Discussion Papers In Economics And Business

Social Recognition and Economic Equilibrium

Ken Urai

Discussion Paper 06-30

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

Social Recognition and Economic Equilibrium

Ken Urai

Discussion Paper 06-30

December 2006

この研究は「大学院経済学研究科・経済学部記念事業」
基金より援助を受けた、記して感謝する。

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

Social Recognition and Economic Equilibrium *

Ken Urai†

Abstract

This paper is an attempt to incorporate the human ability of recognition, especially, the ability to recognize the society to which they belong, with the economic equilibrium theory characterized by a description of society through individual rational behaviors. Contents may be classified into the following three categories: (1) a rigorous set theoretical treatment of the description of individual rationality; (2) set theoretical description of the validity in a society; and (3) rationality as an equilibrium (fixed point) of social recognition.

Keywords : Social Recognition, Rationality, Social Equilibrium, Fixed Point Theorem, Gödel's Incompleteness Theorem.

JEL classification: A10; B40; C60

*I wish to thank professor Mamoru Kaneko (Tsukuba University) for his many researches and enlightenment works on foundation of game theory and social science with which the author is much inspired. My thanks are also due to an anonymous referee of Advances in Mathematical Economics, who has pointed out several mathematical mistakes in the earlier draft of this paper.

†Graduate School of Economics, Osaka University, Toyonaka, Osaka 560-0043, JAPAN. (Internet e-mail address: urai@econ.osaka-u.ac.jp)

1 INTRODUCTION

This work is an attempt to incorporate the human ability to recognize or grasp the world to which they belong with set theoretical view of the society as the totality of rational individuals given by contemporary game and economic theories. In this paper, results are classified into three categories: problems on individuals and rationality (Section 2), society and values in it (Section 3), and equilibrium and social recognition (Section 4).

Section 2, "Individuals and Rationality," is a formal treatment of set theoretic limitations in describing society as consisting of rational individuals. It is shown that the concept of rationality, at least in the sense of the rational acceptability of sentences in a certain formal language, cannot completely be described as long as we require it to be both semantically (introspectively) and logically consistent. The result is obtained as a variation of Tarski's truth definition theorem, a closely related result to Gödel's second incompleteness theorem. I am convinced that the argument could serve as a launching pad for rigorous mathematical treatments on the problematic cognitive features of all mathematical models in social science based on methodological individualism.

Section 3, "Society and Values," deals with problems of describing human society as a whole from macroscopic viewpoints. The description of society, in the sense of a collection of sentences valid for descriptions of society in a certain formal language, cannot be semantically consistent (introspectively complete), as long as we require it to be logically consistent. It follows that for logical consistency, we cannot define society in an introspectively complete manner. In other words, we cannot assure rightness on the validity for descriptions of society, (not to speak of optimality, efficiency, etc.), except for believing it.[1] The results may also be considered a mathematical critique of empiricism (logical positivism) as a methodology of social science.[2]

We may relate the results in Sections 2 and 3 to our ordinary way of "justification" (set theoretically, an *internal* condition — consistency in a view of the world as an *extensional* condition — for true rationality, if it exists,) and "refutability" (that forms an *external* condition — inconsistency on a view of the world for an *intensional* condition — for rationality). Section 4 incorporates these two features into a social equilibrium argument, so that 'rationality,' as a non-refutable justification, is assured to exist as a fixed point (equilibrium) of recognition of the world for each member of society.

The contents in Sections 2 (individual rationality) and 3, (social value) are based on my earlier papers written in 2002 (Urai (2002a), Urai (2002b), and Urai (2002c)), including several important mathematical corrections on theorems as well as in proofs. Section 4 was originally prepared as a separate article on equilibrium. Hence, it would also be possible for readers to read each section independently.

2 INDIVIDUAL AND RATIONALITY

In this section, we see that there is no set theoretical formal description of human society that incorporates our quite natural and important kind of inference (recognition) ability. There are many

[1]The problem is based on the difficulty of defining "judgements for facts" since it should also determine our value judgements concerning what our "facts" are. According to H. Putnam, targets of such value judgements are called *epistemic values* (Putnam, 2002; p. 30) whose existence expounds the collapse of the classical fact/value dichotomy. We should note that mathematical models in social science always implicitly or explicitly presuppose one such value.

[2]W.V.O. Quine in his famous essay, "Two dogmas of empiricism," in 1951 (reprinted in Quine (1953)) treated the problem as the difficulty of defining analyticity.

causes for the impossibility to obtain a 'complete' social model in the sense that every feature of the world is completely described. Indeed, standard economic theory admits many types of 'externality.' There are many unknown structures in the real world, especially in technologies, information, preferences, and expectations, etc. It seems, however, that such problems have been recognized by theorists as merely the gap between an idealized economic model and reality. What I am concerned with here is not the gap between them but the impossibility of the notion of an *idealized model* itself.

If the purpose of economics is to describe human society as theoretical and well-founded mechanisms of 'rational' individuals, an economic model should formalize a *system of rules* that enables each agent's behavior to be called 'rational.' In order to formalize such economic 'rationality,' however, we should premise a restricted view on individual prospects or thoughts about the whole world. If we don't, as we shall see in the following, the view of the world necessarily becomes inconsistent (hence, every action is rational for him). On the other hand, with such a restricted view of the world, agents are not allowed to ask whether the world is exactly as they are thinking (in their view of the model). In other words, a consistent view (description) of the world must be incomplete in the sense that every agent should be convinced in the rightness of the view itself without any proofs.

The result in this section is related to Gödel's second incompleteness theorem. Indeed, the main theorem in this section may be considered a generalized version of Tarski's truth definition theorem, which is another important result of Gödel's lemma for the incompleteness theorem.[3] Note, however, that there is an important difference between the foundation of mathematics (Gödel's theorem) and the foundation of our view on society including ourselves. The former is a problem on what mathematics can do to formalize rationality, and the latter is an argument for formalizing rationality itself. We may change and reconstruct mathematics through our convictions and beliefs. To formalize ourselves, however, any restricted formalization may fail to characterize our total recognition ability; there isn't any simple way or regular routine to formalize our general intelligence.

In this section, *rationality* is treated as an attitude to accept certain kinds of formal assertions written in formal language.[4] The syntax for such a language and semantics (especially for the meanings of rationality) are given by a theory of sets $\mathscr{B} = (L_B, R_B, T_B)$ called an *underlying theory of sets*.[5] We assume that each person i using his/her formal language has a theory $\mathscr{L}_i = (L_i, R_i, T_i)$ that is at least as strong as the underlying theory of sets $\mathscr{B}$.[6] Thus, we are modeling a situation where person i can treat his/her assertion θ in the language of i (theory $\mathscr{L}_i$) as a set theoretic object $\ulcorner\theta\urcorner$ through basic underlying theory $\mathscr{B}$. The problem we treat in this section is whether we may construct a *formula* $P_i(x)$ *of person* i in one free variable x such that $P_i(\ulcorner\theta\urcorner)$ means that θ is a rationally acceptable assertion of i. Of course the answer depends on properties requested for the meanings of 'rational acceptability.' What we are concerned with here are logical *consistency* ($P_i(\ulcorner\theta\urcorner)$ and $P_i(\ulcorner\neg\theta\urcorner)$ never occur simultaneously) and introspective *completeness* ($P_i(\ulcorner\theta\urcorner)$ means $P_i(\ulcorner P_i(\ulcorner\theta\urcorner)\urcorner)$). The main theorem in this section shows that there is no P_i satisfying both of these two critical properties (Theorem 2.3). Theorems in

[3]Mathematical concepts in this section may be found in the standard literature in mathematical logic and/or theory of sets, e.g., see Kunen (1980), Jech (2003), Fraenkel et al. (1973).

[4]Throughout this section, I use linguistic definitions and approaches that may be common in classical arguments in analytical philosophy. From the standpoint of our notions of rationality and truth, however, I am influenced by the recent works of H. Putnam (after Putnam (1983)) and works in cognitive science such as Lakoff (1987), etc.

[5]A precise definition will be given in Section 2.

[6]Of course there must be an appropriate translation between his/her formal language and the language for $\mathscr{B}$.

this section show:

(1) The description of the world under notion P_i cannot be complete as long as we require P_i to be consistent. (Theorem 2.3.)

(2) Especially, we cannot introspectively recognize the consistency and the completeness (of our world view) itself. (Theorem 2.1, Theorem 2.2.)

(3) We cannot define (completely describe) rationality as long as we require it to be consistent. (Theorem 2.3.)

Therefore, all rational economic agents in a standard economic model should believe in their rational choices without knowing whether being rational is (truly) rational. All players in non-cooperative game theory should believe in their own rational behavior as well as their opponents' without without knowing what rationality exactly means. This seems to be a failure in all mathematical models in social science based on the *methodological individualism*. Indeed, the concept of 'rational individual' (consistency) always prevents us from providing a satisfactory answer to the question: 'What exactly is society?' (introspection) (see, Theorem 2.2 (b) and (d)). Therefore, every such agent naturally fails to possess self confidence in his/her rationality. Of course this is not saying that all attempts to describe society as the totality of rational individuals are meaningless. The result does suggest, however, that such attempts can never be completed, even in an asymptotic sense, and that we must allow for the relation between our recognition abilities and our views of the world.

2.1 View of the World

In this paper, we treat explicitly each agent's reasoning to chose an action by identifying individual rationality with consistency of a view of the world. Let $I = \{1, 2, \ldots, m\}$ be the index set of agents. For each $i \in I$, denote by A_i the set of possible actions for agent i. Each action profile, $(a_1, a_2, \ldots, a_m) \in \prod_{i \in I} A_i$, in the economy decides a consequence, c_i, in a set, C_i, for each $i \in I$.

In standard economic arguments and non-cooperative game theory, there are stories (mathematical structures), equilibrium and solution concepts, that enable for each agent i to have a reason for his/her choices of an action a_i. Since there are lots of reasons for (mutually exclusive) actions to be chosen, there may also exist many equilibrium and solution concepts. The rationality (the reason) in this sense crucially depends on the view of the world (equilibrium or solution concept). The purpose of this section is to show that this type of rationality is completely different from our 'true' rationality (thinking) and that the use (merely a part) of our true rationality may lead us to deny any such a specific view of the world and the rationality in the restricted sense.

In the following, we suppose that agent i has a theory (written by a formal language) $\mathscr{L}_i = (L_i, R_i, T_i)$ for obtaining a reason to decide an action a_i. L_i is the list of all symbols for the language, R_i is the list of all syntactical rules including construction rules for terms, formulas, and all inference rules (making a consequent formula from original formulas, e.g., modus ponens, instantiation, etc.), and T_i is the list of all axiomatic formulas for the theory. We assume that each element of L_i may be uniquely identified

with (coded into) an object in a certain basic theory of sets, $\mathscr{B} = (L_B, R_B, T_B)$, written under the first order predicate logic. We call $\mathscr{B}$ an *underlying theory of sets* for $\mathscr{L}_i$.[7]

The first important assumption of this section is that such a set theory is so basic that every agent could develop (understand) it by their own language.

> (A.1) The theory $\mathscr{L}_i = (L_i, R_i, T_i)$ is at least as strong as $\mathscr{B} = (L_B, R_B, T_B)$.[8] (Here, we implicitly assume that there is an appropriate translation between the languages for $\mathscr{L}_i$ and $\mathscr{B}$. Throughout this section, such a translation is assumed to be fixed, and we suppose that each formula φ in $\mathscr{B}$ could be identified with "the same" formula in $\mathscr{L}_i$ without loss of generality.)

The second assumption in this section is that though the theory, $\mathscr{L}_i = (L_i, R_i, T_i)$, of i may be stronger than $\mathscr{B} = (L_B, R_B, T_B)$, the structure of theory $\mathscr{L}_i$, i.e., each rules in list R_i is written in the language of the underlying theory of sets, $\mathscr{B}$. More precisely;

> (A.2) $\mathscr{B}$ describes $\mathscr{L}_i$ in the following sense: (i) Each member of list L_i is a term (a set) in theory $\mathscr{B}$. (ii) List R_i consists of formulas in theory $\mathscr{B}$. Especially, there are formulas in one free variable, $Term_i(x)$, $Form_i(x)$, $Form_i^1(x)$, in two free variables, $Neg(x, y)$, in three free variables, $Sbst(x, y, z)$, in the language of $\mathscr{B}$, maintaining, respectively, that in $\mathscr{L}_i$, x is a term, a formula, a formula in one free variable, a negation formula of formula y, a substitution formula of term z into the single free variable of formula y, based on descriptions of construction rules for them written in theory $\mathscr{B}$.[9] Every inference rule, as a relation among formulas of i, is also written in the language of $\mathscr{B}$ as a well defined set theoretic procedure. (iii) $Axiom_i(x)$ which defines formulas of i belonging to list T_i is written in the language of $\mathscr{B}$ as a well defined set theoretic procedure.

Assumption (A.2) is intended to be a sufficient condition that a combination of inference procedures, such as a proof procedure in theory $\mathscr{L}_i$, may be identified with a set theoretic procedure written in the form of a formula in theory $\mathscr{B}$. It should be noted that each term, formula, and inference procedure (including the proof procedure) of i may not be finitistic (recursive) since the set theoretic methods in $\mathscr{B}$ may be much stronger than the finitistic method. Description for them, however, are given in the language of $\mathscr{B}$ as set theoretical objects and processes that are well defined in set theory $\mathscr{B}$.

Under (A.1) and (A.2), an agent i is possible to treat an assertion (formula) θ in the language of i (theory $\mathscr{L}_i$) as a set theoretic object $\ulcorner\theta\urcorner$ through the underlying theory of sets, $\mathscr{B}$.[10] In the following, we call the theory, $\mathscr{L}_i = (L_i, R_i, T_i)$, satisfying these two assumptions, (A.1) and (A.2), the *world view* of i. The world view may include many features of the real world by adding additional axioms and syntactical rules, if necessary, and we suppose that an agent i chooses a 'rational' action $a_i \in A_i$ under the world view, $\mathscr{L}_i$. The third assumption is on the possibility of such a structure in the world view deciding the 'rationality'.

---

[7] The reader may identify $\mathscr{B}$ with ZF, Zermelo-Fraenkel set theory under the first order predicate logic. Since such a coding argument is usually restricted in the domain of finitistic objects, a minimal theory may be $ZF^- - P - INF$, ZF with the axiom of foundation, the power set, and the infinity are deleted.

[8] That is, every theorem in $\mathscr{B}$ is a theorem in $\mathscr{L}_i$.

[9] By "based on descriptions of construction rules," I mean that the set of formulas in L_i may supposively be closed under such formation rules that are well defined in set theory $\mathscr{B}$. That is, if θ is a formula in L_i, then $\neg\theta$ is also a formula in L_i, if $P(x)$ is a formula in one free variable x in L_i and if t is a term in L_i, then $P(t)$ is also a formula in L_i, and so forth.

[10] For finitistic objects, notation $\ulcorner\ \urcorner$ is called Quine's corner convention.

(A.3) There is a formula, $P_i(x)$, in one free variable, x, in the theory of i to mean that $x = \ulcorner\theta\urcorner$ for a certain formula θ of i and θ is *rationally acceptable* for i. The meaning of $P_i(x)$, as a way to decide such acceptable sentences, is given as a set theoretic property in theory $\mathscr{B}$, (hence, we may not require it to be finitistic), so that $P_i(x)$ may also be identified with a formula in $\mathscr{B}$.

Under (A.2), one of the most typical set theoretic procedure in $\mathscr{B}$ satisfying conditions in (A.3) for $P_i(x)$ (the rational acceptability) may be the proof procedure in $\mathscr{L}_i$, though we do not confine ourselves to this most familiar case. In ordinary settings in economics, such a P_i may be considered as an arbitrary formula allowing, at least, one assertion specifying a certain character of $a_i \in A_i$ as a possible final decision of an agent i, as rationally acceptable. For example, such assertions may be: "final decision $a_i \in A_i$ of i is a price taking and utility maximizing behavior," for an ordinary micro economics settings, "final decision $a_i \in A_i$ of i is a best response given other agents' behaviors," for Nash equilibrium settings, and so on. It follows that, an agent $i \in I$ chooses an action $a_i \in A_i$ only if there is a sentence of i, θ, which is rationally acceptable, $(P_i(\ulcorner\theta\urcorner))$, asserting that agent i is allowed to chose action a_i as his/her final decision.

2.2 Rationality

As stated in the introduction, we are considering that an economic model should incorporate a structure which makes each agent's behavior to be called *rational*. In the previous section, such a structure is represented by the formula, $P_i(x)$, for agent i under the world view, $\mathscr{L}_i = (L_i, R_i, T_i)$, of i. We shall make in this section a further specification on the property $P_i(x)$, the *rationality* of i.

Perhaps, the most important property for P_i to be called as the rationality of i will be the consistency. It seems, however, that there are two kind of such consistency. One is the logical consistency and the other is the semantical consistency. We say that $P_i(x)$ is *logically consistent* if for any sentence θ of i, $P_i(\ulcorner\theta\urcorner)$ and $P_i(\ulcorner\neg\theta\urcorner)$ do not hold simultaneously. The logical consistency of $P_i(x)$ as a fact in the underlying theory of sets, $\mathscr{B}$, is denoted by $CONS(P_i)$. Formally;

(D.1) $CONS(P_i)$ is a formula in $\mathscr{B}$ which is equivalent to saying that $Form_i(\ulcorner\theta\urcorner) \to (P_i(\ulcorner\theta\urcorner) \to \neg P_i(\ulcorner\neg\theta\urcorner))$.[11]

The *semantical consistency* of P_i is the requirement that for any sentence θ of i, $P_i(\ulcorner\theta\urcorner)$ and $\neg P_i(\ulcorner P_i(\ulcorner\theta\urcorner)\urcorner)$ do not hold simultaneously. Since the condition (ordinarily) means that for each sentence θ of i, $P_i(\ulcorner\theta\urcorner) \to P_i(\ulcorner P_i(\ulcorner\theta\urcorner)\urcorner)$, we also call it the *introspective completeness* and denote it (as a fact in the underlying theory of sets) by $COMP(P_i)$. Formally;

(D.2) $COMP(P_i)$ is a formula in $\mathscr{B}$ which is equivalent to saying that $Form_i(\ulcorner\theta\urcorner) \to (P_i(\ulcorner\theta\urcorner) \to P_i(\ulcorner P_i(\ulcorner\theta\urcorner)\urcorner))$.

The logical consistency and the introspective completeness of P_i will be argued in the next section as mostly desirable properties for P_i. The reminder of this section is devoted to define additional basic

[11] As noted in (A.2), we assume that for each formula θ in $\mathscr{L}_i$, $\neg\theta$ is also a formula in $\mathscr{L}_i$, and that the translation process between $\ulcorner\theta\urcorner$ and $\ulcorner\neg\theta\urcorner$ may be written in a formula in $\mathscr{B}$. Note also that as stated in (A.3), $P_i(x)$ is considered as a formula in $\mathscr{B}$.

properties for P_i. In the following, we assume that P_i automatically satisfies all of the following four properties.[12]

(A.4) If $\mathscr{B} \vdash \theta$, then $\mathscr{B} \vdash P_i(\ulcorner \theta \urcorner)$.

That is, each theorem in the underlying theory of sets is rationally acceptable for i.

(A.5) If $\mathscr{B} \vdash Form_i(\ulcorner \theta \urcorner) \wedge Form_i(\ulcorner \eta \urcorner) \wedge \ulcorner \theta \urcorner = \ulcorner \eta \urcorner$, then $\mathscr{B} \vdash P_i(\ulcorner \theta \leftrightarrow \eta \urcorner)$.

This implies that for each two formulas of i which are proved to be equal as set theoretical objects in $\mathscr{B}$, it is rationally acceptable to treat them as equivalent formulas.

(A.6) $\mathscr{B} \vdash Form_i(\ulcorner \theta \urcorner) \rightarrow (P_i(\ulcorner P_i(\ulcorner \theta \urcorner) \urcorner) \rightarrow P_i(\ulcorner \theta \urcorner))$.

The rational acceptability of θ under the rational acceptability of $P_i(\ulcorner \theta \urcorner)$ is quite natural.

(A.7) $\mathscr{B} \vdash (Form_i(\ulcorner \theta \urcorner) \wedge Form_i(\ulcorner \eta \urcorner)) \rightarrow (P_i(\ulcorner \theta \rightarrow \eta \urcorner) \rightarrow (P_i(\ulcorner \theta \urcorner) \rightarrow P_i(\ulcorner \eta \urcorner)))$.

If $\theta \rightarrow \eta$ and θ are rationally acceptable, then η is rationally acceptable. That is, the assumption means that rationally acceptable statements are closed under the modus ponens.

2.3 Incompleteness

In the following, the main result of this section is given in the form of three theorems. These are different aspects of the same fact (a certain kind of incompleteness of P_i) under $\mathscr{B}$ with several auxiliary assumptions. The first theorem says that with additional properties in (A.1)–(A.7), $CONS(P_i) \wedge COMP(P_i)$ is false or is not rationally acceptable.

Theorem 2.1 : Under (A.1)–(A.7),[13]

$$\mathscr{B} \vdash (CONS(P_i) \wedge COMP(P_i)) \rightarrow \neg P_i(\ulcorner CONS(P_i) \wedge COMP(P_i) \urcorner),$$

Proof : Let θ be a formula in one free variable in $\mathscr{L}_i$ and define formula $q(x)$ in one free variable x in $\mathscr{B}$ through the set theoretic process defining formula $q(\ulcorner \theta \urcorner)$ as an equivarent formula of $P_i(\ulcorner \neg \theta(\ulcorner \theta \urcorner) \urcorner)$. (Under condition (A.2), we may assure that the procedure, $\ulcorner \theta \urcorner \mapsto \ulcorner \neg \theta(\ulcorner \theta \urcorner) \urcorner$, is well defined through formulas in $\mathscr{B}$. For example, we may define $q(x)$ as "$Form^1(x) \wedge \exists y(\exists v(Sbst(v, x, x) \wedge Neg(y, v))) \wedge P_i(y)$.") Since $\mathscr{B} \vdash q(\ulcorner q \urcorner) \leftrightarrow P_i(\ulcorner \neg q(\ulcorner q \urcorner) \urcorner)$, by defining Q as $q(\ulcorner q \urcorner)$. Then,

$$\mathscr{B} \vdash Q \leftrightarrow P_i(\ulcorner \neg Q \urcorner).^{14} \tag{1}$$

[12]The following assumptions are written in the form of theorems (or metatheorems on theorems) in $\mathscr{B}$. The symbol $\vdash$ denotes that the right hand side is a theorem under the development of the theory denoted by an expression at the left hand side. Since proofs in $\mathscr{L}_i$ (hence, in $\mathscr{B}$,) may be considered as objects in the underlying theory of sets, an expression such as "$\mathscr{L}_i \vdash \theta$" may also be considered as a formula in the underlying set theory.

[13]More precisely, we are supposing that every facts in (A.1)–(A.7) may be treated as trivial theorems by definitions in the underlying theory of sets, $\mathscr{B}$.

[14]Note that by (A.1)–(A.3), q, P_i, and Q may be considered as formulas in $\mathscr{B}$ as well as $\mathscr{L}_i$ though θ may not be. Since P_i is a formula in $\mathscr{B}$, it may also possible to obtain assertion (1) as an application of, so called, Gödel's lemma (see, e.g., Kunen (1980; p.40, Theorem 14.2)). I have proved it directly merely for the sake of completeness of the paper.

Since $\mathscr{B} \vdash (COMP(P_i) \wedge P_i(\ulcorner \neg Q \urcorner)) \rightarrow P_i(\ulcorner P_i(\ulcorner \neg Q \urcorner)\urcorner)$, by assertion (1), we have

$$\mathscr{B} \vdash (COMP(P_i) \wedge P_i(\ulcorner \neg Q \urcorner)) \rightarrow P_i(\ulcorner Q \urcorner). \tag{2}$$

Therefore,

$$\mathscr{B} \vdash (CONS(P_i) \wedge COMP(P_i)) \rightarrow \neg P_i(\ulcorner \neg Q \urcorner). \tag{3}$$

Then, by (A.4) and (A.7),

$$\mathscr{B} \vdash P_i(\ulcorner CONS(P_i) \wedge COMP(P_i) \urcorner) \rightarrow P_i(\ulcorner \neg P_i(\ulcorner \neg Q \urcorner)\urcorner). \tag{4}$$

By (1), we have also that $\mathscr{B} \vdash \neg Q \leftrightarrow \neg P(\ulcorner \neg Q \urcorner)$. Then, by (A.4), (A.7) and (4),

$$\mathscr{B} \vdash P_i(\ulcorner CONS(P_i) \wedge COMP(P_i) \urcorner) \rightarrow P_i(\ulcorner \neg Q \urcorner). \tag{5}$$

Hence, by (3) and (5), we have

$$\mathscr{B} \vdash (CONS(P_i) \wedge COMP(P_i)) \rightarrow \neg P_i(\ulcorner CONS(P_i) \wedge COMP(P_i) \urcorner), \tag{6}$$

which was to be proved. ■

The next theorem consists of assertions with one more additional property, $CONS(P_i)$ or $COMP(P_i)$, to (A.1)–(A.7). The theorem shows how these two concepts are mutually introspectively inconsistent.

Theorem 2.2 : Assume that (A.1)–(A.7) hold.

(a) $\mathscr{B} \vdash \neg COMP(P_i) \vee \neg CONS(P_i) \vee \neg P_i(\ulcorner CONS(P_i) \urcorner) \vee \neg P_i(\ulcorner COMP(P_i) \urcorner)$.

(b) If $\mathscr{B} \vdash CONS(P_i)$, then $\mathscr{B} \vdash COMP(P_i) \rightarrow \neg P_i(COMP(P_i))$.

(c) If $\mathscr{B} \vdash COMP(P_i)$, then $\mathscr{B} \vdash CONS(P_i) \rightarrow \neg P_i(CONS(P_i))$.

(d) if $\mathscr{B} \vdash CONS(P_i)$, then $\mathscr{B} \vdash \neg P_i(\ulcorner P_i(COMP(P_i)) \urcorner)$.

Proof : Recall equation (1) in the proof of previous theorem. Note that

$$\mathscr{B} \vdash COMP(P_i) \rightarrow (P_i(\ulcorner \neg Q \urcorner) \rightarrow P_i(\ulcorner P_i(\ulcorner \neg Q \urcorner)\urcorner)). \tag{7}$$

Hence, by (1) in the proof of previous theorem toghether with conditions (A.4) and (A.7), we have

$$\mathscr{B} \vdash COMP(P_i) \rightarrow (P_i(\ulcorner \neg Q \urcorner) \rightarrow P_i(\ulcorner Q \urcorner)). \tag{8}$$

It follows that

$$\mathscr{B} \vdash COMP(P_i) \rightarrow (CONS(P_i) \rightarrow \neg P_i(\ulcorner \neg Q \urcorner)). \tag{9}$$

By (A.4) and (A.7), we have

$$\mathscr{B} \vdash P_i(\ulcorner COMP(P_i) \urcorner) \rightarrow (P_i(\ulcorner CONS(P_i) \urcorner) \rightarrow P_i(\ulcorner \neg P_i(\ulcorner \neg Q \urcorner)\urcorner)). \tag{10}$$

Again, by (1) in the proof of previous theorem ($\mathscr{B} \vdash \neg Q \leftrightarrow \neg P_i(\ulcorner \neg Q \urcorner)$) to gether with (A.4) and (A.7), it follows from (10) that

$$\mathscr{B} \vdash P_i(\ulcorner COMP(P_i) \urcorner) \rightarrow (P_i(\ulcorner CONS(P_i) \urcorner) \rightarrow P_i(\ulcorner \neg Q \urcorner)). \tag{11}$$

Hence, by (9) and (11), we obtain that

$$\mathscr{B} \vdash \neg(COMP(P_i) \wedge CONS(P_i) \wedge P_i(\ulcorner COMP(P_i) \urcorner) \wedge P_i(\ulcorner CONS(P_i) \urcorner)). \tag{12}$$

Hence, (a) holds. Assertion (b) and (c) follows immediately from (12) if we consider the fact that $\mathscr{B} \vdash CONS(P_i)$ and $\mathscr{B} \vdash CONS(P_i)$ mean $\mathscr{B} \vdash P_i(\ulcorner CONS(P_i) \urcorner)$ and $\mathscr{B} \vdash P_i(\ulcorner CONS(P_i) \urcorner)$, respectively, under $(A.4)$. By (b), we also have

$$\text{If } \mathscr{B} \vdash CONS(P_i), \text{ then } \mathscr{B} \vdash P_i(COMP(P_i)) \rightarrow \neg COMP(P_i). \tag{13}$$

It follows by (A.4) and (A.7) that

$$\text{If } \mathscr{B} \vdash CONS(P_i), \text{ then } \mathscr{B} \vdash P_i(\ulcorner P_i(COMP(P_i)) \urcorner) \rightarrow P_i(\ulcorner \neg COMP(P_i) \urcorner). \tag{14}$$

Under (A.6), however, it is always true that

$$\mathscr{B} \vdash P_i(\ulcorner P_i(COMP(P_i)) \urcorner) \rightarrow P_i(\ulcorner COMP(P_i) \urcorner). \tag{15}$$

Situations in (14) and (15) show a contradiction in $\mathscr{B}$ (for $CONS(P_i)$) if $P_i(\ulcorner P_i(COMP(P_i)) \urcorner)$ is true. We have, therefore, the last assertion (d). ■

The last theorem is on the inconsistency of all properties (A.1)–(A.7), $CONS(P_i)$, and $COMP(P_i)$, together with the underlying theory of sets, $\mathscr{B}$. It may also possible to understand the theorem as an undefinability theorem of the concept "rationality".

Theorem 2.3 : Under (A.1)–(A.7), it is impossible for theory $\mathscr{B}$ to prove $CONS(P_i)$ and $COMP(P_i)$, simultaneously.

PROOF : If $\mathscr{B}$ proves $CONS(P_i) \wedge COMP(P_i)$, it also proves $P_i(\ulcorner CONS(P_i) \wedge COMP(P_i) \urcorner)$ under (A.4), which contradicts to Theorem 2.1. ■

Remark 2.1 : **(Undefinability of Rationality)** If we change (A.3) so that it states the property of P_i in (A.3) without maintaining the existence of P_i, the above theorem asserts that there is no set theoretically well defined procedure in $\mathscr{B}$ (under (A.1) and (A.2)) for defining P_i, a concept of the rationality, satisfying (A.4)–(A.7), $CONS(P_i)$ and $COMP(P_i)$, i.e., we obtain an *undefinability theorem of rationality*.

Remark 2.2 : **(Tarski's Truth Definition Theorem)** The special case that $\mathscr{B} = \mathscr{L}_i = ZF$ and P_i is considered as a definition of "truth" (which clearly satisfy properties (A.4)–(A.7), $CONS(P_i)$ and $COMP(P_i)$) is Tarski's truth definition theorem (see, Kunen (1980), p.41).

Remark 2.3 : **(Undefinability of Common Knowledge)** Especially, if there are two agents, i and j, having the same rationality in the set theory, ($\ulcorner P_i \urcorner = \ulcorner P_j \urcorner$), then (D.2) is a necessary condition for their rationality to be a *common knowledge*, i.e., $P_j(\ulcorner \theta \urcorner) \rightarrow P_i(\ulcorner P_j(\ulcorner \theta \urcorner) \urcorner)$ and $P_i(\theta) \rightarrow P_j(\ulcorner P_i(\ulcorner \theta \urcorner) \urcorner)$. Hence, the above result may also be interpreted as an *undefinability theorem of rationality* as a *consistent common knowledge*.

3 SOCIETY AND VALUES

In this section, we continue to analyze formal set theoretical limitations in describing the human society. Results in the previous section was that there is no satisfactory way to formalize the human society as long as we identify it with the whole of 'rational' individuals (the *methodological individualism*). The purpose of this section is to show that the problem may not vanish even when we look for a structure which may not necessarily have such a micro foundation.

A description of the society that has no micro foundations needs other types of *verifications* for the validity of the description itself. Indeed, it is a fundamental feature of the *logical positivism* to consider the world (the society) as the whole of logical sentences that may or may not hold, and the purpose of social science, (if it may be called as a science,) is to find assertions that are true (or at least may be called as adequate) for a description of the society. If we require such verifications for the validity, however, there always exists the problem on the introspective (semantical) and logical consistency as is the case with structures for rational individuals. That is, such a social validity cannot be introspectively (semantically) consistent as long as we require it to be logically consistent.

Let us denote here by $P(x)$ the assertion in a certain formal language, $\mathscr{L}$, meaning that "the society is such that the assertion x holds." Suppose that the language, $\mathscr{L}$, may be treated as a list of objects in a certain theory of sets, $\mathscr{B}$, which is also written by formulas in language $\mathscr{L}$. We consider that $\mathscr{B}$ is a set theory under the first order predicate logic. (For the sake of simplicity, one may identify $\mathscr{B}$ with ZF, Zermelo-Fraenkel set theory, under the first order predicate logic.)[15] Hence, we may deal with each formula, θ, in $\mathscr{L}$ as a set theoretical object, $\ulcorner\theta\urcorner$, in $\mathscr{B}$. Moreover, assume that the formula, $P(x)$, in one free variable, x, is a set theoretically well defined property (i.e., we may also identify $P(x)$ as a formula in $\mathscr{B}$,) or (if $\mathscr{B}$ is a sufficiently strong theory) an structural object in $\mathscr{B}$. Then, under several natural conditions, we have the following results:

(1) There exists a mathematical truth ($\mathscr{B} \vdash \theta$) that isn't socially valid ($\mathscr{B} \vdash \neg P(\ulcorner\theta\urcorner)$). (Theorem 3.1.)

(2) Especially, we cannot verify the semantical (introspective) consistency of the description, $P(x)$, itself. (Theorem 3.2.)

(3) We cannot define (formally describe) the society as long as we require it to be logically and semantically consistent. (Theorem 3.3.)

These arguments may also be restated as follows: if we identify the description of the society with deciding what is valid in the society, then the social validity (a value judgement in the society) is always restrictive in the sense that we are not allowed to ask what the society exactly is (as long as we require it to be logically and semantically consistent). Of course, the result may also be interpreted as a general statement on various social values, i.e., we cannot completely describe social norms, justice,

[15]This is the same setting as in the preceding section, except that $\mathscr{L}$ and $\mathscr{B}$ are not private but public language and theory, respectively. As in the previous section, for mathematical concepts, see Kunen (1980), Jech (2003), and Fraenkel et al. (1973). I am convinced in that the linguistic definitions and approaches throughout this paper are so common in standard arguments in philosophical analysis that it is not appropriate to refer to merely a few of such authors. On the standpoint of our notions of rationality and truth, however, I have obtained much from recent works in analytical philosophy and cognitive science, such as Kripke (1972), Putnam (1983), and Lakoff (1987).

and/or validities as well defined structures (mechanisms) as long as we require it to be logically and semantically consistent.

These results are closely related to the arguments in the previous section in which it is the logically consistent *rationality of individuals* that makes description of the society introspectively inconsistent. In this section, it is the logically consistent *values in the society* that makes verification of the society introspectively inconsistent. It can be said that though the truth and/or rationality in our society are determined by ourselves, no single mind is allowed to control or even define them.

3.1 Society

As in Section 2, we assume that all mathematical arguments and theorems are supposed to be given in a certain formal set theory, $\mathscr{B} = (L_B, R_B, T_B)$, where L_B is the list of symbols, R_B is the list of syntactical rules, and T_B is the list of axioms. Moreover, it is also assumed that in describing the society, a language, $\mathscr{L} = (L, R, T)$, is used, where L (the list of symbols), R (the list of syntactical rules), and T (the list of axioms) are sufficient for developing the theory $\mathscr{B}$ under the first order predicate logic in the sense that every formula in $\mathscr{B}$ may be identified with a formula in $\mathscr{L}$. That is, we assume the following:

> (B.1) $\mathscr{B}$ may be identified with a set theory under the first order predicate logic. Every symbols, terms, formulas, inference rules, and logical (non-mathematical) axioms in $\mathscr{B}$ are written by formulas in $\mathscr{L}$.

Moreover, we assume that $\mathscr{L}$ is formalized under $\mathscr{B}$. Precisely:

> (B.2) $\mathscr{B}$ describes $\mathscr{L}$ in the following sense: (i) Each member of list L is a set in theory $\mathscr{B}$. (ii) List R_i consists of formulas in theory $\mathscr{B}$. Especially, there are formulas in one free variable, $Term(x)$, $Form(x)$, $Form^1(x)$, $Neg(x, y)$ and $Sbst(x, y, z)$ describing, respectively, "x is a terms of $\mathscr{L}$," "x is a formula of $\mathscr{L}$," "x is a formula in one free variable," "x is a negation of y," and "y is a formula in one free variable, and x is the formula obtained by substituting a term z into y." Every inference rule, as a relation among formulas in $\mathscr{L}$, is also written in the language of $\mathscr{B}$. (iii) $Axiom(x)$ which defines formulas of $\mathscr{L}$ belonging to list T is also a formula in $\mathscr{B}$.

Assumption (B.2) enables us to treat each assertion θ in $\mathscr{L}$ as a set theoretical object $\ulcorner\theta\urcorner$ in theory $\mathscr{B}$. Since every terms and formulas in $\mathscr{B}$ is also in $\mathscr{L}$ by (B.1), through theory $\mathscr{B}$, language $\mathscr{L}$ may be formalized in $\mathscr{L}$ itself.

In this section, we assume that the concept of the *society* is given in a logical formula, $P(\ulcorner\theta\urcorner)$, in one free variable $\ulcorner\theta\urcorner$, in $\mathscr{B}$, maintaining that "the assertion θ in $\mathscr{L}$ is valid as a description for the society." That is, we identify the problem, "what is the society" with the problem "which assertion holds in the society." Hence, if there is a complete description of the society, we may obtain all the relevant assertions on what the society is, what we are in the society, and what we should do in the society. We suppose that such a structure of the society, i.e., the meanings of P, is given in the underlying theory of sets, $\mathscr{B}$. Formally:

> (B.3) There is a formula, $P(x)$, in one free variable x in the theory of sets, $\mathscr{B}$, asserting that "$x = \ulcorner\theta\urcorner$ for a certain assertion θ in $\mathscr{L}$ which is *valid* for a description of the society."

Of course, by (B.1), every formula in $\mathscr{B}$ is also in $\mathscr{L}$, so that the formula $P(x)$ is in $\mathscr{L}$ as well as in $\mathscr{B}$.[16] The "validity" stated in the above will be discussed axiomatically in the next section. Assumption (B.3) at least maintains, however, the standpoint that we identify the world with the whole of valid logical formulas whatever the meaning of the validity is.[17] Hence, in this sense, we identify the society with the whole of values in the society.

3.2 Social Validity and Mathematical Truth

As stated in the previous section, we are considering in assumption (B.3) that to define the *society* is nothing but to decide what the valid descriptions for the society are, hence, is nothing but to decide what the *validity* in the society is. That is, we are considering that all *values* in the society are closely related to the description of the society itself.[18] Hence, the problem on P we have seen in the following of this section is nothing but a problem on the (formally and mechanically defined) values in the society.

As a mechanism which defines the validity in the society, it will be natural for us to expect P having the following properties.[19]

(C.1) (Logical Consistency: CONS) $P(\ulcorner\theta\urcorner) \rightarrow \neg P(\ulcorner\neg\theta\urcorner)$.

(C.2) (Semantical Consistency: COMP) $P(\ulcorner\theta\urcorner) \rightarrow P(\ulcorner P(\ulcorner\theta\urcorner)\urcorner)$

(C.3) There is at least one formula φ such that $\mathscr{B} \vdash \varphi$ and $\mathscr{B} \vdash P(\ulcorner\varphi\urcorner)$.

(C.4) If $\mathscr{B} \vdash \varphi \rightarrow \psi$, then $\mathscr{B} \vdash P(\ulcorner\varphi\urcorner) \rightarrow P(\ulcorner\psi\urcorner)$.

In the following, we see that if we use (C.1) (C.2) and (C.4) as defined properties of P, (i.e., (C.1) (C.2) and (C.4) are automatically proved in $\mathscr{B}$ by definition), then we have a certain mathematical truth that cannot be a socially valid statement.

Theorem 3.1 : Under (B.1) (B.2) (B.3) (C.1) (C.2) and (C.4), there is a statement, ψ, that is mathematically true ($\mathscr{B} \vdash \psi$) though it is not socially valid ($\mathscr{B} \vdash \neg P(\ulcorner\psi\urcorner)$).

Proof : Let θ be a formula in one free variable of $\mathscr{L}$ and define formula q in one free variable in $\mathscr{B}$ through the process identifying $q(\ulcorner\theta\urcorner)$ with an equivalent formula of $P(\ulcorner\neg\theta(\ulcorner\theta\urcorner)\urcorner)$ as in the proof of Theorem 2.1 (assertion (1)), and Q be the formula $q(\ulcorner q\urcorner)$. Then, as before, we have

$$\mathscr{B} \vdash Q \leftrightarrow P(\ulcorner\neg Q\urcorner), \tag{16}$$

$$\mathscr{B} \vdash \neg Q \leftrightarrow \neg P(\ulcorner\neg Q\urcorner). \tag{17}$$

[16] Indeed, as in the preceding section, such a formula is more appropriate to be regarded as a formula in $\mathscr{L}$ than $\mathscr{B}$ even if it is written in $\mathscr{B}$. It is the "meaning" of P that is given in the theory $\mathscr{B}$, so that the formula P itself is more natural to be considered as a formula in $\mathscr{L}$.

[17] Or, at least, we are considering that a complete description of the society should decide (in the sense of $\mathscr{B}$) a set of logical formulas that are valid view of the society.

[18] The results in this section, however, holds even if there is no relation between such a validity and a description of the society. In such a case, the results may be considered as criticism for such a concept of "validity," i.e., for the *logical positivism*.

[19] Note that the following assumptions are written in the form of formulas in $\mathscr{B}$ ((C.1),(C.2)) or assertions on theorems in $\mathscr{B}$ ((C.3),(C.4)). The symbol, $\vdash$, (as in (C.3), (C.4), etc.), denotes that the right hand side is a theorem under the development of the theory denoted by an expression (as theory $\mathscr{B}$ with or without some additional axioms) at the left hand side.

Since, by (C.2), $\mathscr{B} \vdash P(\ulcorner \neg Q \urcorner) \rightarrow P(\ulcorner P(\ulcorner \neg Q \urcorner) \urcorner)$, we have by (16) together with (C.4),

$$\mathscr{B} \vdash P(\ulcorner \neg Q \urcorner) \rightarrow P(\ulcorner Q \urcorner). \tag{18}$$

Therefore, by (C.1),

$$\mathscr{B} \vdash \neg P(\ulcorner \neg Q \urcorner). \tag{19}$$

By (17) together with (C.4), however, statement (19) also implies

$$\mathscr{B} \vdash \neg P(\ulcorner \neg P(\ulcorner \neg Q \urcorner) \urcorner). \tag{20}$$

Let ψ be the formula, $\neg P(\ulcorner \neg Q \urcorner)$. Then, by (19) and (20), ψ satisfies the condition of the theorem. ■

The mathematical truth which cannot be socially valid in the above theorem, may not have any serious meanings in view of social science. There seems to exist, however, an important kind of such assertions with respect to the structure of P itself. Denote condition (C.1) and condition (C.2) by $CONS$ and $COMP$ respectively. If we assume (C.3) and (C.4) together with (B.1), (B.2), and (B.3), we can see that $CONS$ and $COMP$, themselves, may be classified into such an important kind of sentences, as the next theorem asserts.

Theorem 3.2 : Assume (B.1), (B.2), (B.3), (C.3), and (C.4).

(a) $\mathscr{B} \vdash \neg COMP \vee \neg CONS \vee \neg P(\ulcorner COMP \urcorner) \vee \neg P(\ulcorner CONS \urcorner)$.

(b) If $\mathscr{B} \vdash COMP$, we have $\mathscr{B} \vdash CONS \rightarrow \neg P(\ulcorner CONS \urcorner)$.

(c) If $\mathscr{B} \vdash CONS$, we have $\mathscr{B} \vdash COMP \rightarrow \neg P(\ulcorner COMP \urcorner)$.

Proof : Let Q be the same formula defined in the proof of Theorem 3.1. Note that (16) and (17) are also true under the setting of Theorem 3.2. Since $\mathscr{B} \vdash (COMP \wedge P(\ulcorner \neg Q \urcorner)) \rightarrow P(\ulcorner P(\ulcorner \neg Q \urcorner) \urcorner)$, by equation (16) together with (C.4), we have

$$\mathscr{B} \vdash COMP \rightarrow (P(\ulcorner \neg Q \urcorner) \rightarrow P(\ulcorner Q \urcorner)). \tag{21}$$

Therefore,

$$\mathscr{B} \vdash COMP \rightarrow (CONS \rightarrow \neg P(\ulcorner \neg Q \urcorner)). \tag{22}$$

Then, by (C.4),

$$\mathscr{B} \vdash P(\ulcorner COMP \urcorner) \rightarrow (P(\ulcorner CONS \urcorner) \rightarrow P(\ulcorner \neg P(\ulcorner \neg Q \urcorner) \urcorner)). \tag{23}$$

Hence, by (17) and (C.4),

$$\mathscr{B} \vdash P(\ulcorner COMP \urcorner) \rightarrow (P(\ulcorner CONS \urcorner) \rightarrow P(\ulcorner \neg Q \urcorner)). \tag{24}$$

Under (22) and (24), if $COMP \wedge CONS \wedge P(\ulcorner COMP \urcorner) \wedge P(\ulcorner CONS \urcorner)$ is true, a contradiction follows (in theory $\mathscr{B}$), so that we have (a). Under (C.3) and (C.4), by considering φ in (C.4) as the formula whose existence is assured in (C.3), we have for any formula ψ in $\mathscr{B}$, if $\mathscr{B} \vdash \psi$, then $\mathscr{B} \vdash P(\psi)$. Therefore, $\mathscr{B} \vdash COMP$ implies $\mathscr{B} \vdash P(\ulcorner COMP \urcorner)$ and $\mathscr{B} \vdash CONS$ implies $P(\ulcorner CONS \urcorner)$, so that assertions (b) and (c) immediately follow form (a). ■

Lastly, we see the inconsistency of all properties (B.1)–(B.3) and (C.1)–(C.4) together with the underlying theory of sets, $\mathscr{B}$. It may also possible to understand the theorem as an undefinability theorem of the concept "social validity".

Theorem 3.3 : Under (B.1),(B.2),(B.3),(C.3) and (C.4), it is impossible for theory $\mathscr{B}$ to prove $CONS$ and $COMP$, simultaneously.

Proof : Assume that $\mathscr{B}$ proves $COMP$ and $CONS$. Then, by (C.3) and (C.4), $\mathscr{B}$ also proves $P(\ulcorner COMP \urcorner)$ and $P(\ulcorner CONS \urcorner)$, which contradicts to (a) in the previous theorem. ■

Remark 3.1 : **(Undefinability of Social Validity)** If we change (B.3) so that it asserts merely the property of P without maintaining the existence, the above theorem maintains that there is no set theoretical possibility (in $\mathscr{B}$ under (B.1) and (B.2)) for defining a concept of the social validity, P, satisfying (C.1)–(C.4).

4 RATIONALITY AS A FIXED POINT FOR VIEWS OF THE WORLD

In this section, by incorporating the arguments in Sections 2 and 3 of the previous chapter, rationality and social validity are figured out as an equilibrium of social model in which cognitive features of members are treated explicitly. Intuitively, the model in this section describes the situation in which each member is possible to chose an arbitrary finite number of models of the society, the *possible worlds*, to approximate the 'real' world. Even though candidates for such models of the society for each member may not be finite and members are not convinced in his/her approximation to be complete, we can expect the existence of the list of each person's view of the world and 'rational' behaviors based on them, which are also compatible with each person's view of the world in the light of their experiences and beliefs for the validity of the model, as long as the total space for behaviors is not so large, e.g., a compact Hausdorff space.

Results in the previous two sections tell us how it is difficult to describe 'individual rationality' and/or 'social validity' as artifical objects (i.e., conditions for them are determined conventionally by ourselves). In this section, by considering the space of action configurations of all members, $X = \prod_{i \in I} X_i$, as the set of *rigid designators* that are identified across possible worlds, we characterize a list of *rational behaviors* ("justified" under a view of the world) and a possible world ("compatible" with such behaviors) as an equilibrium situation for recognition of the society.

By arguments in Section 2, such a "justification" procedure for rational actions (under a view of the world like an economic model) may be consistent though it may not be so complete to justify what is the "possibility" for views of the world. Judgements for the "compatibility" (to determine what are appropriate to be called as possible worlds), therefore, should be another social validity based on the same view of the world recognizing it as valid. Section 3 tells, however, that we cannot expect such a "validity" to be so strong as maintaining the validity of itself as long as we require it to be consistent. It follows that any attempt to determine "possibility of the world" rigorously through sets of necessary and sufficient conditions seems to be incorrect. We have to leave the extension of such a concept somewhat open to the situation of self reliance of our minds, say, to our willingness, to seek another possibility of the world.

To describe and assure the existence of equilibria, we use a certain kind of mixed strategy (together with expected utility) settings for the sake of simplicity and familiarity of arguments. A *view of the world* in this paper, therefore, is taken as a mixture of possible worlds (e.g., a probability measure on several possible worlds), and the justification for behaviors and the judgement for compatibility, firstly associated merely with each possible world, are supposed to be extended on over all such mixtures, views of the world.[20] The essential feature of this section's approach, however, does not depend on such a special framework. The central ideas discussed here is to characterize human's "rationality," at least in the sense of "rational behaviors" for game theoretic settings in social science, and "validity" for a society, not as terms or objects fixed by a set of criteria laid down in advance, but as *references* to determine the extensions of the terms that refer to them by using classes of laws the whole of which we do not exactly know.[21]

4.1 Individual and Society

Let $I = \{1, 2, \ldots, m\}$ be the index set of members of society. For each $i \in I$, denote by X_i the set of possible *behaviors* for individual i. For the sake of simplicity, we assume that each X_i is a non-empty compact convex subset of a Hausdorff topological vector space. We also assume that X_i covers all the behaviors that are observable to others for each $i \in I$ and that each *behavior profile* $(x_1, x_2, \ldots, x_m) \in \prod_{i \in I} X_i$ is sufficient to decide a *consequence*, $c_i \in C_i$ for each $i \in I$.

4.2 Languages and Possible Worlds

Each member, i, is assumed to have a set of *logical formulas*, T_i, *inference rules*, R_i, and a *language*, $\mathscr{L}_i$, that may be considered as the list of symbols describing the formulas. Triplet $(\mathscr{L}_i, R_i, T_i)$ is called a *possible world* of i. Member i may have a lot of (we suppose possibly denumerably many) possible worlds, $W_i^0 = (\mathscr{L}_i^0, R_i^0, T_i^0)$, $W_i^1 = (\mathscr{L}_i^1, R_i^1, T_i^1)$, $W_i^2 = (\mathscr{L}_i^2, R_i^2, T_i^2)$, ... We assume they may not mutually be consistent nor may not even be translatable one another.

Let us consider the inductive limit of abstract simplices, $W_i = \varinjlim_n \overline{W_i^0 \cdots W_i^n}$, where each $\overline{W_i^0 \cdots W_i^n}$ is identified with n-dimensional standard unit simplex Δ^n. A point, $w \in W_i$, may be considered as representing a special standpoint of i's thought. We call it as i's *view of the world*.

Each i cannot use W_i as a formal object of his/her theory to understand the world. We do not except, however, the case that W_i^m is treated in a certain W_i^n as a formal object.

4.3 Possible Worlds and Behaviors

For each $i \in I$, possible worlds of i, $W_i^0, W_i^1, W_i^2, \ldots$, represent various kinds of reasoning for a certain behavior, $x_i \in X_i$ to be considered better than other others. For example, W_i^n may be a possible world

[20] One may ask why such a view of the world (in the above sense of the mixture) itself is not classified into one of the possible worlds. Of course, we may call it as a possible world as long as it is prepared as a candidate for his/her possible worlds from the beginning.

[21] The idea may be restated that those terms are treated as*natural kind words* in the sense of Kripke (1972) and Putnam (see, e.g., Putnam (1983), 4.Reference and Truth). Their approaches (independently proceeded in 1960's and 1970's) are called the *new theory of refenrence*; and at least in this section, the concept of "possible world" is used in relation to this context.

of a Nash Equilibrium, i.e., under W_i^n, i is convinced in that his/her choice of behavior x_i is reasonable since it is a part of a Nash equilibrium strategy profile for a certain game theoretic model of society which is completely described and treated as valid in $(\mathscr{L}_i^n, R_i^n, T_i^n)$. (In such cases, i's estimation on thoughts of other persons are also described completely and treated as valid under axioms for individuals in $(\mathscr{L}_i^n, R_i^n, T_i^n)$.) It is also possible to consider W_i^n as a world of cooperation equilibrium, (i.e., i thinks that his/her behavior may be considered as a reasonable action since everyone's behavior may be classified as choices to decide a consequence in a core of the game defined in W_i^n), a world of an incomplete information game, a world of an abstract economy (in which the constraint correspondence is described as a rule in T_i^n), and so on. (It should be remarked that we are assuming that all behaviors which is possible for i is completely listed in X_i. Hence, each X_i is so defined as to include a mixed strategy if such behaviors are allowed to exist in the formalized model.)

4.4 Justification and Refutability

Each possible world W_i^n of $i \in I$ defines for a given profile of behaviors, $x = (x_j)_{j\in I} \in \prod_{j\in I} X_j$, the set of *justified* behaviors, $\Phi_i(W_j^n, x) \subset X_i$, as behaviors that are better than x_i under W_i^n, and the set of *incompatible* profiles of behaviors, $\Theta_i^X(W_i^n) \subset \prod_{j\in I} X_j$, under W_i^n. When (an observable fact) x is incompatible with W_i^n, we say that W_i^n is *refutable* for i under x. In the following, we shall treat *rationality* of i based on a justification for behavior x_i of profile $x = (x_j)_{j\in I} \in \prod_{j\in I} X_j$ under a certain possible world that is not refutable under x. (Note that i may treat sets $\Phi_i(W_i^n, x)$ and $\Theta_i^X(W_i^n)$ in his/her formal theory, W_i^n, though he/she may not recognize them as relations on or into W_i.)

As stated before, we consider that i's view of the world is a point, $w_i \in W_i = \varinjlim_n \overline{W_i^0 \cdots W_i^n}$. Person i has no formalized theory on the rightness among possible worlds in W_i. (This is not saying that we prevent i from having formal treatments among finitely many possible worlds, $W_i^0, \ldots, W_i^n$, in a certain W_i^m.) Hence, we may interpret a point, $w_i \in \overline{W_i^0 \cdots W_i^n} \subset W_i$, as a representation of the state of i's thought in the form of a degree of confidence among possible worlds, $W_i^0 \cdots W_i^n$. In this sense, we suppose that Φ_i and Θ_i^X may adequately be extended as correspondences on W_i. Of course, a certain $w_i = (w_i^0, \ldots, w_i^{\ell(i)}) \in W_i$, may justify or refute behaviors in various ways based on $W_i^0, \ldots, W_i^{\ell(i)}$. We may assume, however, that it would be natural for Φ_i and Θ_i^X to satisfy following conditions.

($A.1$) $x \in \Phi_i(v_i, z)$ and $y \in \Phi_i(w_i, z)$ implies that for all $\lambda \in [0,1]$, there exists $\hat{\lambda} \in [0,1]$, such that $\lambda x + (1-\lambda)y \in \Phi_i(\hat{\lambda} v_i + (1-\hat{\lambda})w_i, z)$.

($A.2$) $x \notin \Theta_i^X(v_i)$ and $x \notin \Theta_i^X(w_i)$ implies that for all $\lambda \in [0,1]$, $x \notin \Theta_i^X(\lambda v_i + (1-\lambda)w_i)$.

Two conditions would be quite natural. Condition ($A.1$) says that a mixture of better strategies will be supported by a certain mixture of the two possible worlds. It should be noted here that $\hat{\lambda}$ may be different from λ. Condition ($A.2$) asserts that if x does not refute two possible worlds, then it does not refute standpoints of their mixture.

By considering the fact that there is no complete description of the world (results in earlier sections), it would be appropriate to treat such a standpoint, a view of the world, as a candidate for, say, a *wide sense of rationality*. The next section is devoted to show the existence of such rationality as a fixed point of social recognition.

4.5 Equilibrium under Social Recognition

We are assuming that there are possible worlds, $W_i^0, W_i^1, W_i^2, \ldots$, for each $i \in I$, which means that person i does not have any formalized ideas on the relation among 'all' of such possible worlds. In other words, $W_i^0, W_i^1, W_i^2, \ldots$, are all of the formalized ideas of i with respect to the society. At the same time, we have supposed that for each $i \in I$, there are relations Φ_i and Θ_i^X on $W_i = \varinjlim_n \overline{W_i^0 \cdots W_i^n}$ and X into W_i, based on justifications and refutabilities defined in each of W_i^n's. As stated before, they are not written in the theories of i.

Let $X = \prod_{i \in I} X_i$ and $W = \prod_{i \in I} W_i$. Given behavior profile $x = (x_1, \ldots, x_m) \in X = \prod_{i \in I} X_i$ and view of the world $w_i \in W_i$ of i, we denote by $\varphi_i(w_i, x) \subset W \times X$, the set of pairs, (v_i, y_i) such that $x \notin \Theta_i^X(v_i)$ and $y_i \in \Phi_i(v_i, x)$, i.e., the set of pairs of non refutable view v_i of the world under x and justified behavior y_i as better than x_i under v_i. Pair $(w, x) = (w_1, \ldots, w_m, x_1, \ldots, x_m) \in W \times X$ is said to be an *equilibrium (under social recognition)* for society $((W_i, X_i, \Phi_i, \Theta_i^X)_{i \in I})$ if $\varphi_i(w_i, x) = \emptyset$ for all $i \in I$. Adding to it, if $x \notin \Theta_i^X(W_i)$ for all $i \in I$, the equilibrium, (w, x), is called *rational*.

Theorem 4.1 : *Society $((W_i, X_i, \Phi_i, \Theta_i^X)_{i \in I})$ has an equilibrium if* $(A.1)$, $(A.2)$, *and the following condition for each Φ_i is satisfied .*

$(A.3)$ *For each $(v_i, y_i) \in W_i \times X_i$, $\Phi_i^{-1}(v_i, y_i) \subset W \times X$ is open.*

The equilibrium is rational if for each $i \in I$ and $x \in X$, there is at least one $w_i \in W_i$ such that $x \notin \Theta_i^X(w_i)$.

Proof : Assume the contrary. Then, for each $(w, x) = (w_1, \ldots, w_m, x_1, \ldots, x_m) \in W \times X$, there exists at least one $i \in I$ such that $\varphi_i(w_i, x) \neq \emptyset$, i.e., there are $i \in I$ and a pair, $(v_i^x, y_i^x) \in W_i \times X_i$, such that $x \notin \Theta_i^X(v_i^x)$ and $y_i^x \in \Phi_i(v_i^x, x)$. Under $(A.3)$, there is an open neighborhood, $V(w) \times U(x)$, of (w, x) such that $(v_i^x, y_i^x) \in \Phi_i(v_i, z)$ for all $(v, z) \in V(w) \times U(x)$. Note that the definition of φ_i enable us to chose $V(w)$ as $v(w) = W$ for all w. Then, since X is compact, we may assume that there are finite points, $x^1, \ldots, x^n \in X$, and their open neighborhoods, $U(x^1), \ldots, U(x^n)$, with indices of persons, $i(1), \ldots, i(n) \in I$, and their thoughts and behaviors, $(v_{i(1)}, y_{i(1)}) \in W_{i(1)} \times X_{i(1)}, \ldots, (v_{i(n)}, y_{i(n)}) \in W_{i(n)} \times X_{i(n)}$, satisfying for each $t = 1, \ldots, n$, that $(v_{i(t)}, y_{i(t)}) \in \Phi_i(v_i, z)$ for all $z \in U(x^t)$. Let $\alpha_1 : X \to [0, 1], \ldots, \alpha_n : X \to [0, 1]$, be the partition of unity subordinated to $U(x^1), \ldots, U(x^n)$. Denote by $N(i) \subset \{1, 2, \ldots, n\}$ the subset of indices for neighborhoods associated with i, i.e., $N(i) = \{t \,|\, i(t) = i\}$ for each $i \in I$. Moreover, denote by U_i the set $\bigcup_{t \in N(i)} U(x^t)$. Of course, $\{U_i | i \in I\}$ covers X. On each $U_i \subset X$, $i \in I$, define mapping $\psi_i : U_i \to X_i$ as $\psi_i(x) = \sum_{t \in N(i)} \alpha_t y_{i(t)}$ and correspondence $\Psi_i : U_i \to X$ as $\Psi_i(x) = \{\psi_i(x)\} \times \prod_{j \in I, j \neq i} X_j$. Under $(A.1)$ and $(A.2)$, Ψ_i is non-empty convex valued correspondence satisfying $x_i \notin \Psi_i(x)$ for each $x = (x_1, \ldots, x_m) \in U_i$. It also has the open lower section property under $(A.3)$. Define correspondence $\Psi : X \to X$ as

$$\Psi(x) = \bigcap_{i \in \{j \in I | x \in U_j\}} \Psi_i(x).$$

Then, Ψ is a non-empty convex valued correspondence having no fixed point, which is impossible since X is non-empty compact convex set. (Non-empty valued convex correspondence into itself having lower intersection property has a fixed point.) Hence, there is an equilibrium point $x^* \in X$. The last assertion is clear since the fact $\varphi(w_i, x^*) = \emptyset$ for all i does not depend on $w = (w_1, \ldots, w_m)$. ■

REFERENCES

Fraenkel, A. A., Bar-Hillel, Y., and Levy, A. (1973): *Foundations of Set Theory* Second edn. Elsevier, Amsterdam.

Jech, T. (2003): *Set Theory* Third edn. Springer-Varlag, Berlin.

Kripke, S. A. (1972): *Naming and Necessity*. Hauvard University Press, Cambridge, Massachusetts.

Kunen, K. (1980): *Set Theory: An Introduction to Independence Proofs*. North Holland, Amsterdam.

Lakoff, G. (1987): *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. The University of Chicago Press, Chicago and London.

Putnam, H. (1983): *Realism and Reason* vol. 3 of *Philosophical Papers*. Campridge University Press, New York.

Putnam, H. (2002): *The Collapse of The Fact/Value Dichotomy and Other Essays*. Harvard University Press, Cambridge, Massachusetts.

Quine, W. V. O. (1953): *From a Logical Point of View: 9 Logico-Philosophical Essays, Second Edition, Revised 1961*. Harverd University Press.

Urai, K. (2002a): "Why there isn't a complete description of the human society: The rationality and individuals," Discussion Paper No. 02-04, Faculty of Economics and Osaka School of International Public Policy, Osaka University.

Urai, K. (2002b): "Why there isn't a complete description of the human society II: The society and value," Discussion Paper No. 02-05, Faculty of Economics and Osaka School of International Public Policy, Osaka University.

Urai, K. (2002c): "Why there isn't a complete description of the human society I: The individual and rationality," Kokyuroku No. 1264, Research Institute for Mathematical Sciences, Kyoto University.