.2, is similar to NM1-NM2, though it is more interactive than NM1-NM2.

In the case of an unsolvable game G with a Nash equilibrium, there are multiple candidate sets of possible decisions and predictions, even though the decision criterion and game structure are commonly understood between the players. Each maximal pair (E_1, E_2) satisfying N1-N2 may be a candidate, but it requires further information for the players to choose among them. Thus, N1-N2 alone is not sufficient to provide a definite recommendation in unsolvable games. Theorem 3.2 gives a demarcation between the cases of having a definite recommendation and not.

One possible way to reach a recommendation for an unsolvable game is to impose additional criterion, such as the symmetry requirement in Nash [17], to select a certain subset of Nash equilibria. The game of Table 2.2 is unsolvable, but it has a unique symmetric equilibrium $(\mathbf{s}_{11}, \mathbf{s}_{21})$. Hence, if we add the symmetry criterion, we convert an unsolvable game to a solvable game.

Table 3.1

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$
$\mathbf{s}_{11}$	$(5,5)^n$	$(0,5)^n$
$\mathbf{s}_{12}$	$(5,0)^n$	$(0,0)^n$

Table 3.2

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$
$\mathbf{s}_{11}$	${}^n(5,-5)$	${}^n(0,-5)$
$\mathbf{s}_{12}$	${}^n(5,0)$	${}^n(0,0)$

Another possible criterion is Pareto-optimality. In the thought process of decision making, the players may add the (strong) Pareto-criterion to their decision criterion. In the game of Table 3.1, $(\mathbf{s}_{11}, \mathbf{s}_{21})$ (weakly) Pareto-dominates the other equilibria, and $(\mathbf{s}_{12}, \mathbf{s}_{21})$ does in the game of Table 3.2. We obtain a unique decision in both games. This suggests a possibility to obtain the value of the 2-person game for each player, as discussed in Nash [17]. Indeed, in general, if the Nash solution exists in a 2-person game, we can select a unique payoff vector by the Pareto criterion.

To achieve solvability by adding additional criteria seems difficult in general. Nevertheless, N1-N2 serves the starting point which allows further investigation of their compatibility with additional principles in specific classes of games, which may become a fruitful direction for future research.

One alternative to obtain a definite recommendation in unsolvable games other than additional criteria is to introduce pre-play communication between the players. This requires a development of a language to communicate about which subsolution would be played. This approach, however, meets conceptual issues regarding modeling communication. The game of Table 3.3 has three subsolutions indexed by $(1), (2), (3)$: To communicate which subsolution would be played requires the information of all the elements of the targeted subsolution. The success of such a communication depends upon the choice of names or language referring to subsets of strategies or subsolutions. In this paper, we do not touch this problem.

Table 3.3

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$	$\mathbf{s}_{23}$
$\mathbf{s}_{11}$	${}^{(1)}(1,1)$	${}^{(1)}(1,1)^{(2)}$	$(0,0)$
$\mathbf{s}_{12}$	$(0,0)$	${}^{(3)}(1,1)^{(2)}$	${}^{(3)}(1,1)$

3.2. Rationalizable Strategies

Let us consider the following modification of N1-N2: for E_1 and E_2,

R1: for each $s_1 \in E_1$, Best$(s_1; s_2)$ holds for some $s_2 \in E_2$;

R2: for each $s_2 \in E_2$, Best$(s_2; s_1)$ holds for some $s_1 \in E_1$.

This criterion differs from N1-N2 only in that the quantifier "for all" before players" predictions in N1-N2 is replaced by "for some". In fact, R1-R2 is the pure-strategy version of the BP-property given by Bernheim [3] and Pearce [18]. The greatest pair (E_1, E_2) satisfying R1-R2 exists and 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 [3] (Proposition 3.1); we include the proof for self-containment.

Theorem 3.3 (Rationalizability): $(R_1(G), R_2(G))$ is the greatest pair satisfying R1-R2.

Proof. Suppose that (E_1, E_2) satisfies R1-R2. 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-R2, there is an $s_j \in E_j$ such that $\mathrm{Best}(s_i; s_j)$ holds. Because $E_j \subseteq R_j^\nu(G)$, we have $s_j \in R_j^\nu(G)$. Thus, $s_i \in R_i^{\nu+1}(G)$.

Conversely, we show that $(E_1(G), E_2(G))$ satisfies R1-R2. 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 $\mathrm{Best}(s_i; s_j^\nu)$ 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 $R_j(G) = \bigcap_{\nu=0}^{\infty} R_j^\nu(G)$. Also, $\mathrm{Best}_i(s_i; s_j^*)$ holds. Thus, $(R_1(G), R_2(G))$ satisfies R1-R2. ■

Existence of a Theoretical Prediction: Theorem 3.3 and Lemma 2.2 imply that the greatest pair satisfying R1-R2 exists and consists of nonempty sets. Interchangeability is automatically satisfied by construction. In this respect, the rationalizability theory may appear preferable to the Nash theory, since it avoids issues due to the emptiness or nonexistence of the Nash solution. However, we can/should take a different view: Emptiness or nonexistence involved in the Nash theory may help identify situations where additional principles other than best-response against predictions are required to obtain a recommendation. The Nash theory may be more useful than the rationalizability theory in that it demarcates between those two cases. We will return to this issue once more in Section 4.2.

Set-theoretical Relationship to the Nash Solutions: It follows from Theorem 3.3 that each strategy of a Nash equilibrium is a rationalizable strategy. Hence, the Nash solution, if it exists, is a subset of the set of rationalizable strategy profiles. However, the converse does not necessarily hold. Indeed, consider the game of Table 3.4, where the subgame determined by the *2nd* and *3rd* strategies for both players is the "matching pennies".

Table 3.4

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$	$\mathbf{s}_{23}$
$\mathbf{s}_{11}$	$(5, 5)$	$(-2, -2)$	$(-2, -2)$
$\mathbf{s}_{12}$	$(-2, -2)$	$(1, -1)$	$(-1, 1)$
$\mathbf{s}_{13}$	$(-2, -2)$	$(-1, 1)$	$(1, -1)$

Table 3.5

	$\mathbf{s}_{21}$	$\mathbf{s}_{22}$	$\mathbf{s}_{23}$
$\mathbf{s}_{11}$	$(5, 5)$	$(1/2, 1/2)$	$(1/2, 1/2)$
$\mathbf{s}_{12}$	$(1/2, 1/2)$	$(1, -1)$	$(-1, 1)$
$\mathbf{s}_{13}$	$(1/2, 1/2)$	$(-1, 1)$	$(1, -1)$

This game has a unique Nash equilibrium, $(\mathbf{s}_{11}, \mathbf{s}_{21})$. Hence, the set consisting of this equilibrium is the Nash solution.

It follows from the above observation that both $\mathbf{s}_{11}$ and $\mathbf{s}_{21}$ are rationalizable strategies. Moreover, the other four strategies, $\mathbf{s}_{12}, \mathbf{s}_{13}$ and $\mathbf{s}_{22}, \mathbf{s}_{23}$ are also rationalizable: Consider $\mathbf{s}_{12}$. It is a best response to $\mathbf{s}_{22}$, which is a best response to $\mathbf{s}_{13}$, and $\mathbf{s}_{13}$ is a best response to $\mathbf{s}_{23}$, which is a best response to $\mathbf{s}_{12}$. That is, we have the following relations:

$$\text{Best}(\mathbf{s}_{12}; \mathbf{s}_{22}),\ \text{Best}(\mathbf{s}_{22}; \mathbf{s}_{13}), \text{Best}(\mathbf{s}_{13}; \mathbf{s}_{23}),\ \text{and Best}(\mathbf{s}_{23}; \mathbf{s}_{12}).$$

By Theorem 3.3, those four strategies are rationalizable. In sum, all the strategies are rationalizable in this game.

This example shows that even for solvable games, the Nash solution may differ from rationalizable strategies. As we shall see later, the game of Table 3.4 becomes unsolvable if mixed strategies are allowed, while the rationalizable strategies remain the same.

3.3. Mixed Strategy Versions

Theorems 3.2 and 3.3 can be carried out in mixed strategies without much difficulty. The use of mixed strategies may give some merits and demerits to each theory. Here, we give comments on the mixed strategy versions of the two theories.

The mixed strategy versions can be obtained by extending the strategy sets S_1 and S_2 to the mixed strategy sets $\Delta(S_1)$ and $\Delta(S_2)$; where $\Delta(S_i)$ is the set of probability distributions over S_i. Requirements N1-N2 are modified as follows: for $E_i \subseteq \Delta(S_i)$, $i = 1, 2$,

N1^m: for each $s_1 \in \sigma(E_1)$, Best$(s_1; m_2)$ holds for all $m_2 \in E_2$,

N2^m: for each $s_2 \in \sigma(E_2)$, Best$(s_2; m_1)$ holds for all $m_1 \in E_1$,

where $\sigma(E_i)$ is the support of E_i, i.e., $\{s_i \in S_i : m_i(s_i) > 0 \text{ for some } m_i \in E_i\}$, for $i = 1, 2$. As stated above, the rest of the discourse is modified by replacing S_i with $\Delta(S_i)$ for $i = 1, 2$. That is, we can modify Theorem 3.2 by replacing pure strategies by mixed strategies.

The mixed strategy version of rationalizability is the original in Bernheim [3] and Pearce [18]. Especially, the use of mixed strategies is crucial for their interpretation of beliefs about the other player's decisions. The pure strategy version discussed in this paper is known as *point-rationalizability* due to Bernheim [3]. However, the interpretation of mixed strategies is not the main issue in this paper.

The mixed strategy version of rationalizability is also characterized by modifying N1^m-N2^m as follows: for $E_i \subseteq \Delta(S_i)$, $i = 1, 2$,

R1^m: for each $s_1 \in \sigma(E_1)$, Best$(s_1; m_2)$ holds for some $m_2 \in E_2$;

R2^m: for each $s_2 \in \sigma(E_2)$, Best$(s_2; m_1)$ holds for some $m_1 \in E_1$,

Theorem 3.3 can be obtained for R1^m-R2^m without much difficulty.

A simple observation is that a rationalizable strategy in pure strategies (i.e., in the sense of Section 2) is a rationalizable strategy in mixed strategies. Similarly, since a Nash equilibrium in pure strategies is also a Nash equilibrium in mixed strategies, it may be conjectured that if a game G has the Nash solution E in the pure strategies, it might be a subset of the Nash solution in mixed strategies. In fact, this conjecture is answered negatively.

Consider the game of Table 3.4. This game has seven Nash equilibria in mixed strategies:

$$((1,0,0),(1,0,0)),\ ((0,\tfrac{1}{2},\tfrac{1}{2}),(0,\tfrac{1}{2},\tfrac{1}{2})),\ ((\tfrac{4}{18},\tfrac{7}{18},\tfrac{7}{18}),(\tfrac{4}{18},\tfrac{7}{18},\tfrac{7}{18}))$$
$$((\tfrac{1}{8},\tfrac{7}{8},0),(\tfrac{3}{10},\tfrac{7}{10},0)),\ ((\tfrac{1}{8},0,\tfrac{7}{8}),(\tfrac{3}{10},0,\tfrac{7}{10})),\ ((\tfrac{3}{10},\tfrac{7}{10},0),(\tfrac{1}{8},0,\tfrac{7}{8})),\ ((\tfrac{3}{10},0,\tfrac{7}{10}),(\tfrac{1}{8},\tfrac{7}{8},0)).$$

This set does not satisfy interchangeability (2.2). For example, $((1,0,0),(1,0,0))$ and $((0,\frac{1}{2},\frac{1}{2}),(0,\frac{1}{2},\frac{1}{2}))$ are Nash equilibria, but $((0,\frac{1}{2},\frac{1}{2}),(1,0,0))$ is not a Nash equilibrium. Thus, (2.2) is violated, and the set of all mixed strategy Nash equilibria is not the Nash solution.

This result depends upon the choice of payoffs: In Table 3.5, $(\mathbf{s}_{11},\mathbf{s}_{21})$ is the unique Nash equilibrium in mixed strategies, while all strategies are still rationalizable.

4. Evaluations of N1-N2 and R1-R2 as Prediction/Decision Criteria

In our unified approach, we found that the difference between the Nash and rationalizability theories is the choice of quantifier "for all" or "for some" for each player's predictions. Based on this, we evaluate these theories from the viewpoint of *ex ante* individual decision making, and their logical coherence with the conceptual bases of game theory. First we consider the principles of prediction/decision making in general and focus on the distinction between decision and prediction and the resulted infinite regress in particular. Then, we return to the pinpointed difference between the two theories. We start with viewing Johansen's [10] argument on the Nash theory, which is the first attempt to understand *ex ante* decision/prediction making in game theory. Our discussions help clarify some of his postulates, and *vice versa*.

4.1. Johansen's Argument

Johansen [10] gives the following four postulates for decision/prediction making in games and assert that the Nash noncooperative solution can be derived from those postulates[4]

[4]He assumed that the game has the unique Nash equilibrium for his assertion (p.435), but he noted that interchangeability is actually enough (p.437) for it.

for solvable games.

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 any[5] 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.

Here we understand Johansen's assertion that the Nash solution is characterized by these postulates as the statement of Theorem 3.2.(1), though Johansen [10] does not give a mathematical formulation of argument.

First, we clarify a key difference between Johansen's [10] argumentation and other justifications for solution concepts in the game theory literature. In many other justifications, the term "rationality" is synonymous to "payoff maximization." This is not the case for Johansen's, since the two terms "rational" and "payoff maximization against prediction" appear in J2 and J4. Indeed, it is more faithful to his argumentation to regard these four postulates together as an attempt to define "rationality", while "payoff maximization" is just a component of it[6].

We now compare J1-J4 with N1-N2 and R1-R2. Postulate J1 is well taken in N1-N2 and R1-R2 if we ignore its subtle part "only on the basis of...", because both criteria are described only with the components of the game structure $G = (N, \{S_i\}_{i \in N}, \{h_i\}_{i \in N})$. We will revisit this subtle part a few times below. Postulate J2 requires the decision criterion be symmetric between one player and his imaginary player; this is an intrapersonal coherency requirement in the mind of each player. Postulate J3 requires each player's prediction about the other's decision be correctly made; this is an interpersonal coherency requirement. Both systems N1-N2 and R1-R2 satisfy these two requirements. Finally, Postulate J4 corresponds to the requirement that actions in E_i maximize player i's payoff against elements in E_{-i} predicted by player i; the difference in the quantifiers before the predicted decisions that appear in N1-N2 and R1-R2 will be discussed in great detail later.

Postulate J1 is relevant in several different places in our approach. One apparent

[5] This "any" was "some" in Johansen's orginal Posutlate 3. According to logic, this should be "any". However, this is mistakenly expressed as "some" in many scientists (even mathematicians).

[6] This may be the reason for Bernheim's [4], p.486, criticism against these postulates as too ambiguous to avoid various different ways of reading. We will comment on the "common knowledge of rationality" for rationalizability in the companion paper [8].

place is the consideration of the greatest set satisfying N1-N2 in Theorem 3.2: A choice of a proper subset needs some additional information to the given game structure. We will return to Postulate J1 when it becomes relevant.

The difference between J2 and J3 is delicate and we give some explanation here. Take the criterion N1-N2 (the argument is for R1-R2 is symmetric). Postulate J2 implies that from player 1's perspective, N2 is symmetric to N1, while J3 requires that if player 1 uses N2 for his prediction, i.e., he believes that player 2's decisions are determined by N2, player 2 actually uses N2 for his decision. If we drop J2, it would be possible to have the combination N1-R2 for player 1: That is, player 1 uses N1 for his decision making but R2 for his prediction making. Similarly, we can assume that player 2 uses R2 for his decision making and N1 for prediction making. These violate J2 but still satisfy J3[7]. Criterion NM1-N2 in Section 2 can be interpreted in a similar manner.

On the other hand, the elimination of J3 allows the possibility that the players use entirely different criteria: For example, 1 adopts N1-N2, while 2 adopts R1-R2. These possibilities require arguments about interpersonal thinking that are beyond the scope of the present framework, but will be discussed in the companion paper [8].

4.2. Prediction/Decision Criterion

Here, we discuss some principles for prediction making and decision making, and highlight the inevitable infinite regress in decision/prediction making in a game situation with N1-N2 or R1-R2. For this discussion, the difference between N1-N2 and R1-R2 is not significant; we focus on N1-N2, but will comment on R1-R2 as we proceed.

Prediction Making (Putting Oneself in the Other's Shoes): System N1-N2 is understood as describing both prediction making and decision making: from player 1's perspective, N1 requires a property for his possible decisions E_1, but it includes his predictions about player 2's decisions E_2. N1 alone does not determine E_1, but needs other criterion to determine E_2. Player 1 makes prediction about 2's possible decisions, by putting himself into player 2's shoes. This argument could not stop here; since 1's prediction about player 2's predictions affects 1's own ultimate decisions, player 1 needs to know how player 2's predictions are made, and to assume that 2 assumes that 1 uses N1. Continuing this argument *ad infinitum,* we meet the infinite regress described in Diagram 4.1: The determination of E_1 by N1 needs the determination of E_2 by N2, which needs the further determination of E_1 by N1, and so on. This infinite regress is encountered by R1-R2 as well.

Double Uses of N1-N2: The infinite regress in Diagram 4.1 is made from the viewpoint of player 1's decision making. A symmetric argument from player 2's viewpoint is constructed. Here, N1 is a decision criterion for 1 and is a prediction criterion for 2,

[7] So far, we have no example where N1-R2 yields a different outcome from either N1-N2 or R1-R2.

while N2 is a decision criterion for 2 and a prediction criterion for 1. Thus, both N1 and N2 are used both as decision and prediction criteria. This is required by J2, as discussed in Section 4.1.

Diagram 4.1

N1		N1		N1		$\cdots$
$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$	
N2		N2		N2		

$\Longrightarrow$

Diagram 4.2

N1
$\downarrow$ $\uparrow$
N2

Diagram 4.3

R1		R1		R1		$\cdots$
$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$	
R2		R2		R2		

$\Longrightarrow$

Diagram 4.4

R1
$\downarrow$ $\uparrow$
R2

In our formalism, no explicit distinction can be made between player 1's and 2's perspectives, which remains in interpretation. Without this distinction, the infinite regress in Diagram 4.1 collapses into a system of simultaneous equations described by Diagram 4.2. As shown in Theorem 3.2, circularity in N1-N2 corresponds to the concept of Nash equilibrium. The theory of rationalizability is parallel in this respect; although the original definition of rationalizable strategies, given in Section 2, takes the form of Diagram 4.3, Theorem 3.3 states that it collapses to Diagram 4.4.

One way to avoid the collapses from Diagram 4.1 (3, respectively) into Diagram 4.2 (4) is reformulated our considerations in an epistemic logic framework, in which we can explicitly discuss the relationship between the above infinite regress and the common knowledge of N1-N2 (R1-R2). This will be given in the companion paper [8].

***Ex Ante* Decision Making, Inferences, and Solvability**: Given that our goal is to understand *ex ante* decision-making, each player has to reach his possible decisions based on his prediction about the other's decisions. To reach those decisions and predictions, each player infers the other's decisions as well as his own decisions from his criterion and knowledge of the game structure. Rationality is understood as an assumption about the criterion and his ability to conduct such inference.

This inference takes place in a player's mind. We allow him to reach multiple decisions from the inference; thus the outcome of the inference is the set of all such decisions. Moreover, for both systems N1-N2 and R1-R2, the outcome of this inference includes not only his final decisions but also prediction about the other's decisions.

In situations such as the Battle of Sexes game, individualistic decision making is incapable of recommending a set of definite decisions without communication between the players. Theorem 3.2 exactly demarcates between the case where individualistic decision making does serve a definite set of decisions (when the game is solvable) and the case where it does not (when the game is unsolvable). For example, according to Theorem 3.2.(2), the Nash solution does not give a definite recommendation in the Battle of Sexes game. On the other hand, the theory of rationalizability tells no difficulties: Indeed, the recommendation from R1-R2 is the set of rationalizable strategies, which is always nonempty.

4.3. The Free-will Postulate vs. Complete Determinism

The difference between N1-N2 and R1-R2 lies in the choice between the quantifiers 'for all' and 'for some' for one's predictions about the other's possible decisions. Here, we argue that this difference stems from a methodological choice between two meta-theoretical foundations: One foundation is the free-will postulate, and the other is complete determinism.

The Free-will Postulate: This is a very basic principle in game theory, stating that players are free to make any choice by their own will. Another basic assumption in game theory, utility maximization, may effectively void this postulate, but even if each player is very smart, it is still possible that one's own decision together with prediction about the other' decision cannot be completely determined. This possibility is first argued in Morgenstern [16]. Moreover, whenever the social science involves value judgements for an individual being and/or society, it relies on the free-will postulate as a foundation[8]. Here, we argue that the Nash theory is consistent with this postulate, while the rationalizability theory is encountered with some difficulties to be reconciled with this postulate.

First, we consider two applications of the postulate at two different layers in terms of interpersonal thinking:

(i): It is applied by the outside observer to the (inside) players;

(ii): It is applied by an inside player to the other player.

In application (i), the outside theorist respects the free will of each player; the theorist can make no further refinement than the inside player. This corresponds to the 'greatestness' requirement for (E_1, E_2) in Theorems 3.2.(1) and Theorem 3.3. In (ii), when one player has multiple predictions about the other's possible decisions, the free-will postulate, applied to interpersonal decision making, requires the player to take all possible predictions into account. Apparently, N1-N2 is formulated in this manner. However, R1-R2 violates the postulate in (ii); each player can use any arbitrary decision of the other to rationalize his own decision.

However, this analysis may not fully appreciate the methodological assumptions behind the rationalizability theory. Indeed, although, as for N1-N2, the rationalizability

[8] The free-will postulate is needed for deontic concepts such as responsibility for individual choice and also for individual and social efforts for future developments.

theory recommends the greatest set that satisfies R1-R2, its foundation differs from that for N1-N2. We argue in the following that it would be natural to interpret the rationalizability theory from the viewpoint of complete determinism.

Complete Determinism: The quantifier 'for some' in R1-R2 can have two different interpretations:

(a): it requires only the mere existence of a rationalizing strategy;

(b): it suggests a specific rationalizing strategy predetermined for some other reason.

From the definition of rationalizability alone, It is not possible to judge which of (a) and (b) is more faithful to the theory. Consider first interpretation (a): The rationalizing strategy is arbitrarily chosen among the candidate predictions, and this treatment reminds us the Aesops' *sour grapes*; the fox finds one convenient reason to persuade himself. Interpretation (a) states that the player can make a decision if he finds any reason for it. This interpretation of "rationalization" is at odds with the purpose of a theory of *ex ante* decision-making for games, for which the theory is serious about a best choice responding to prediction about the other's decisions. Such a theory is supposed to provide a rationale for players' possible decisions as well as predictions. Interpretation (a) avoids this issue, with which we cannot take rationalizability seriously as a theory of *ex ante* decision-making.

Interpretation (b) is more serious: According to (b), there are some further components, not explicitly included in the game description G and R1-R2, which determine a specific rationalizing strategy. Unless this rationalizing strategy is homogeneous for each step in Diagram 4.3, the rationalization process generates an infinite regress. A specific rationalizing strategy for each step is uniquely predetermined for some reason. Uniqueness is crucial, for otherwise the player would have to arbitrarily choose among different strategies or to look for a further reason to choose some of them. In this sense, (b) is understood as presuming a complete description of mental states about interpersonal thinking corresponding to the steps in Diagram 4.3. Indeed, there is a literature, beginning from Aumann [2][9] to justify the rationalizability theory or alike along this line (see Tan-Werlang [19]).

Taking the existential quantifier 'for some' seriously, we meet a deterministic view. Determinism has been tenanted as the foundational standpoint of natural science. However, Determinism is incompatible with the free-will postulate, and also the additional background information is incompatible with Johansen's postulate J1. To think about

[9] In the problem of common knowledge in the information partition model due to Robert Aumann, the information partitions themselves are assumed to be common knowledge. He wrote in [1], p.1237: "*Included in the full description of a state ω of the world is the manner in which information is imparted to the two persons*". This can be interpreted as meaning that the primitive state ω includes every information. A person receives some partial information about ω, but behind this, everything is predetermined. This view is shared with Harsanyi [6] and Aumann [2].

this incompatibility, we should reflect upon the deterministic view from the viewpoint of social science and science as a whole. For simplicity we shall consider only complete determinism.

It is inevitable to encounter value judgments in social science. However, if we take complete determinism, social problems can and should be analyzed purely from the viewpoint of causal relationships as in natural science; it is incapable of discussing value judgement and associated rights/obligations from this perspective. In this sense, complete determinism is not compatible with social science.

Besides this foundational difficulty, it is not very fruitful as a methodology for social science in general, which is aptly described by Hayek [7], Section 8.93: *"Even though we may know the general principle by which all human action is causally determined by physical processes, this would not mean that to us a particular human action can ever been recognizable as the necessary result of a particular set of physical circumstances."* In fact, complete determinism is justified only because of its non-refutability by withdrawing from concrete problems into its own abstract world.

In fact, neither complete determinism nor the free-will postulate can be justified by its own basis. Either should be evaluated with coherency of the entire scope and the scientific and/or theoretical discourse.

Starting Postulates for a Scientific Study and Basic Beliefs for a Player: A scientific study is an attempt not to describe every detail of the world, but to draw a simplified picture of a target problem so as to captures an (seemingly) essential structure of it. A discourse of a scientific study starts with basic postulates, with which a simplified picture of the target world is drawn. Even though the researcher may have more basic views or principles, each study must take this form of a discourse. For example, a study in chemistry needs its own starting postulates. The claim, made by complete reductionism, that every study in chemistry could be treated from the viewpoint of quantum mechanics is absurd from the viewpoint of practical research in chemistry. Similarly, complete determinism may explain practically nothing in human behavior, as quoted in Hayek [7]. We take the free-will postulate, which is automatically assumed when discussing decision-making, as a basic component of game theory.

Similar to the above argument, we view a player as having some understanding of the game structure which he is now playing. A normal form game $G = (N, \{S_i\}_{i \in N}, \{h_i\}_{i \in N})$ is regarded as abstracted from a real social situation by taking relevant information. Johansen's postulate J1 requires the discourse of a study of *ex ante* decision making starts with the descriptive elements in G. Player's understanding of G is parallel to the theorist's understanding of his theory. This is one possible starting assumption. From this perspective, we conclude that the Nash noncooperative theory is a faithful description of *ex ante* decision-making, while it would be natural to regard the rationalizability theory as presuming determinism and it does not give a coherent picture of decision-making for games.

5. Conclusions

We presented the unified framework of the Nash noncooperative theory and rationalizability theory. Then, we pinpointed that the difference between them is the choice of the quantifier "for all" or "for some" for predictions about the other player's possible decisions. In Section 4, we discussed various conceptual problems by viewing the quantifier "for all" or "for some" from the perspectives of the free-will postulate and complete determinism.

As already stated, in our current framework no formal distinction is made for one's predictions and the other's decisions. Similarly, the knowledge of the game structure and rationality is also purely interpretational here. To formalize the distinction between decisions and predictions, and to evaluate the (common) knowledge requirements more explicitly, we need a certain extended framework. In the companion paper [8], we will adopt the epistemic logic approach. Specifically, we will use the (propositional) *common knowledge logic* CKL. It enables us to study the meaning of the common knowledge of the structure of the game and the player's rationality, stated in the quotation from Mas-Collel, *et al* [13]. We can also discuss the relationship between the infinite regress mentioned in Section 4.2 and the common knowledge of N1-N2 (or R1-R2).

The CKL approach sheds different lights on the problems of the free-will postulate and/or complete determinism. We adopt the proof theoretical (syntactical) formulation of the logic, where each player is facilitated with some logical inference ability and infers logical conclusions for his decisions from his basic beliefs. This view is coherent to what we described in Section 4.3. Also, we provide the semantic approach to prove some unprovability results. In sum, the logical approach is needed so as to have more explicit and extensive discussions on the problem of *ex ante* individual decision making than in those in Section 4.

References

[1] Aumann, R. J., (1976), Agreeing to Disagree, *Annals of Statistics* 4, 1236-1239.

[2] Aumann, R. J. (1987), Correlated Equilibrium as an Expression of Bayesian Rationality? *Econometrica* 55, 1-18.

[3] Berheim, D. (1984), Rationalizable Strategic Behavior, *Econometrica* 52, 1007-1028.

[4] Berheim, D. (1986), Axiomatic Characterizations of Rational Choice in Strategic Environments, *Scan. J. of Economics* 88, 473-488.

[5] Brandenburger, A., and E. Dekel (1993), Hierarchies of Beliefs and Common Knowledge, *Journal of Economic Theory*, 59, 189-198.

[6] Harsanyi, J. C. (1967/68), Games with Incomplete Information Played by Bayesian Players, I-III, *Management Science* 14, 159-182, 320-334, 486-502.

[7] Hayek, F. A., (1952), *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology*, The University of Chicago Press, (1952).

[8] Hu, T., and M. Kaneko (2012), Common Knowledge Logic Approach to the Nash Noncooperative Theory and Rationalizability, being prepared.

[9] Jansen, M. J. M., (1981), Maximal Nash Subsets for Bimatrix Games, *Naval Research Logistics Quarterly* 28, 147-152.

[10] Johansen, L., (1982), On the Status of the Nash Type of Noncooperative Equilibrium in Economic Theory, *Scand. J. of Economics* 84, 421-441.

[11] Kaneko, M., (1999), Epistemic Considerations of Decision Making in Games. *Mathematical Social Sciences* 38, 105–137.

[12] Kaneko, M., (2004), *Game Theory and Mutual Misunderstanding*, Springer, Heidelberg.

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

[14] Mertens, J., and S. Zamir (1985), Formulation of Bayesian analysis for games with incomplete information, *International Journal of Game Theory* 14, 1-29.

[15] Millham, C. B., (1974), On Nash Subsets of Bimatrix Games, *Naval Research Logistics Quarterly* 21, 307-317.

[16] Morgenstern, O., (1935), Perfect Foresight and Economic Equilibrium. Zeitschrift für Nationalökonomie 6, 337-357.

[17] Nash, J. F., (1951), Non-cooperative Games, *Annals of Mathematics* 54, 286-295.

[18] Pearce, D., (1984), Rationalizable Strategic Behavior and the Problem of Perfection, *Econometrica* 52, 1029-1050.

[19] Tan T. C., and S. R. D. C. Werlang, (1988), The Bayesian Foundations of Solution Concepts of Games, *Journal of Economic Theory* 45, 370-391.

[20] von Neumann, J. and O. Morgenstern, (1944), *Theory of Games and Economic Behavior*, Princeton University Press, Princeton.