

# ON JOHN RAWLS'S TWO PRINCIPLES OF JUSTICE

————— Their theoretical development —————

by RYOJI TOYA (戸谷良二)

John Rawls is Professor of Philosophy at Harvard University. He has devoted his scholarly career over thirty years exclusively to studies of theoretical problems of social justice. His major work, *A THEORY OF JUSTICE* (henceforth TJ), appeared in 1971. Because of its quality of originality and of thoroughness, TJ has aroused wide and lively debates among philosophers, economists and other social scientists since then. TJ is surely the touch stone for contemporary discussions in this area of investigation.

But it is no easy task to have a full understanding of TJ, for it deals a wide variety of topics, from various arguments for basic principles to their application to politics and social problems. Also, discussions and arguments it includes are quite complicated, sometimes not clear, and often seem to be inconsistent. TJ is not so systematized or unified as Rawls intends to be.

However, the central point to his theory is very much clear, that is, to construct and justify the two principles of justice under the "original position", a hypothetical contractarian situation of choice. This fundamental idea has been held consistently since 1958 article "JUSTICE AS FAIRNESS" (henceforth JF), in which the two principles of justice received their first formulation. A dozen years between JF and TJ were spent on amending or deepening his theory to meet objections and criticisms to the first formulation of the two principles in JF. Complexities mentioned above in TJ can be seen as a result of this theoretical development which Rawls had to embark on.

So, my aim in this paper is, first, to compare the first formulation of the two principles of justice in JF with the final one in TJ and to trace the course Rawls took to make his theory more plausible; and second, to examine two crucial notions Rawls introduced in the later version of his theory; namely, the notion of "the veil of ignorance" and that of "the difference principle". I hope this brief survey will be of some help for grasping the heart of his theory.

## 1. BACKGROUND

Before proceeding the main issue, I will take a glance at the historical-

theoretical background from which Rawls began with his problem.

In his preface to TJ, in explaining his aim and intention, Rawls mentions utilitarianism and intuitionism as two major rival schools of normative ethical theory which have been dominant in modern Anglo-American ethics, and says, "The outcome is that we often seem forced to choose between utilitarianism and intuitionism".<sup>①</sup> But, for him this choice is dilemmatic; each of them has strength and weakness as follows:

Utilitarianism

- Strength: can enunciate a rational rule or procedure for settling ethical disputes because of its teleological nature with one ultimate standard, e. g. the principles of utility that defines the good as satisfaction of (rational) desire.
- weakness: cannot explain fully our common convictions concerning the priority of justice, for it defines the right as maximizing the good and thus subordinates the former to the latter.

Intuitionism (interpreted broadly as pluralism; namely as the doctrine that there is a plurality of first principles to be weighed against one another by intuition.)

- strength: can define the right independently of the good, and so match our moral sentiments.
- weakness: cannot show any constructive and recognizable ethical criteria; in case of moral conflicts it only tells us to rely on our each intuition.

Morally, Rawls wants to preserve the priority of the right with intuitionists, but methodologically he is on utilitarian's side. Intuitionism, failing to show any method, any priority rules for weighing conflicting principles, seems after all only to express one's moral conviction.

So, his problem is to construct a substantial theory of justice without relying on intuitive principles, in other words, a theory of justice with rationally recognizable criteria. At this point he finds a way out in the tradition of the social contract theory:

. . . What I have attempted to do is to generalize and carry to a higher order of abstraction the traditional theory of the social contract as represented by Lock, Rousseau, and Kant . . . . this theory seems to offer an alternative systematic account of justice that is superior, or so I argue, to the dominant utilitarianism of the tradition. <sup>②</sup>

The strategy Rawls adopts is to construct principles of justice as agreed on by self-interested rational persons who are situated in the contractarian initial situation of choice. Rawls assigns quasi-formal constraints to this situation and persons so that any principles agreed on unanimously under such circumstances will be, by the very nature of the agreement, "just". For Rawls the theory of justice is "part of the theory of rational choice".

## 2. THE FORMULATION OF TWO PRINCIPLES OF JUSTICE IN "JUSTICE AS FAIRNESS"

In JF, Rawls confines his concern with justice to the sense "only as a virtue of social institutions, or what I call practices".<sup>③</sup> The word "practice" is defined here as "any form of activity specified by a system of rules which define offices, roles, moves, penalties, defences, and so on, and which give the activity its structure".<sup>④</sup> Questions of justice in this sense arise "when conflicting claims are made upon the design of a practice and where it is taken granted that each person will insist, as far as possible, on what he considers his rights".<sup>⑤</sup> Then the principles of justice are to be invoked "for formulating restrictions as to how practices may define positions and offices, and assign thereto powers and liabilities, rights and duties".<sup>⑥</sup> They deal with some aspects of distributive justice.

A practice is an institution or a system within a society. So, to show the process that leads to the two principles of justice, Rawls starts by assuming "a society of persons amongst whom a certain system of practices is already established", and persons are also assumed to have the following characteristics:

- 1) They are mutually self-interested; their allegiance to their practices is founded on the prospect of self-advantage.
- 2) They are rational; they know their own interests, and are capable of tracing out the likely consequences of adopting one practice rather than another.
- 3) They are free from the fault of envy; the bare knowledge or perception of the difference between their condition and that of others is not a source of great dissatisfaction.
- 4) They have roughly similar needs and interests; fruitful cooperation among them is possible. They are also sufficiently equal in power and ability; none is able to dominate others.

Under these circumstances, each person proposes by turns the principles upon which his complaints as well as those of others are to be tried, and by

which he will be bound in future occasions. This procedure keeps going until everyone is of one mind as to how complaints are to be judged, and thus acknowledges a set of principles of general kind that should be applied impartially to each one's conduct and constitute constraints upon one's pursuit of interests.

The two principles of justice are settled on as the solution to the above bargaining situation. They are formulated in JF as follows:

First, each person participating in a practice, or affected by it, has an equal right to the most extensive liberty compatible with a like liberty for all, and second, inequalities are arbitrary unless it is reasonable to expect that they will work out for everyone's advantage, and provided the positions and offices to which they attach, or from which they may be gained, are open to all. These principles express justice as a complex of three ideas: liberty, equality, and reward for services contributing to the common good. ⑦

The first principle expresses "the initial position of equal liberty", the starting point of bargaining, and departing from which will always require a justification. However, the persons are always ready to move to more advantageous positions because of their self-interestedness as conditioned above. So, if they find that a certain practice with an inequality may guarantee a higher level of benefit than the same practice without the inequality, it is reasonable for them to prefer unanimously the former to the latter. This necessitates the stipulation of the second principle. "The second principle defines what sorts of inequalities are permissible; it specifies how the presumption laid down by the first principle may be put aside".<sup>⑧</sup> An inequality here must be understood as meaning the difference in the rewards or burdens assigned by rules to roles and positions, not the difference yielded from the exploitation.

And the second principle holds that "an inequality is allowed only if there is reason to believe that the practice with the inequality, or resulting in it, will work for the advantage of every party engaging in it".<sup>⑨</sup>

Now, from a brief view of the two principles above, we can say that the originality of Rawls's theory lies in his devising how to balance the egalitarian point of view with the claim for more efficient arrangement of social institutions and systems. Then, the second principle is the key of his theory.

### 3. QUESTIONS ABOUT THE FIRST FORMULATION IN JF

There have been many criticisms and objections to the formulation of two principles of justice in JF, most of which are focused on how to interpret the second principle. In this paper I will mention three of them that seem important to later development of Rawls's theory.

1) How does the second principle relate to the first principle? The first principle concerns the equal liberty of persons in the initial situation, but, by "liberty" what does Rawls mean? This point is not clear in JF. In view of the discussion on the priority of liberty in TJ, where he speaks of the basic liberties of citizens such as freedom of thought and liberty of conscience, we can say that in JF also Rawls's "liberty" means such political liberties as secured by modern democratic political systems. On the other hand, the second principle clearly concerns the distribution of economic advantage. Then, does it mean that the economic advantage might override the equal liberty or the basic civil rights?

2) How is the economic surplus which inequalities may yield to be distributed? Is any way of distribution, if it will work out for everyone's advantage, just? The reason for introducing an inequality into a practice is that it can contribute to the better-off of all members. If so, there should be the economic surplus yielded from the inequality. Clearly there can be many ways of distributing it. For example, suppose a practice with five positions, filled by five persons, A-E. Now, 1,000 units of surplus income are gained by introducing an inequality, which assigns burdens and rewards to each position so that the practice would run most efficiently. Five persons are arguing how to distribute this surplus, and A proposes such a distribution plan as (A:956, B:1, C:1, D:1, E:1), B proposes (A:1, B:956, C:1, D:1, E:1), and so on. The second principle cannot solve this dispute, for it is indifferent among these plans so far as any proposed one may guarantee the minimum increase in income of every position. But, the redistribution of social benefit is surely one of the most important subjects of justice.

3) Would the second principle with the competitive interpretation of "open to all" clause be unanimously agreed on? Rawls comments in JF that "open to all" means "must be won in a fair competition". However, in JF, persons are assumed to have a good knowledge of talents and abilities of oneself and of the others. Each person can know the relative position to the others in competition. So, more talented persons will surely agree to the second principle, and the less talented will also surely reject it because they cannot have any prospect of better-off position. Then the bargaining will come to a break down.

#### 4. THE REVISED FORMULATION OF TWO PRINCIPLES IN "A THEORY OF JUSTICE"

In response to questions such as above cited, Rawls has made many theoretical modifications, alterations and qualifications which are summed up in the final formulation of two principles of justice in TJ:

##### First Principle

Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.

##### Second Principle

Social and economic inequalities are to be arranged so that they are both:

(a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and

(b) attached to offices and positions open to all under conditions of fair equality of opportunity. ⑩

Major points Rawls has elaborated answering to above three questions are as follows:

1) The meaning of "liberty" is made explicit as "a complex of rights and duties defined by institutions". It is a system of basic rights of citizens secured by the concept of the rule of law. And Rawls gives priority to the equal liberty of citizenship over social and economic advantages; the two principles are arranged in "lexical order", that is, the first principle in the ordering must be satisfied before moving on to the second, the second before the third, and so on. Therefore, "liberty can be restricted only for the sake of liberty".

2) Rawls admits the ambiguity of "for everyone's advantage" clause, and shows two interpretations of it; one is the principle of efficiency and the other is the difference principle. The principle of efficiency, which is simply the principle of Pareto optimality applied to the basic structure of society, holds that a distribution is efficient when there is no way to change it so as to make at least one person better off without making other persons worse off. But, as we have seen, there are many efficient distributions, and this principle cannot single out one of them as best. To remove this indeterminateness, Rawls chooses the least advantaged persons as a position from which social and economic inequalities are to be judged, and says:

Assuming the framework of institutions required by equal liberty and fair equality of opportunity, the higher expectations of those better situated are just if and only if they work as part of a scheme which improves the expectations of the least advantaged members of society. The intuitive idea is that the social order is not to establish and secure the more attractive prospects of those better off unless doing so is to the advantage of those less fortunate. ⑩

Rawls calls this interpretation the difference principle, as stipulates it as the second principle (a). The difference principle shows us more egalitarianistic approach to the distribution problem.

3) In TJ, Rawls sets aside the competitive interpretation of “open to all” clause, and qualifies it as meaning “fair equality of opportunity”. But more important is his introducing the notion of “the veil of ignorance”. This notion is an analytical device to rule out the possibility of tailoring the principle to one’s own particular contingencies. The persons in the original position are assumed to be behind the veil of ignorance:

First of all, no one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like. Nor again, does any one know his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism. More than this, I assume that the parties do not know the particular circumstances of their own society . . . . The persons in the original position have no information as to which generation they belong . . . . ⑪

While thus lacking the knowledge about themselves and their stations, the persons are also assumed to know “the only particular facts that their society is subject to the circumstances of justice and whatever this implies”, plus the general facts that affect the choice of the principles of justice such as the principles of economic theory, the basis of social organization, the laws of human psychology, and so on. Under these limitations on knowledge, the persons ought to evaluate principles solely on the basis of generality.

However, if the persons have no knowledge about themselves, then they will not have any interests, desires and purposes, then there will not arise the question of justice, for they lack the motivation for bargaining. At this point, Rawls

introduces two notions; the notion of "the rational life plan" and that of "primary goods".

Rawls adapts J. Royce's thought that "a person may be regarded as a human life lived according to a plan", and develops his theory of goodness as rationality:

The main idea is that a person's good is determined by what is for him the most rational long-term plan of life given reasonably favorable circumstances. A man is happy when he is more or less successfully in the way of carrying out his plan. To put it briefly, the good is the satisfaction of rational desire. <sup>13</sup>

Though plans may vary with one's abilities and circumstances, there are those things as would be preferred more of than less by a rational man whatever his life-plan may be. These are "primary goods". Primary goods are necessary means in carrying out one's life-plan. The persons in the original position don't know what their rational plans are in detail, but they know that they have their own rational life-plans, and also know that they prefer more social primary goods rather than less.

The social primary goods are, Rawls mentions, rights and liberties, opportunities and powers, income and wealth, and a sense of one's own worth. Positions and offices in systems of a society are to be evaluated in terms of expectations of these social primary goods.

Rawls had to make one more important alteration by lowering the veil of ignorance; the scope of the principles of justice has shifted from practices to "the basic structure of society". This shift is necessary because a practice is an institution or a system within a society. Before designing a practice there must be an already established society. So, the persons cannot enter a practice without some amount of knowledge about themselves and their society. Behind the veil of ignorance they are obliged to restrict their concern to the most fundamental arrangement of society. A society is, according to Rawls, "a more or less self-sufficient association of persons who in their relations to one another recognize certain rules of conduct as binding and who for the most part act in accordance with them".<sup>14</sup> The role of the principles of justice is to assign rights and duties to the basic structure of society.

## 5. FURTHER QUESTIONS

Now we have an outline of Rawls's two principles of justice in their full



version. As we have seen, Rawls has devised new notions one after another to meet objections and criticisms. Of these notions, much-discussed and most problem-laden are the notion of the difference principle and that of the veil of ignorance. Here, I will take up these two notions to see what each of them has solved, and, on the other hand, what it newly poses.

### The Difference Principle

By introducing the difference principle, one can choose a system of society as “just”; single out the least advantaged person or group of a system of society, then compare his conditions of well-being with those of the least advantaged person of an alternative system of society, and see which of them is better off.

Besides this, the difference principle plus the notion of “primary goods” has another merit of avoiding difficulties of interpersonal comparisons of utility that have long troubled utilitarians. As long as the least advantaged person can be identified, the difference principle asks us only ordinal judgments of well-being:

If we can decide whether a change in the basic structure makes him better or worse off, we can determine his best situation. We do not have to know how much he prefers one situation to another. The difference principle, then, asks less of our judgment of welfare. We never have to calculate a sum of advantages involving a cardinal measure. <sup>(19)</sup>

Still more, comparisons finding the least advantaged are made in terms of expectations of primary social goods, that is, “to compare men’s situations solely by reference to things which it is assumed they all prefer more of”. Thus, the notion of primary goods can establish a common objective measure in comparisons.

The difference principle, however, poses a serious problem; it has certain implications which contradict our intuitive judgment. Suppose two cases below:

There is a society with five members, A-E. (Each one’s condition of well-being is indicated numerically.) Now, two plans of social re-arrangement are proposed.

	Plan 1
Present	(A:11, B:11, C:11, D:11, E:100)
(A:50, B:40, C:30, D:20, E:10)	
	Plan 2
	(A:100, B:100, C:100, D:100, E:9.9)

E is the least advantaged in the present system. On Plan 1 E is far better off, but other four members are worse off though not worse than E in the present system. On Plan 2 other four members than E are better off, but E is slightly worse off.

The difference principle should adopt Plan 1, and rule out Plan 2. It implies that any benefit to the least advantaged member can outweigh any loss to a better off member as long as it does not reduce the second below the first<sup>(16)</sup>, and it also implies that other members than the least advantaged cannot be better off unless the least advantaged member is better off. The total society will be dominated by the least advantaged member. These implications seem to be unacceptable.

### The Veil of Ignorance

By lowering the veil of ignorance Rawls can force the persons in the original position to adopt the general point of view in choosing the principles. Rawls's original intention is to show that morally substantial principles can be reached through pursuing prudential reasoning with certain constraints. The veil of ignorance is a device to attain this end; it adjusts the original position appropriate to make judgment from "the moral point of view":

We want to define the original position so that we get the desired solution. If a knowledge of particulars is allowed, then the outcome is biased by arbitrary contingencies . . . . If the original position is to yield agreements that are just, the parties must be fairly situated and treated equally as moral persons. <sup>(17)</sup>

Also, by denying to the persons any knowledge of their place in the distribution of natural talents and abilities it shows us Rawls's view that talents and abilities are social, rather than personal, resources.

However, despite its attractiveness, we must say that this notion has complicated the matter insolubly. On account of limited space I will itemize some crucial points:

(1) Rawls forbids the persons in the original position to have any knowledge of particular facts about themselves and their stations; no one knows who he is. Then, there cannot be any distinction between persons. Where can we seek for each one's identity? If so, why a contract or a bargaining be needed? Is there any difference between an impartial and rational legislator's designing a system of society and such abstract persons' assembly for agreement on that matter?

(2) Rawls assumes that the persons know the general facts about society

such as the principles of economic theory, and the laws of human psychology. The principles of justice must be applied to any society regardless of time or locality. General facts, the knowledge of which the persons are to have, also must be. But, no economics or no psychology has yet reached such a universal stage. Rather, by the very nature of empirical sciences, it is simply impossible; economics is deeply embeded in the social system from which it has grown, for example. Those general facts which Rawls posits are nonexistent.

(3) Rawls introduces the notion of primary goods so that a bargaining under the veil of ignorance might be possible. Primary goods are those things which every rational man wants more of rather than less. But, how Rawls can identify primary goods? What a Bushman wants more of differs from what a citizen in Tokyo wants more of. By burying this difference in one's way of life under the veil of ignorance, and then by listing certain things as primary goods, doesn't Rawls force us a certain pattern of life or a certain type of man as a norm? Further, Rawls mentions rights and liberties, opportunities and powers, wealth and income and a sense of one's own worth as social primary goods. There are category differences between them. Rawls often speaks of "an index of primary goods", but, although wealth and income may be easily indexed, how are we to index a sense of one's own worth? They must be integrated in order, but this poses as difficult a problem as that of interpersonal comparisons of utility.

## Notes

- ① John Rawls, *A Theory of Justice* (Oxford University Press, 1971), preface viii
- ② *Loc. cit.*
- ③ J. Rawls, "Justice As Fairness," *Philosophical Review*, 67 (1958), p. 164
- ④ *Loc. cit.* notes
- ⑤ *ibid.* p. 172
- ⑥ *ibid.* p. 164
- ⑦ *ibid.* p. 165
- ⑧ *ibid.* p. 167
- ⑨ *ibid.* *Loc. cit.*
- ⑩ Rawls, *op. cit.*, p. 302

- ⑪ *ibid.*, p. 75
- ⑫ *ibid.*, p. 137
- ⑬ *ibid.*, pp. 92–93
- ⑭ *ibid.*, p. 4
- ⑮ *ibid.*, p. 92
- ⑯ See Kenneth Arrow, “Some Ordinalist-Utilitarian Notes on Rawls’ Theory of Justice, *Journal of Philosophy*, 70 (1973), p. 251
- ⑰ Rawls, *op. cit.*, p. 141

(とや・りょうじ 京都大学文学部研修員)