Why the society and social values cannot be completely described

（社会および社会的価値の完全な記述はなぜ存在し得ないか）

要旨 **(2002.9.14)**：浦井 憲（大阪大学大学院経済学研究科）

先の論文（合理性と個人）において扱われたのは、「合理的な主体」とそういった主体の構成する「社会」についての記述、すなわち「ミクロ的基礎を持った社会の記述」が、集合論的観点から持たなければならない限界であった。本稿において考察の対象となるのは、必ずしもミクロ的基礎を持たない「社会」の記述である。この場合、重要な問題となってくるのが、そのような「社会の記述」の「正当性」という問題である。記述が公理的、ミクロ的な構成として行われないのであれば、記述の正当性はそれに代わる何らかの「価値概念の特定化」の下で与えられなければならない。

前論文と同様に、ここでの結論はその「正当性」という概念に対して通常我々の持つ非常に自然な要請と、その概念の完成した記述ということとの間にある「数学的に必然な（集合論的）制約」として得られる。前論文同様の言語的な設定（１階の述語論理下で語られたＺＦのように、「自らを語る言語を自らの対象物として取り扱うことができる」ような集合論 $\mathscr{B} = (L_B, R_B, T_B)$ と、その集合論を語る言語 $\mathscr{L} = (L, R, T)$）を考えよう。$\mathscr{B}$ の Term や Formula はすべてこの言語 $\mathscr{L}$ の Term および Formula であり、数学的公理を除く公理もまた $\mathscr{L}$ のものであるとする。同時に言語 $\mathscr{L}$ の Term および Formula はすべて $\mathscr{B}$ における set とみなせるものであり、「$\mathscr{L}$ の term である」「$\mathscr{L}$ の formula である」「$\mathscr{L}$ の 1 変項論理式である」「$\mathscr{L}$ の論理式 x は論理式 y の否定である」「$\mathscr{L}$ の論理式 z は 1 変項論理式 y に項 x を代入したものである」をそれぞれ表す $\mathscr{B}$ の論理式が存在するものと仮定する。つまり、この意味において $\mathscr{B}$ は $\mathscr{L}$ を記述するようなものであり、また $\mathscr{B}$ を語ることのできる言語 $\mathscr{L}$ は、$\mathscr{B}$ を語ることを通していわば自らを客体視することができる。この言語 $\mathscr{L}$ とは即ち、「社会」を語る「我々」の視点そのものであり、「社会」とはこのような「言語を持った主体による自己およびその類型の客体視である」という側面を認めていることが、第一論文と共通の重要な視点である。

以上を前提にすると、第一論文とほぼ同一線上に本稿で論じられる問題を位置づけることができる。前論文で見たものは個人がその内面において社会を理解する段階における個人の合理性の問題であった。ここではそういった個人は存在せず、社会はあらかじめ全体としてとらえられるのであるが、それでも問題が消えて無くなるわけではない。問題はそういった社会像の評価、検証ということに現れる。言語 $\mathscr{L}$ の下で（必ずしもミクロ的な基礎を持たない形で）「社会」が語られているとは、$\mathscr{L}$ の陳述 θ の中で「社会についての叙述として正当」なものとは何かが定められていることであると考えよう。つまり、社会についての叙述のうちで正当なものとそうでないものを見分ける集合論的にきちんと定義された手続き（すなわち $\mathscr{B}$ の一変項論理式）P が存在していることであるとし、$P(\ulcorner\theta\urcorner)$ が「θ は社会についての叙述として正当」ということを表しているものとする。「社会」を語るということは、ここではその「社会における（叙述の）正当性」を語るということと同一視されており、そのとき我々がこの「正当性」に対してどのような要請をおくことが可能であるかということが、前論文と同様に考察の対象となる。本稿の結論は、この「正当性」についての論理的整合性（$P(\ulcorner\theta\urcorner)$ と $P(\ulcorner\neg\theta\urcorner)$ が同時に成り立たないこと）が、その意味論的整合性（$P(\ulcorner\theta\urcorner)$ ならば $P(\ulcorner P(\ulcorner\theta\urcorner)\urcorner)$ であること）と、一般に両立しないということである。

この議論を第一論文と同様にまとめるならば、(a) 自己の客体視を認めないほどに（学問的意味で）我々の言語を制限するか (b) 論理的整合性の成り立たないような正当性概念に甘んずるか (c) 意味論的（内観的）に否定されるような正当性概念に甘んずるか、いずれかの選択が我々に必要となる。社会における「正当性」という概念上、(b) および (c) を支持することは通常不可能であり、それらを拒否するならば立場は (a) しかない。言うなれば上記集合論的な結論からの当然の要請として、「社会の記述の完成」と引き替えに、我々は「自らの正当性への問い」を必ず放棄せねばならない、ということになるのである。

Why the society and social values cannot be completely described as well defined mathematical objects*

by

Ken URAI†

January 2002
(Revised September 2002)

Abstract

In this paper, we treat problems on formal set theoretical limitations in describing the human society. We see that a description of the society, at least as a set of sentences that are valid for descriptions of the society in a certain formal language, cannot be semantically (introspectively) consistent as long as we require it to be logically consistent. In other words, we cannot assure the rightness on the validity for descriptions of the society itself except for believing it. The result may also be considered as a rigorous mathematical investigation for a problematic feature in the logical positivism.

Keywords: Social values, Recognition, Mathematical logic, Theory of sets, Tarski's truth definition theorem, Gödel's incompleteness theorem, Logical positivism.

JEL classification: A10; A13; B40; C60; C70; E00

* Research was supported by the Japanese Ministry of Education Grant 12730013. The earliest version of this paper was presented in the mathematical economics seminar at Keio University (February 4, 2002). The author thanks to the participants at the seminar, my colleagues, and all of my laboratory members.

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

1. INTRODUCTION

In this paper, and the preceding paper of the author, Urai (2002), we see a formal set theoretical limitation in describing the human society as a rigorous mathematical object. The result in the previous paper 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 purpose of this paper 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 the language, $\mathscr{L}$.[1] 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 always exists a formula, θ, that is set theoretically valid ($\mathscr{B} \vdash \theta$) but is socially invalid ($\mathscr{B} \vdash \neg P(\ulcorner\theta\urcorner)$). (Theorem 1.)
(2) Especially, we cannot verify the semantical (introspective) consistency of the description, $P(x)$, itself. (Theorem 2.)
(3) We cannot define (formally describe) the society as long as we require it to be logically and semantically consistent. (Theorem 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, 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 Urai (2002) in which it is the logically consistent *rationality of individuals* that makes description of the society introspectively inconsistent. In this paper, 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.

[1] This is the same setting as in the preceding paper, except that $\mathscr{L}$ and $\mathscr{B}$ are not private but public language and theory, respectively. For mathematical concepts in this paper, see Kunen (1980), Jech (1997), and Fraenkel, Bar-Hillel, and Levy (1973). I am convinced in that the linguistic definitions and approaches throughout this paper are so common in classical arguments in the 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 the work of H. Putnam (1983) and G. Lakoff (1987).

2. THE SOCIETY

Throughout this paper, 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 $\mathscr{B}$. More precisely, we assume the following:

(B.1) Every symbols, terms, formulas, inference rules, and logical (non-mathematical) axioms in $\mathscr{B}$ are also in $\mathscr{L}$.

Moreover, we assume that $\mathscr{L}$ is formalized under $\mathscr{B}$. More 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 paper, 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 $\mathscr{B}$.[2] 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.[3] Hence, in this sense, we identify the society with the whole of values in the society.

[2] Indeed, as in the preceding paper of the author, Urai (2002), 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 $\mathscr{L}$.

[3] 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.

3. THE 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.[4] Hence, the problem on P we have seen in the following of this paper 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.[5]

(C.1) (Logical Consistency) $Form(\ulcorner\theta\urcorner) \to (P(\ulcorner\theta\urcorner) \to \neg P(\ulcorner\neg\theta\urcorner))$.
(C.2) (Semantical Consistency) $Form(\ulcorner\theta\urcorner) \to (P(\ulcorner\theta\urcorner) \to P(\ulcorner P(\ulcorner\theta\urcorner)\urcorner))$
(C.3) If $\mathscr{B} \vdash (y = z)$, then $\mathscr{B} \vdash P(y) \leftrightarrow P(z)$.
(C.4) If $\mathscr{B} \vdash \varphi \to \psi$, then $\mathscr{B} \vdash P(\ulcorner\varphi\urcorner) \to P(\ulcorner\psi\urcorner)$.
(C.5) $\mathscr{B} \vdash Form(\ulcorner\theta\urcorner) \to (P(\ulcorner P(\ulcorner\theta\urcorner)\urcorner) \to P(\ulcorner\theta\urcorner))$.

In the following, we see that if we use (C.1)–(C.3) as defined characters on P, we may obtain a mathematical (set theoretical) truth that cannot be valid in the society.

THEOREM **1.** Under (B.1)–(B.3) and (C.1)–(C.3), there is a sentence ψ such that $\mathscr{B} \vdash \psi$ and $\mathscr{B} \vdash \neg P(\psi)$.

PROOF. Let θ be a formula in one free variable in $\mathscr{L}$, $q(\ulcorner\theta\urcorner)$ be the formula $P(\ulcorner\neg\theta(\ulcorner\theta\urcorner)\urcorner)$, and Q be the formula $q(\ulcorner q\urcorner)$. Then,

$$\mathscr{B} \vdash \ulcorner Q\urcorner = \ulcorner P(\ulcorner\neg Q\urcorner)\urcorner, \tag{1}$$

$$\mathscr{B} \vdash \ulcorner\neg Q\urcorner = \ulcorner\neg P(\ulcorner\neg Q\urcorner)\urcorner. \tag{2}$$

Since, by (C.2), $\mathscr{B} \vdash P(\ulcorner\neg Q\urcorner) \to P(\ulcorner P(\ulcorner\neg Q\urcorner)\urcorner)$, we have by equation (1) together with (C.3),

$$\mathscr{B} \vdash P(\ulcorner\neg Q\urcorner) \to P(\ulcorner Q\urcorner). \tag{3}$$

Therefore, by (C.1),

$$\mathscr{B} \vdash \neg P(\ulcorner\neg Q\urcorner). \tag{4}$$

By substituting (2) to (4), we have

$$\mathscr{B} \vdash \neg P(\ulcorner\neg P(\ulcorner\neg Q\urcorner)\urcorner). \tag{5}$$

Let ψ be the formula $\neg P(\ulcorner\neg Q\urcorner)$. Then, by (4) and (5), ψ satisfies all the necessary conditions. ■

The mathematical truth which cannot be socially valid in the above theorem may be a statement which does 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. As in the preceding paper (Urai (2002)), denote condition (C.1) and (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 see that such assertions as CONS and COMP may not be socially valid.

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

[5] Note that the following assumptions are written in the form of theorems ((C.1),(C.2),(C.5)) or metatheorems on theorems ((C.3),(C.4)) in $\mathscr{B}$. The symbol $\vdash$ in (C.3)–(C.5) denotes that the right hand side is a theorem under the development of the theory denoted by an expression at the left hand side.

THEOREM **2.** Suppose that (B.1)–(B.3), (C.3), and (C.4) hold.

(a) Assume $COMP$, then $\mathscr{B} \vdash CONS \to \neg P(\ulcorner CONS \urcorner)$.
(b) Assume $CONS$, then $\mathscr{B} \vdash COMP \to \neg P(\ulcorner COMP \urcorner)$.
(c) Assume $CONS$, then $\mathscr{B} \vdash \neg COMP$.
(d) Assume $COMP$ and (C.5), then $\mathscr{B} \vdash \neg CONS$.

PROOF. Let Q be the same formula as is defined in the proof of Theorem 1. Note that (1) and (2) are also true under the setting of Theorem 2. Since $\mathscr{B} \vdash (COMP \wedge P(\ulcorner \neg Q \urcorner)) \to P(\ulcorner P(\ulcorner \neg Q \urcorner) \urcorner)$, by equation (1) together with (C.3), we have

$$\mathscr{B} \vdash (COMP \wedge P(\ulcorner \neg Q \urcorner)) \to P(\ulcorner Q \urcorner). \tag{6}$$

Therefore,

$$\mathscr{B} \vdash (CONS \wedge COMP) \to \neg P(\ulcorner \neg Q \urcorner). \tag{7}$$

Then, by (C.4),

$$\mathscr{B} \vdash P(\ulcorner CONS \wedge COMP \urcorner) \to P(\ulcorner \neg P(\ulcorner \neg Q \urcorner) \urcorner). \tag{8}$$

By substituting (2) to (8),

$$\mathscr{B} \vdash P(\ulcorner CONS \wedge COMP \urcorner) \to P(\ulcorner \neg Q \urcorner). \tag{9}$$

Hence, by (7) and (9), we can see that $CONS \wedge COMP$ in (7) and $P(\ulcorner CONS \wedge COMP \urcorner)$ in (9) cannot hold simultaneously. We obtain assertion (a) and (b), respectively, by deleting conditions $COMP$ and $CONS$ in the above argument. By (b) and (C.4), we have

$$\mathscr{B} \vdash P(\ulcorner COMP \urcorner) \to P(\ulcorner \neg P(COMP) \urcorner). \tag{10}$$

Moreover, by $CONS$, we also have

$$\mathscr{B} \vdash P(\ulcorner \neg P(COMP) \urcorner) \to \neg P(\ulcorner P(\ulcorner COMP \urcorner) \urcorner). \tag{11}$$

Hence, by (10) and (11),

$$\mathscr{B} \vdash P(\ulcorner COMP \urcorner) \to \neg P(\ulcorner P(\ulcorner COMP \urcorner) \urcorner). \tag{12}$$

Hence, assertion (c) holds. By applying (C.4) twice on (a),

$$\mathscr{B} \vdash P(\ulcorner P(\ulcorner CONS \urcorner) \urcorner) \to P(\ulcorner P(\ulcorner \neg P(\ulcorner CONS \urcorner) \urcorner) \urcorner). \tag{13}$$

Moreover, by (C.5),

$$P(\ulcorner P(\ulcorner \neg P(\ulcorner CONS \urcorner) \urcorner) \urcorner) \to P(\ulcorner \neg P(\ulcorner CONS \urcorner) \urcorner). \tag{14}$$

By (13) and (14), we have assertion (d). ■

Lastly, we see the inconsistency of all properties (B.1)–(B.3) and (C.1)–(C.5) together with the underlying theory of sets, $\mathscr{B}$ as in the preceding paper. It may also possible to understand the theorem as an undefinability theorem of the concept "social validity".

THEOREM **3.** Under (B.1)–(B.3) and (C.1)–(C.5), the theory $\mathscr{B}$ is contradictory.

PROOF. In this case, $\mathscr{B}$ proves $CANS \wedge COMP$ by assumptions as well as $\neg CONS \wedge \neg COMP$ by Theorem 2, (c) and (d). Hence, we have a contradiction. ■

If we change (B.3) so that it asserts the property of P in (B.3) without maintaining the existence of P, the above theorem maintains that there is no possibility for defining a concept of the social validity satisfying (C.1)–(C.5), i.e., we obtain an *undefinability theorem* of the truth, justice and/or a 'complete' social validity.

(Graduate School of Economics, Osaka University, Japan)

References

Fraenkel, A. A., Y. Bar-Hillel, and A. Levy (1973) : *Foundations of Set Theory*, 2nd Revised Ed. (Elsevier/Amsterdam).

Jech, Thomas (1997) : *Set Theory*. 2nd ed. (Springer/Berlin).

Kunen, Kenneth (1980) : *Set Theory*. (Elsevier/Amsterdam).

Lakoff, George (1987) : *Women, Fire, and Dangerous Things*: What categories reveal about the mind. (The University of Chicago Press / Chicago).

Putnam, Hilary (1983) : *Realism and Reason*, Philosophical Papers Volume 3. (Cambridge University Press/New York).

Urai, Ken (2002) : "Why there isn't a complete description of the human society I: The rationality and individual," Graduate School of Economics, Osaka University, mimeo.