



Behind the Veil of Ignorance

«Behind the Veil of Ignorance»

by János Kis

Source:

Croatian Journal of Philosophy (Croatian Journal of Philosophy), issue: 5 / 2002, pages: 131-159, on www.ceeol.com.

Behind the Veil of Ignorance

JÁNOS KIS

Central European University, Budapest

The paper examines consensual contractarianism John Rawls proposed in his A Theory of Justice, and develops the following criticism. The veil of ignorance device requires but cannot secure the neutrality of the primary goods. In the Rawlsian 'original position' of contract, the only relevant information the hypothetical choosers are allowed to have is that they all prefer to have some 'primary goods' rather than not to have any, and that they prefer to have more rather than less of the primary goods. This stipulation entails that the 'primary goods' are neutral with regard to the diverging preferences of the choosers. In other words, for the Rawlsian contract to yield acceptable results, neutrality of the primary goods must hold. It cannot, however. Hence Rawls' account of a consensual contract is untenable. The paper suggests that the difficulty is not rooted in the particular features of the Rawlsian theory but in the very idea of a consensual contract.

The idea of a contract involves that of agreement. The notion of 'agreement', however, has a shifting scope. Sometimes it denotes 'compromise', at other times 'consensus'. Compromise agreement emerges where individuals with conflicting interests find a mutually beneficial trade-off. Consensual agreement, on the other hand, is one in which the individuals concerned share a common objective, or pursue compatible goals that do not necessitate trade-offs for their realization.¹ Contractarian theories, which present the notion of agreement in terms of compromise, argue from a process of bargaining; consensualist theories model the contract on common deliberation. In order to yield determinate results, compromise theories must specify the characteristics of the parties to the contract. These characteristics include, for example, the bargaining capacities, and the internal and external resources of the parties involved. Consensus theories, on the other hand, can neglect the particular features of the participants; moreover, at least some features, such as those which would distort the consensus or prevent it from emerging, must be eliminated from the hypothetical *status quo* in which the parties are supposed to seek agreement.

¹ For the distinction, see A. Hamlin: "Liberty, Contract and the State," in A. Hamlin and P. Pettit (eds.), *The Good Polity* (Oxford: Blackwell, 1989), 90.

This is an awkward move, however, and the present article will center on the difficulty resulting from it. I will focus on the version of consensual contractarianism John Rawls developed in *A Theory of Justice*. My argument will go like this. The only relevant information the hypothetical choosers are allowed to have, in the Rawlsian ‘original position’ of contract, is that they all prefer to have some ‘primary goods’ rather than not to have any, and that they prefer to have more than less of these primary goods. This stipulation entails that the ‘primary goods’ are *neutral* with regard to the diverging preferences of the choosers. In other words, the neutrality of primary goods must hold for the Rawlsian contract to yield acceptable results. But this does not work. Hence Rawls’s account of a consensual contract is untenable. I will suggest that the difficulty consists not in the particular features of the Rawlsian theory, but in the very idea of a consensual contract.

Section I offers an explanation as to why Rawls opts for a consensualist version of contractarianism. Section II shows how the neutrality thesis is implied by the premises of the Rawlsian theory. Section III raises the objection of incoherence, whilst Section IV examines three attempts at salvaging the neutrality claim and shows them to be unsuccessful. Section V addresses the question as to whether the said difficulty arises out of the particular form of the Rawlsian theory, or whether it is damaging to the very idea of consensual contractarianism. Here I will briefly consider Rawls’s attempt at solving the problem in his later work.

I

Society is a cooperative venture for mutual advantage, says Rawls at the beginning of *Theory*.² By coordinating their action, individuals produce a surplus of outcomes that would otherwise not obtain in the absence of cooperation. That surplus might include a wide diversity of goods: peace and security, freedoms, income and wealth, and so on. The problem of justice, then, is how the fruits of cooperation should be distributed among the members of society so that nobody should have a defensible complaint about being treated unfairly.

Suppose self-interested individuals want to find out which scheme of cooperation is most advantageous to them. Suppose also that each individual has the right of veto, so that arrangement A cannot be enforced unless everybody prefers A to any of its alternatives, which means that if there is no agreement about A the search must start again. Call this the unanimity condition. If the unanimity condition is met, then cooperation will be to the *mutual advantage* of the participants; for no rational individual would endorse a particular scheme of distribution unless it is better (or at least not worse) for him to cooperate under that scheme than to join a different system of cooperation or to refuse cooperating altogether.

² See J. Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), 4.

One is tempted to assume that if cooperation results in the best possible outcome for all, and is subject to the choices of other parties who also seek the best possible outcome for themselves, then nobody can complain of being treated unfairly. That is to say, voluntary cooperation for the sake of mutual advantage seems to be necessarily fair, and the principles which regulate and justify such cooperation would be principles of justice. Now, if this is the case, then it is plausible to regard a theory of justice as a theory of compromise agreement, insofar as it assumes that the cooperating parties each seek out their own self-regarding interests, and that there is no *a priori* guarantee that these various interests will be directly co-possible.

Nevertheless, Rawls abandons the compromise model early in his argument, and proceeds to spell out a consensualist theory. Suppose, he says, the parties to the contract bargain so as to procure mutual advantage. The outcome of the bargaining process will vary with the initial conditions under which the contract is made. These conditions range over the personal capabilities of the contractors (e.g. their negotiating skills) and their impersonal resources (e.g. their financial assets). The distribution of personal and impersonal resources, however, might vary depending on empirical contingencies, and such variations do not leave the outcome of the bargaining process unaffected. In other words, the outcome of the agreement will reflect the initial conditions. Almost any conceivable outcome could be agreed upon by rational individuals, provided that those initial conditions are suitably defined. "To each according to his threat advantage is not a principle of justice," says Rawls.³

Mutual advantage does not yield, in itself, outcomes which would be either determinate or morally privileged. Universal consent is a necessary, but not a sufficient condition for the principles accepted to be just. It does not identify justice unless the consenting people make their choice under a *fair* initial distribution of bargaining resources. So in order for us to isolate principles of justice within a larger set of possible principles of cooperation, the conditions of fair bargaining must be specified. This is why Rawls calls his theory *justice as fairness*: the claim is that if and only if unanimous consent is reached from a position which is fair, the principles universally accepted will be principles of justice. What makes them just is not their inherent properties, but the way in which they are identified. Thus whatever rule the parties come to agree upon in a fair bargaining situation is a principle of justice. The spelling-out of the conditions of fairness circumscribes what Rawls calls the *original position*.

The Rawlsian theory of justice, then, is founded on a two-pronged argument: principles of justice are chosen by individuals who are 'mutually disinterested', i.e. "not taking an interest in one another's interests," and the choices are made under constraints which pin down

³ *Ibid.*, 145.

the requirements of morality.⁴ It is the spelling out of the fair initial conditions which shifts the contract from a bargaining paradigm to a paradigm of consensus-seeking. One obvious strategy for identifying the original position of fairness would be to repeat the thought experiment of the hypothetical contract, and to claim that those contract-making rules which rational individuals, who pursue self-regarding preferences, would unanimously consent to are the ones that are fair. Rawls believes, however, that this procedure would end up in an infinite regress: one would require a second contract to establish the rules under which the contracting rules are made, then a third contract for the rules of the second, and so on *ad infinitum*. So we are faced with the task of stopping the regress, and Rawls solves the problem by denying that the conditions of fairness could be determined by a contractarian argument. They should be identified, he notes, with the help of a wholly different procedure which he calls the method of *reflective equilibrium*.

Reflective equilibrium works in the following way. In the course of our moral reflections, we often make judgments about particular acts and omissions (Rawls calls these 'moral intuitions'). We start the argument by collecting these judgments together and examining them so that, in the end, we obtain those which survive critical scrutiny. Next, we construct general principles which give structure to our separate judgments and explain them within a framework of a unified moral theory. Once the principles are established, we then return to the particular judgments in order to see whether they fit into the general theory. The examination might show that this schema is sufficiently comprehensive, so that we have good reasons to accept the theory, but that some initial intuitions are nonetheless in conflict with it. If so, then the requirement of consistency demands that we revisit these intuitions and modify them accordingly, so that the conflict disappears. Some particular judgments might yield to the pressure; some others, however, might resist it very strongly. Consequently, we go back to the principles themselves in an attempt to reach consistency by introducing changes at this end of our moral conception. And we continue this back-and-forth movement between particular judgments and general principles until a point is reached where neither our considered judgments nor our abstract principles yield to any further attempt at revision. This will be the point of reflective equilibrium.

Rawls claims to pin down the original position by means of the method of reflective moral reasoning. He starts from the intuition that human beings are to be individually treated as free and equal rational persons. He then explains what freedom and equality mean. Individuals are free beings in that they have the capacity to form, revise and pursue their own *conception of the good*, and are endowed with a *sense of justice*. Namely, humans do not blindly follow their wants and desires: they have the reflective capacity to 'stand back' from their imme-

⁴ *Ibid.*, 13, 584.

diate motives and to ask whether the things desired are worthy of being desired, whether it is really good to pursue them, and whether it is just to reach a personal goal given the costs its pursuit imposes on others. Briefly, human individuals have the capacity of testing their wants and desires against their conception of the good and principles of justice, and to subject their conduct to the results of this critical test. They are autonomous persons. And they are equals insofar as they all have the capacity to be autonomous, at least to a required minimum degree. No social arrangement can be just unless it treats all the individuals as equals in this sense.

But what does it mean to treat individuals as equals? One possible interpretation is offered by utilitarianism, a moral and political theory Rawls strongly objects to. When we face competing options, utilitarians say, nobody's preferences should be excluded from consideration, nor given a smaller weight than anybody else's. Everybody should count for one, nobody should count for more than one. But what are we to do with the preferences, equally weighed across persons? According to the utilitarians, they are to be aggregated, and priority should be given to that alternative under which the aggregate (or average) preference satisfaction of all individuals concerned is greatest. Thus equal regard means impartially counting the wants and desires of everybody as inputs to the utilitarian calculus, which is supposed to identify the maximum aggregate (or average) preference satisfaction or welfare as its output.

Rawls objects to this understanding of equal treatment because it does not take seriously the distinctness of human lives. Treating people as equals must cope with the important fact that no individual is indifferent about his life and the myriads of lives of other human beings: each person has his own life to lead, and the success of this particular life is not, for him, the same as the success of other lives, which are lead by different people. The normative significance of the separateness of the lives of human persons prohibits, Rawls believes, aggregating preferences across individuals. Aggregate or average utility cannot be the standard of justice for a society which treats its members as free and equal rational persons. This constitutes the most telling reason why Rawls proposes contractarianism as the appropriate method to tackle the problem of the principles of justice. Each party is represented as having the right of veto, and the individuals's right to veto prevents any trade-offs between the interests of one, single individual and the aggregate interests of many other individuals which is at the core of the utilitarian reasoning. In other words, contractarianism excludes preference aggregation across individuals, and thereby awards appropriate weight to the distinctness of human lives. This is why Rawls maintains that the idea of contract provides a correct account of what equality requires.

Once the intuitive ideas of freedom and equality are spelled out and connected to the contract paradigm, the initial conditions for a fair agreement can then be determined. Rawls's reasoning is as follows.

Because they have the capacity of reflection, human individuals can be held responsible for their acts and omissions. And if they can be held responsible, then it is not morally arbitrary for somebody to enjoy advantages or suffer disadvantages so long as these reflect his own choices and efforts. A distribution of costs and benefits, however, which is affected by factors independent of personal choices and efforts, is morally arbitrary. The contingent position the individual occupies in the structure of class and social status is an example of morally arbitrary advantages or disadvantages: nobody can be held responsible for the economic and cultural capital he inherits by being born into a particular family. More interestingly, Rawls also includes among the morally arbitrary—and therefore unfair—advantages and disadvantages the place the individual has in the chance distribution of natural talents and disabilities. For the original position to be fair, he claims, such arbitrary sources of advantage and disadvantage should not be allowed to influence the bargaining process. Legitimate social inequalities must depend on such variables for which individuals can be held responsible, i.e. which reflect their choices, actions and omissions. The distribution of the cooperative surplus should not vary with circumstances which do not belong to the domain of individual responsibility. ‘Undeserved inequalities’ ought not to affect the determination of principles of justice.

The fairness of the original position entails, therefore, a distribution of the bargaining resources in such a way that nobody enjoys undeserved advantages or suffers disadvantages for which he cannot be held responsible. But, claims Rawls, if the parties to the agreement are well-informed about their future class situation, or the future value of their inherited talents, then this knowledge can introduce an unfair bias into the bargaining process. The only way to make the hypothetical original position conform to the requirements of fairness is to impose moral constraints as informational constraints, as limitations on the knowledge available to the participants. What cannot be known by the choosers in the original position cannot influence their decisions. Hence the parties to the agreement must have as little information as possible in the original position. They must proceed from behind a *veil of ignorance*.⁵ Principles of justice, then, are those rules of social cooperation which rational contractors, deliberating behind the veil of ignorance, would agree to.

Now, we can see why Rawls abandons the compromise model in favor of the consensual model. The veil of ignorance renders the parties to the contract fully interchangeable, in the sense of having exactly the same interests (given their knowledge about the world and themselves). Thus if we succeed in showing that a particular individual A would choose, behind the veil of ignorance, a set of principles as maximally advantageous for himself, then it follows that individuals B, C, ..., N would also choose the same set of principles. The model only

⁵ *Ibid.*, 12, 137.

requires one, single individual to make decisions for all the people concerned. Each party is seen to be “convinced by the same arguments,” so that “we can view the choice in the original position from the standpoint of one person selected at random.”⁶ Once that person ‘selected at random’ is shown to agree with a set of principles, it can be directly inferred that all the persons will also agree. No room is left open for bargaining and compromise.

The veil of ignorance, of course, sets out artificially designed conditions. As a rule, real individuals do not deliberate in the absence of all information about their particular characteristics (be they morally arbitrary or not). The theory of justice cannot be complete, therefore, unless an account is given of the authority the hypothetical contract and its outcomes have for people in ordinary life. Two interpretations of this can be put forward. First, the hypothetical contract should be seen as a limiting case which real-world decisions are to mimic and approximate themselves to as much as possible. On this reading, the hypothetical contract sets an *ideal* to guide the moves of real people making real contracts without, however, imposing *obligations* on them. The outcomes of real agreements, and these alone, have binding force. And then, ultimately, the justification of the principles of justice would not be consensualist, because real individuals, endowed with sufficient information about their diverging interests, cannot reach any agreement other than through bargaining. But once we understand the idea which is latent in ‘justice as fairness’, we should observe that the limiting case picture does not provide an accurate interpretation of *Theory*. There are no guarantees regarding the distribution of resources, as well as the distribution of information about this distribution, so that the participants of the contracting process will bargain from fair initial circumstances, which are at the core of justice as fairness. So we have to opt for a second interpretation, and Rawls himself is clear about this.

According to this second interpretation, once the original contract is made and the veil of ignorance removed, people are not allowed to remain in a position of mutual disinterestedness. The reason why they are permitted to follow their self-regarding interests, and nothing else, behind the veil of ignorance is because the veil constrains individual choices in such a way as to secure the satisfaction of the basic requirements of fairness, whatever the choosers’s decisions may be (provided they are rational). “[T]he combination of mutual disinterest and the veil of ignorance achieves the same purpose as benevolence,” asserts Rawls.⁷ But fairness does not automatically hold in ordinary life: “In practical affairs an individual does have the knowledge of his situation and he can, if he wishes, exploit the contingencies for his advantage.”⁸ Therefore, the principles of justice cannot effectively constrain choices

⁶ *Ibid.*, 139.

⁷ *Ibid.*, 148.

⁸ *Ibid.*

in the real world unless an individual is motivated by the desire “to advance the interests of others and to see their ends attained,”⁹ so that “[h]e voluntarily takes on the limitations expressed by [...] the moral point of view.”¹⁰

Rawls’s assumption is that the intuitive ideas of freedom and equality are shared by a sufficiently large number of real people and, thus, after due moral reasoning, a reflective equilibrium based on these ideas, appropriately spelled out and reconsidered, can be reached. People who agree upon the fundamental moral idea that society should treat its members as free and equal rational persons will also agree that social cooperation should be regulated by principles which are fair, in the sense of respecting everybody’s freedom and equality. They will agree that individuals should be allowed to pursue their own self-interests within the constraints of the principles of justice only. This is why Rawls opens *Theory* with the statement that “[j]ustice is the first virtue of social institutions.”¹¹

So rather than mimic the self-interested search of the hypothetical choosers for principles which maximize their personal advantage, actual individuals are driven by a common desire to find shared principles of justice, and ought to use the hypothetical contract as a *test* which any proposed principle must pass to qualify as fair and, therefore, just. Call this the ‘veil of ignorance test’. The role of the veil of ignorance test, then, is to help people motivated by a shared concern for freedom and equality to identify the requirements of justice. For us, real individuals acting in the real world, a proposed principle is not a principle of justice unless we come to see that hypothetical individuals, negotiating from a hypothetical position of fairness, would unanimously endorse it as serving their personal advantage.¹² This is the deepest sense in which the Rawlsian contract rests on the logic of consensus, and not that of compromise.

II

For the veil of ignorance test to yield determinate results, however, two further assumptions must be made. Rawls is aware of these, and he is confident that they can obtain. The need for the first assumption might be explained in the following manner.

Suppose the hypothetical contractors have a choice between two alternative institutional schemes, P and Q, and suppose that in both there are two income classes, the rich and the poor. In P the rich are

⁹ *Ibid.*, 129.

¹⁰ *Ibid.*, 148.

¹¹ *Ibid.*, 3.

¹² Note that this understanding of justice as fairness contradicts Rawls’s claim that his theory falls under the heading of pure procedural justice: we have a criterion of fairness, independent of the procedures of choice, and the principles of just procedure will be those which can be shown to comply with this substantive criterion.

better-off than they would be in Q, but the poor are much worse-off. Behind the veil of ignorance, nobody knows where, on the income scale, he will land in the real world; one cannot even attach probabilities to the chance of ending up rich or poor. Now, if an individual is strongly risk averse, then it is rational for him to choose that system in which the poor are better-off, and disregard the difference between the relative success of the rich in P and Q. But if his risk aversion is very weak, then reason will tell him to choose that system under which he would fare best, provided he ends up at the top, no matter what alternatives he would face by ending up as one of the poor. Thus the hypothetical contract cannot be expected to yield a determinate choice of institutions unless we can attribute a characteristic attitude towards risk as something uniquely rational to any contractor. Indeed, Rawls makes the assumption that behind the veil of ignorance, where people cannot even attach probabilities to the various possible outcomes, the rational choice strategy is to be extremely prudent. Under the condition of complete uncertainty, the advantages of landing in a better-off group are overridden by the dangers of ending up among the worst-off. Thus reason commands those persons choosing principles of social cooperation behind the veil of ignorance to opt for a choice strategy which Rawls calls *maximin*. Maximin recommends that, if we face two different options, we should focus on the worst possible outcome in each, and choose that option under which the worst outcome is the least detrimental.¹³

To see why the second assumption is necessary, we have to look more closely at the information the veil of ignorance is said to suppress. According to Rawls, this information comprises, besides one's inherited social position and natural talents, one's psychological propensities and particular conceptions of the good life. The last assumption seems, on the face of it, to be odd, because Rawls makes it very clear that his theory does not treat humans as 'passive carriers of desires' and that people can be held responsible for their conceptions of the good. As beings endowed with the capacity to subject their beliefs to reflective scrutiny, humans are responsible for the conception of the good they hold insofar as they are expected to be open to re-examining it in the light of critical objections, and to adjust their actions and plans to those convictions they would hold on the basis of the best available reasons.¹⁴

¹³ Without this assumption, no plausible argument can be made in favor of the famous 'difference principle', which states that resources should be distributed equally across individuals unless inequality is necessary to improve the situation of the worst-off.

¹⁴ *Ibid.*, 94. See "Social Unity," 169. Rawls implicitly rejects the interpretation of neutrality as the neutrality of the outcome. This is made explicit in "Priority of Right and Ideas of the Good." See his *Political Liberalism* (New York: Columbia University Press, 1993), 190 ff.

So Rawls cannot maintain, in the straightforward sense, that a person's conception of the good is determined by natural hazard or social contingency. He must have something else in mind. Unfortunately, he is not very explicit on this point. But the following seems to be consistent with the general tenets of his theory: What, with respect to natural talents or social endowments, should be excluded from consideration as morally arbitrary is not the absolute level of intelligence or physical strength, or inherited wealth for that matter, but one's place in the chance distribution of these resources. And the same thing holds for conceptions of the good also. Even if a person can be held responsible for the fact that she clings to a particular conception of the good life, she bears no responsibility for the fact that other people hold the same or some different conception. The distribution of personal ends, values and desires is a social contingency, and these are thus the advantages and disadvantages stemming from the fact that an individual holds a conception which is shared by a smaller or larger number of people. The influence of such contingent numbers on the choice of the principles of justice is, I suspect, what Rawls wants to 'nullify' by suppressing the knowledge of particular conceptions. If the parties to the agreement knew their own conceptions, then the majority would enjoy a bargaining advantage which enables them to push through principles that are biased in favor of their conception. Or two groups could form a coalition, and so on. Rawls obviously wants to exclude any such opportunities because, as we have seen, he wants the hypothetical contract to emerge not from a process of bargaining, but from a convergence of views, wherein each party is susceptible to the same arguments.

Be that as it may, the elimination of conceptions of the good life raises difficult questions. If the hypothetical choosers remain in the dark about their future values and preferences, how then can they have any specific goals to maximize? And, given the absence of such a goal, how can they reach any rational decision? Once conceptions of the good have been excluded, what can serve as the basis for setting such a goal? For his part, Rawls suggests that there is a limited class of all-purpose means which are necessary to further any conceivable goals: "[W]hatever one's system of ends," these "are necessary means."¹⁵ Such all-purposive means he calls *primary goods*. They are not objects of desires, wishes or likings but those of needs: that is, their availability is a general requirement for individuals to lead their lives in society, whatever their system of ends may be. Some of the primary goods are given by nature, like health and vigor, intelligence and imagination. Others are provided by society. On the list of social primary goods Rawls includes rights and liberties, opportunities and powers, income and wealth, and the social bases of self-respect. So far as primary goods are concerned, or so *A Theory of Justice* claims, it is

¹⁵ *Ibid.*, 93.

generally true that whatever else a rational person may want, he will want these as well, and that it is always rational for him to desire more rather than less of the primary goods.¹⁶ What the parties to the agreement are supposed to maximize, therefore, is the expectations of primary goods available to them.

Now, Rawls intimates that the stipulation of primary goods does not put any constraints on the preferences and life plans of the cooperating individuals: "I make no restrictive assumptions about the parties' conceptions of the good except that they are rational long-term plans."¹⁷ *Whatever else* a person might want, it is rational for him to want the bundle of primary goods, and to want more than less of these types of goods.¹⁸ But the implicit position Rawls takes, as entailed by his characterization of primary goods, is different. The assertion that it is always rational for the hypothetical choosers to maximize their expectation of primary goods implies a strong postulate about the desires hidden by the veil of ignorance. The hypothetical choosers are not supposed to know anything about their particular preferences, but they are supposed to know at least this: Nobody in ordinary life can possibly hold a preference such that it might conflict with the desire of having more, rather than less, primary goods (or with the manner the preferences for primary goods are ordered) and which, in cases of conflict, would override them. Otherwise the goal of maximizing the expectation of primary goods would not always be rational, and the suppression of information regarding real-world preferences would bar the choosers from knowing when it is and when it is not so. In other words, the set of possible preferences is constrained by the device of the original position in such a way that the primary goods remain *neutral* with respect to any two ways of ordering them. Call this the assumption of the neutrality of primary goods in relation to possible preferences, or, in shorthand, the assumption of neutrality. Without this assumption, the hypothetical choosers cannot have a goal to maximize, and would thereby not be in a position to rationally choose a unique set of principles for their cooperation. On the other hand, the making of this assumption does more than just facilitate rational choice behind the veil of ignorance; it also provides for the impartiality of the principles chosen, because it implies that the outcomes of the hypothetical process of choosing are not affected by variations in the preferences and conceptions individuals hold about the good. Whatever the conceptions, and whatever their correct ordering (if there is one), the principles of justice chosen in the original position remain the same. Thus neutrality makes a tremendous difference in justice as fairness—provided that the procedure suggested is really neutral in the requisite sense.

¹⁶ *Ibid.*, 92f., 413.

¹⁷ *Ibid.*, 129.

¹⁸ *Ibid.*, 93.

With these premises at his disposal, Rawls proceeds to show that the hypothetical choosers will unanimously agree on the following two principles. First, each individual should enjoy the most extensive liberty compatible with the similar liberty of others.¹⁹ This principle concerns the distribution of basic rights and liberties, such as freedom of speech and assembly, freedom of conscience and thought, freedom from arbitrary arrest and seizure, and the right to private property. The second principle reads as follows: Social and economic inequalities are to be arranged so that they are both (a) attached to positions and offices open to all, and (b) are reasonably expected to advance the position of the worst-off.²⁰

The idea that positions and offices should be open to all is not new, of course. It echoes the old liberal claim of equal opportunity. Rawls, however, gives a special twist to this requirement. Originally, equal opportunity was marshaled against aristocratic privileges, and involved an appeal to the independence of natural talent from social status and class. This means that the barriers obstructing talented people of humble origins should be removed so that they can compete on equal terms with children of the nobility. But this did not demand the reduction of unfair disadvantages of birth through the collective efforts of society. Rawls interprets condition (a) in this more demanding sense. He calls it *fair equality of opportunity*, and depicts it as “roughly equal prospects of culture and achievement for everyone similarly motivated and endowed.”²¹ Under fair equality of opportunity, the hierarchical division of labor is organized in such a way that “positions are open not only in the formal sense, but all have a fair chance to attain them [...] regardless of the initial place in the social system.”²²

Condition (b) is Rawls’s famous *difference principle*, which is based on the general presumption that the outcome of social cooperation must be distributed equally. But, according to the difference principle, departures from this presumption are still permissible, provided that the worst-off fares better under the unequal scheme of distribution than he would under strict equality. It is the worst-off who bears the burden of inequality; unequal distribution is justifiable, therefore, if and only if it makes the worst-off better-off—all things considered (including the disadvantages inherent in his inferior position)—than in a situation of flat equality.

Thus the two principles include three separate norms that regulate the institutions of a just society: maximum equal liberty, fair equality of opportunity, and giving priority to the worst-off (the difference principle). Owing to the plurality of these regulative principles, a problem

¹⁹ In his later works, Rawls substitutes rights and liberties with the ‘most extensive liberty’.

²⁰ *Ibid.*, 60.

²¹ *Ibid.*, 73.

²² *Ibid.*

of weighing arises. There are many different answers to this problem, and Rawls explicitly considers and rejects two of them. Value pluralists (whom he calls 'intuitionists') might hold that the three norms are incommensurable, and that any ordering of them is therefore as acceptable as any other. Utilitarians might claim that we should always give priority to that principle under which the aggregate welfare is maximal. Against the 'intuitionists', Rawls maintains that we should not renounce the possibility of having a theory capable of yielding determinate answers to questions of weighing. To the utilitarians, he objects that they, by making their choices dependent on aggregating the preference satisfactions across different persons, fail to take seriously enough the distinctness of individual lives; namely, they permit sacrificing *one* human being's most important interests (his liberty, or even his life) in order to promote the interests of *other* individuals (e.g. to increase marginally their level of material well-being), which are separately insignificant, but once summed together command great import. As an alternative to intuitionism and utilitarianism, Rawls proposes a solution which gives a determinate answer to the problem of weighing, but avoids aggregating goods across different persons and allows for the inviolability of individual rights. His solution consists in postulating a relationship of lexicographical ordering between the three norms: the first principle is lexicographically prior to the second (at least in societies where starvation is not a problem) and, within the second principle, condition (a) is lexicographically prior to condition (b). Lexicographical ordering requires that, if a principle precedes another, the second should not be implemented until the first is fully satisfied. That is to say, a principle of lexicographic priority should never be satisfied to a lesser degree in order to secure greater satisfaction to another principle. Because of the postulated priorities, one should not interfere with basic freedoms for the sake of greater equality of opportunity, nor should fair equality of opportunity be traded-off against greater material equality. It is this strict ordering which, Rawls hopes, makes his theory capable of giving determinate answers to particular contested issues of politics. He also believes that this keeps his theory firmly within the liberal tradition, notwithstanding its redistributionist bent, because the priority of the first principle excludes the overriding of basic liberal rights either by the aim of fair equality of opportunity or by the benefiting of the worst-off.

III

Some critics assert that, even if the assumptions of the maximin strategy and the neutrality of primary goods obtain, the two principles of justice and their lexicographical ordering do not follow from the premises spelled out by the original position device. For the aims of the present discussion, we can set these kinds of objections aside. Others object that the unique status attributed to the maximin strategy is arbitrary and ill-founded. This involves a much deeper argument be-

cause, if it holds, it might imply that the veil of ignorance recommends the maximization of average utility rather than the pursuit of the maximin strategy. In other words, the Rawlsian idea of justice as fairness and the rejection of utilitarianism cannot go together. For present purposes, I do not want to address these types of claims.

I want to raise some doubts about the assumption of neutrality. Is it true that it is rational to prefer the Rawlsian list of primary goods over any other bundle, no matter what conception of the good one holds?²³ Is it true that whatever our conception may be, we would still prefer the same mix of goods on the list? If the answer to the second question is “no”, then the outcome of the procedure is indeterminate, because we have no criteria for rank ordering the different mixes of primary goods. And if the answer to the first question is “no”, then the choice of the Rawlsian primary goods is based on an implicit commitment to a special class of conceptions of the good life; and therefore the procedure which distributes these primary goods among members of society will not be neutral in the requisite sense.²⁴

Rawls does not ignore that the heterogeneity of primary goods gives rise to the problem of weighing these against each other, and he is conscious about the absence of an index of primary goods in his theory.²⁵ He believes, nevertheless, that for at least the most important purposes the task of defining such an index can be avoided. First, the lexical priority of the principle of equal basic liberties and rights over the other two principles, and that of fair equality of opportunity over the difference principle, solves the problem for rights, liberties and opportunities. The task of weighing primary good becomes, then, one of weighing income and wealth, power and authority, and the social bases of self-respect against each other. Secondly, even this remaining difficulty is made easier by the fact that the weights must not be given for the entire population, but only for the worst-off group because, bearing in mind the aims of the difference principle, it is this group's position alone which matters. This simplification has two advantages. In the case of the worst-off group, income and wealth, on the one hand, and powers and prerogatives of office, on the other, tend to be associ-

²³ Rawls makes a qualification about this claim: he supposes that the choosers in the original position are ‘mutually disinterested’, i.e. they are not driven by other-regarding preferences. Conceptions of the good which give free rein to the desire of domination over others (530), and to envy or rancor (143f.), are excluded in advance, as are those conceptions which entail altruism and benevolence (143f.). There is no need to examine here the bases of these limitations, because our argument is valid whether they are accepted or not.

²⁴ For the thesis that primary goods are not neutral, see T. Nagel, “Rawls on Justice,” *Philosophical Review*, 82 (1973), 220-234, and A. Schwartz, “Moral Neutrality and Primary Goods,” *Ethics*, 83 (1973), 294-307. Both Nagel and Schwartz contend that an individualistic bias is built into Rawls's conception of primary goods.

²⁵ J. Rawls, “Social Unity and the Primary Goods,” in A. Sen and B. Williams (eds.), *Utilitarianism and Beyond* (Cambridge, Mass. and London: Cambridge University Press, 1982), 163.

ated, so that no trade-offs requiring weights or priority arise between these two classes of goods. And the position of the worst-off group can be modeled on its representative member. These two simplifications jointly imply that no interpersonal comparisons should be made. In terms of justice and fairness, it is sufficient to compare the different positions that one representative individual belonging to the worst-off group reaches under different distributions of income and wealth.²⁶

Surely, once the worst-off group is determined, the remaining task seems easy enough to handle. But before that, the worst-off group needs to be identified, and its determination requires comparisons with other groups in society. Furthermore, on those pages of *A Theory of Justice* where Rawls discusses the indexing problem, he is remarkably silent about the primary good of self-respect, though in other places he makes the claim that this is not just one of the primary goods, but is “perhaps the most important primary good.”²⁷ Thus we are left with two unanswered questions: How do the social bases of self-respect compare with the other primary goods, such as income and wealth, in the case of the representative individual of the worst-off group; and how is the worst-off group to be identified against the rest? Suppose that these problems admit a unique solution, provided that there is a unanimous agreement about the description of the primary goods on the list. Even in this case, the issue of indexing may reappear if it should turn out that the primary goods admit of more than one list, depending on what kind of conception of the good an individual accepts for himself. Identifying the primary goods is the key problem.

Suppose we face the question of what institutional structure fits best the principles of justice. To be more exact, suppose that a choice is to be made between an individualist and a communalist social order. We'll assume that both systems are constrained by basic rights and freedoms, but where the predominance of private property, market economy, and voluntary associations of purpose characterize the individualist order, the communalist order is characteristically based on overlapping ownership rights, networks of reciprocal claims, and communities of belonging. These terms can be used in the following way: private property is ownership of any good where the owner has near exclusive control over the use and alienation of that good; a market is a system of exchanges where any two parties can be anonymous to each other; an association of purpose is a combination of individuals who are free to join or leave the group at any time; overlapping ownership denotes a set of criss-crossing claims made by different individuals and/or the group, its subgroups, and its individual members over the same resources, where each individual bears duties towards other individuals and/or groups in their personal network of reciprocal claims; a network of reciprocal claims is a system of exchanges where named persons (such as, for example, members of the same genealogical lin-

²⁶ *Theory*, 93f.

²⁷ *Ibid.*, 396.

eage, people related to each other through marriage, and individuals belonging to the same neighborhood) bear unspecified obligations towards each other; a community of belonging is a group into which the individual is born, and which sets the permanent framework for his different pursuits.

If we agree with the Rawlsian account of primary goods, including the social bases of self-respect, then a preference for the individualist order seems to be secured. This is because Rawls claims that the sense of one's worth has two sources in society: first, the publicly affirmed distribution of fundamental rights and liberties which we, as citizens, enjoy on an equal footing with everybody else; and, secondly, associations of shared interests which we freely establish with others, and where we find our endeavors confirmed by the co-members.²⁸ The institutional requirements for self-respect tend, thus, to go together with those of income and wealth, these latter being more likely to increase under a market system than in an economy largely embedded in reciprocity networks and communities.

Someone, however, might challenge the Rawlsian interpretation of the social bases of self-respect on the grounds that voluntary associations are not the proper bases for a human being's sense of worth. The reason why Rawls gives associations of purpose such a key role in securing self-respect is because self-respect and mutual respect require, according to him, that different individuals's "common plans be both rational and complementary."²⁹ But, a critic might object, this is typical of an individualist way of respecting ourselves and each other, and individualism is not the right perspective for self-esteem. Self-respect and mutual respect, in the appropriate sense, are communalistic attitudes: they are based on belonging, not action; on traditions, not rational plans; on communal identity, not on the choices of our associates on the basis of common purposes. Notice that the argument is ethical, not psychological: it is not drawn from empirical assumptions about the circumstances under which the psychic sense of one's worth is more secure, but from evaluative ideas about the conditions under which self-respect, if given, is appropriate. It belongs, therefore, to that domain where conflicting positions often give rise to what the later Rawls calls 'reasonable disagreement': namely, to such kinds of contests where, even after duly considering all the arguments of the other side, none of the parties is unreasonable to persist in his or her conviction.³⁰

²⁸ Ibid., 544, 442. Though Rawls employs here the term 'community', he makes it clear that he has in mind a group created for special purposes which the individual is free to join or leave. See also 441.

²⁹ Ibid., 441.

³⁰ See Rawls, *Political Liberalism*.

IV

Three lines of argument have been proposed to meet this objection. Perhaps there are more. But the way these three fail suggests that none of them is likely to succeed. The first is offered by Allen Buchanan. In an article published in the mid-1970s,³¹ Buchanan contends that, whatever one's particular plan of life is, a strong preference for those goods which Rawls lists as primary goods is uniquely rational. Rationality requires one to recognize that conceptions of the good are no less fallible than scientific theories. Therefore, it is rational to "maintain an attitude of critical revisability toward one's own conception and of open-mindedness toward competing conceptions."³² If this holds, then it is rational "to attempt to satisfy the epistemic conditions necessary for the effective expression of an attitude of critical revisability."³³ Likewise, it is rational "to attempt to provide for the implementation of those new or revised conceptions of the good which one may develop" in the future.³⁴

According to Buchanan, two consequences follow from these prudential principles. First, the veil of ignorance is not simply a device suited to securing that the principles of justice will not arbitrarily favor some ways of life over others. With respect to a single individual, it "insures that the principles of justice will allow that individual maximum freedom to develop and implement new or modified conceptions of the good."³⁵ Secondly, a rational person will want to have more rather than less of the Rawlsian primary goods, either as the epistemic conditions necessary for the effective expression of an attitude of critical revisability, or as maximally flexible assets required for implementing whatever new or modified conceptions one may develop.³⁶ Buchanan mentions freedom of speech as an obvious item on the list of primary goods, whose purpose is to maintain the requisite epistemic conditions, and wealth as an obvious case of maximally flexible goods.

Now, it is indeed rational to concede the fallibility of one's beliefs and to maintain a critical attitude towards them. But it is a great leap from the principle of revisability and its corollaries to reach the two theses Buchanan wants to defend. It is not clear that allowing one's conception of the good to be hidden under a veil of ignorance is the appropriate attitude for an individual, "viewed as a critical pursuer of ends over time," to take. The idea of revisability entails that a conception of the good is not a drive towards pleasurable states of mind, where pleasure is determined by biological factors, but is a set of de-

³¹ A. Buchanan, "Revisability and Rational Choice," *Canadian Journal of Philosophy*, 5 (1975), 395-408.

³² *Ibid.*, 399.

³³ *Ibid.*, 401.

³⁴ *Ibid.*, 402.

³⁵ *Ibid.*, 406.

³⁶ *Ibid.*, 402.

sires mediated by beliefs and convictions. A conception of the good is revisable because beliefs and convictions are revisable. But if we hold a belief, then we hold it to be true. To take the attitude of critical revisability towards a belief does not include suspending one's commitment to its truth—and it is precisely this sort of suspension upon which the veil of ignorance is modeled. Somebody who wants to occupy the position of an impartial observer with respect to the beliefs he now endorses as true, and to the beliefs which he rejects as false (but might, perhaps, embrace in the future), is not really committed to the truth of his beliefs, nor is he really critical towards them. He is not really committed, namely, because one cannot be simultaneously convinced of the truth of a belief and allow a conflicting belief to have an equal chance of being true. And he is not really critical because, if he does have a system of beliefs, this system sets for him a particular perspective of critical revisability which is not the same as the perspective of critical revisability open to other rational persons committed to a different system of beliefs. The veil of ignorance model of critical revisability suggests, however, that the objectivity we can reach by distancing ourselves from our present beliefs is an impersonal view from nowhere.

But if objectivity does not imply an impersonal view from nowhere, then the claim that rationality as such uniquely requires a strong preference for the primary goods Rawls enumerates cannot be maintained. Suppose a person is convinced, in the mode of the old conservatives, that the only reason which privileges one form of life against another is that the first is embedded in an already existing tradition, while the second is only a mental blueprint. Let this person believe also that too much public scrutiny might subvert the existing tradition and lead to something resembling ethical anarchy. It is rational, even for such an individual, to entertain the fallibility of his beliefs and, therefore, leave them open to critical examination. But it is not rational for him to want as much freedom of speech as possible, as a uniquely optimal condition for keeping open the possibility of revising his particular conception of the good. Instead, he will maintain, in the spirit of Leo Strauss, that censorship is necessary to avoid general value skepticism. For an old conservative, the appropriate way to preserve the possibility of critical thinking is not to secure freedom of speech to the general public, but to grant freedom of thought to the wise, so long as they are prudent enough to restrict themselves to an esoteric discourse.³⁷

Or consider a person committed to the idea that, in view of the enormous misery in the world, the only life worth living is that of an ascetic, who works very hard and gives away everything he earns except what is essential for bare subsistence. Suppose this person allows for the possibility that at some time in the future—perhaps at an age when his productive capacities are exhausted—he might be caused, by

³⁷ See L. Strauss, "Liberal Education and Responsibility," in Strauss, *Liberalism Ancient and Modern* (Ithaca and London: Basic Books, 1968), 9-25.

reasons not available to him now, to convert to hedonism. Buchanan's argument would make him want to set aside some of the income and wealth he earns in order to cover his possible hedonistic desires. But this would make him incapable of abiding by his present altruistic convictions as closely as he should. For such a person, the choice between saving a part of his income for his possible future desires or giving away everything does not reflect the rational intertemporal distribution of resources, which is dictated by the uncertainty of his future beliefs and wants, but is expressive of a profound ethical decision. Either he has no reason not to hold his present convictions as true, and must therefore condemn some hedonist turn in the future, as well as reject the idea of setting aside sufficient resources for later desires; or this suggestion is sound, and so his present commitments should be immediately abandoned, because they are not supported strongly enough against reasons of prudence. If he persists in his present beliefs, then the thought of the possibility of him changing his views in the future must bring him feelings of regret and shame.³⁸ And if he does not persist in his present beliefs, then the decision to save and accumulate is not based on prudence and the refusal to adjudicate between actual and possible beliefs, but on taking sides against the beliefs held up to that moment. In sum, Buchanan's defense of the claim that the Rawlsian primary goods are neutral with respect to competing conceptions of the good does not withstand close scrutiny.

Another argument is suggested by Will Kymlicka. In an article which addresses the idea of primary goods as being biased in favor of individualist, and against communalist, conceptions of the good life,³⁹ Kymlicka makes the following remark:

[W]hile it is true that Rawls's theory makes the costs of a particular choice dependent on the extent to which other people's aims coincide or conflict, that does not show that the primary goods scheme is biased against communal ways of life. For the extent to which other people share one's end will depend on the judgments the others freely make when considering the various ways of life available to them. If socialists are unable to convince others of the worth of that way of life, then it will be difficult to acquire the resources necessary to start up a socialist community.⁴⁰

³⁸ Recall the case of Derek Parfit's nineteenth century Russian nobleman who, in several years' time, will inherit vast estates. As a young idealist, he presently intends to give the land to the peasants. But he knows that, by the time he will inherit the estate, he will have become conservative and will probably refuse to give the family estate away. Because he disapproves of this foreseeable change in his beliefs and attitudes, he signs an advance directive to the effect that the land should be distributed among the peasants, and makes this document revocable only with his wife's consent. See D. Parfit, "Later Selves and Moral Principles," in A. Montefiore (ed.), *Philosophy and Personal Relations* (London: Routledge, 1973), 145.

³⁹ W. Kymlicka, "Liberal Individualism and Liberal Neutrality," *Ethics*, 99 (1989), 883-905.

⁴⁰ *Ibid.*, 890.

If, under an individualist order, enough people can be convinced to cooperate in communalist ways of life, then communities will flourish. If, however, there are insufficient arguments and other incentives to induce large enough shifts in cooperation propensities, then communalism would be too costly in terms of the advantages other people would have to give up in order to join the proposed scheme of cooperation. Therefore, those persons whose conception of the good is dependent for its realization on others bearing these costs cannot justifiably claim that their ends should be given equal consideration through the distribution of primary goods.

Some may object that, perhaps, communalists do not fare ill not because the majority prefer individualistic arrangements to communalistic ones, irrespective of what the initial conditions may be, but because individualism is already instituted in the first place. So Kymlicka's reasoning is incomplete. But it can be completed. A further argument can be offered to show that the uncoerced evolution of the social order would converge, whatever the original structure, upon the same individualistic institutions. Suppose that, where the original system is communalistic, free individual choices progressively transform its institutions into an individualistic order (perhaps because anonymous market exchanges are more efficient than those regulated by the reciprocal claims of named individuals, or because private property secures more solid expectations than overlapping ownership, or for some other reason). In this case, the direction of social change would be path-independent, and the objection against Kymlicka has no bite. Communalists have no reasonable complaint against the structure of primary goods as prescribed by the individualist regime, because this structure would be preserved by uncoerced human interaction if it were instituted in the first instance, and also because uncoerced interaction would eventually converge in such a structure if the point of departure were a different structure. So regardless of which type of system comes first temporally, after a sufficiently long period of time society will reach an individualist system based on the predominance of private property, market economy, and voluntary associations; and hence it is better to start immediately with the individualist system in order to save the costs incurred by the transition from one regime to the other.

Yet these conclusions derive their plausibility from a tacit assumption which, as soon as it is made explicit, proves not to be plausible at all: namely, when in the first scenario communalists are induced by individualists to increase their participation in market exchanges, they decide to give up (first in part, then in total) the assets they acquired from communities and personal networks so as to share the more efficient institutional resources of the market; and individualists, in like manner, are invited in the second scenario to give up their institutional order for a communalistic system. Anthropology and history suggest a different account of the transition to market economy and

private property. I am not thinking here of the role coercion, fraud and violent expropriation have played in the victorious advance of the market in real-world history. These types of blatant injustices are ruled out by the assumption that basic rights and liberties are recognized and enforced. There remains a problem, nevertheless. When people loyal to personal networks and communities of belonging accept a market economy, they must not perceive the deal as a trade-off between commodity exchange and the reciprocity-regulated flow of goods and services.

Consider the following scenario. Participants in a system of reciprocities mobilize their leisure time in order to complement the set of goods already within their reach with other additional commodities. Nobody entertains, at the beginning, the prospect of withdrawing some of the resources at his disposal from the flow of reciprocal gifts in order to divert them to market exchanges. A new equilibrium is created where some economic actors are better off in terms of income (earned at the cost of reduced leisure, which they value less than the commodities acquired), and no actor is worse off in terms of the goods accessible to them through the chain of gifts. This equilibrium, however, proves to be unstable. As a reaction to the growth of income, the population also starts to grow, and its increase exerts pressure on the system of reciprocities such that the system begins to erode, so that the market remains the only workable device of economic co-ordination which can secure survival. In this model of transition, the switch to private property and the predominance of anonymous contracts is an unintended consequence of changes which were not foreseen, let alone chosen, by anybody. Or imagine another scenario. People who earned a surplus in year t withdraw a part of it to enter into advantageous market exchanges. They do this believing it is an exception: all adjustments to market exchange are performed under the expectation that others holding surpluses do not follow suit, or even if they do, their willingness and ability to be at the disposal of those who are currently suffering shortages will not diminish. In other words, the surplus-holders hope the reciprocity networks will remain intact, so that if, in the year $t + n$, they themselves incur a shortage, there will be sufficient opportunities for them to make claims on the resources of those who will hold surpluses at that time. Yet this expectation fails because it is erroneous to believe that the turn to the market will not be general. As a result of mass defection among surplus-holders, those with shortages are now reduced to seeking credit in the market in order to survive, and so they will not have surpluses in years to come to support the needy. The flow of reciprocal gifts collapses, and market exchanges penetrate every economic relationship without anybody having the *ex ante* aim to opt out of the reciprocity system. The capacity of the market to displace other institutional structures of coordination is due, in large measure, to the willingness of people to enter into market exchanges in the hope of being free-riders with respect to their

communal bonds; and free-riding subsequently gives rise to processes people do not foresee, let alone endorse, at the beginning.

A complementary remark suggests itself in relation to the second scenario. When members of an individualistic society refuse to join communalistic projects in large enough numbers, they do not face a choice between an individualistic and a communalistic order. The choice is whether to participate in communalistic projects within a larger society organized along individualistic lines. To reject the option of partial communalism does not necessarily entail rejecting communalism as an all-inclusive alternative to the individualistic order. The choice of partial communalism, one might suggest, would bring with it special costs which are largely absent in a totally communalistic society. For example, one could argue that the opportunities open to adherents of communalistic subgroups may be much more restricted if these, instead of being integral parts of a totally communalistic society, are enmeshed with an individualistic environment. Thus what the failure of communalistic overtures demonstrates is not that the choice of total communalism involves unacceptable costs to the majority, but that partial communalism involves such costs as these, and there is no way of inferring the first from the second.

The evolutionary advantages of private property and the market economy do not provide a good argument for their moral superiority. Responding to the expansionary potential of private property and the market, communalists might say, as they often do, that precisely because communalism is more worthy of human existence than individualism, permissible ownership rules should be constitutionally restricted so as to protect people against their own willingness to free-ride. I think the communalist claim is erroneous, but it is simply question-begging to argue that if individuals are free to choose, then the market and private property will win the day. So Kymlicka's defense of the neutrality of primary goods does not withstand criticism any better than Buchanan's.

Let us turn now to the third argument. This originates in Rawls's contention that justice depends on the expectations of the worst-off. The famous difference principle states that "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."⁴¹ From this it follows that, when comparing two alternative institutional arrangements, we have to compare the expectations of the worst-off group in the first with those of the worst-off group in the second. If the least advantaged members of an individualistic society are better off than the least advantaged members of a communalistic society (and *vice versa*), then it is rational to opt for the individualistic order (and *vice versa*). Thomas Scanlon's 'reasonable rejection test' of moral principles suggests the same procedure.

⁴¹ *Theory*, 75.

Scanlon offers this test as the key to his contractualist account of moral wrongness. "An act is wrong," he writes, "if its performance would be disallowed by any system of rules for the general regulation of behavior which no one could reasonably reject as a basis for informed, unforced general agreement."⁴² A principle can be reasonably rejected by (or on behalf of) a person if there is at least one alternative principle under which nobody has to bear as great a burden as this person has to bear under the principle in question. And, conversely, "it would be unreasonable ... to reject a principle because it imposed a burden on you when every alternative principle would impose much greater burdens on others."⁴³

A Theory of Justice proposes to identify the worst-off group in terms of income and wealth. So let us first make an estimate of the distributions of income the two systems would generate over time, and identify that group in each system whose members occupy the least advantaged positions in the respective income distribution. Next, let us select within the lowest income group of the individualistic society those members who hold a communalist conception of the good life and, within the lowest income group of the communalistic society, members who hold an individualist conception. These subgroups can be said to be worse-off than the rest because the system, besides placing their members in the lowest income positions, denies them that kind of social support they believe to be appropriate for securing a firm sense of their own worth. Suppose, for the sake of argument, that the income position of the two groups is the same. So if the incongruence between an individualist's understanding of the kind of social support necessary to reach an appropriate form of self-respect and the kind of support a communalist society offers him imposes a greater burden than the parallel incongruence experienced by the communalist in an individualistic order (or *vice versa*), then individualists have a reasonable complaint against the communalistic order, and communalists have no such complaint against the individualistic order (or *vice versa*). We have, thus, a criterion for the ordering of positions in terms of self-respect, a criterion which is neutral in the sense of being insensitive to the orderings suggested by the two competing conceptions.

But again, this solution is question-begging. To have a standard solution which can decide whose burden is greater, one must first independently determine which conception is valid. This is because, as already noted, the burden imposed on self-respect is an ethical issue, and not a psychological one. The question is not who has the more intense feeling of frustration, but whose frustration is appropriate or justified. If a person's sense of deprivation is based on false beliefs, then his claim to the resources society denies him is simply mistaken. In the light of communalism, the individualist's complaint against the

⁴² T. M. Scanlon, "Contractualism and Utilitarianism," in A. Sen and B. Williams (eds.), *Utilitarianism and Beyond*, 110.

⁴³ *Ibid.*, 111.

communalistic order is unreasonable. The same judgment is made on the basis of individualist conceptions about communities of belonging as sources of self-respect. To establish what a real burden is, and what the proper weights for measuring real burdens are in relation to each other, we need a conception capable of deciding the individualist/communalist contest.⁴⁴ Neutrality is certainly out of the question as a possible candidate.

V

All three arguments in favor of the Rawlsian primary goods fail, and they fail for parallel reasons. In order to show that primary goods are neutral with respect to conflicting conceptions of the good life, these conceptions must assume some further relationship of neutrality which proves not to hold. Buchanan assumes that an individual can be neutral about the different convictions he holds in succession through time. Kymlicka assumes that uncoerced social evolution is path-neutral, in the sense that whatever the initial structure of primary goods, the choices between the alternative arrangements will bring about the predominance of a structure that is adequate to the Rawlsian list. Rawls and Scanlon assume that the burdens defining the worst-off in two different institutional arrangements can be established neutrally. If the defense of the neutrality of primary goods claim must be based on further assumptions of some neutral relationship, and if this additional neutrality is no more likely to hold than the primary goods themselves, then it is highly unlikely that the neutrality of primary goods will withstand criticism. So I think we have good reasons to abandon it.

But if the neutrality assumption is dropped, we then face a dilemma. Either we can drop the assumption of the veil of ignorance as well, and so the very idea of justice as fairness will be compromised; or we can preserve the veil of ignorance, and so the hypothetical choice will not yield any determinate results. Consider the second horn of the dilemma: the veil of ignorance is not abandoned. In this case, the hypothetical choosers will know that their preference for the primary goods is *not* neutral in relation to other preferences they might hold. But the knowledge they may have about these other preferences will be suppressed. As a consequence, the choosers cannot know when (if ever) it is to their advantage to seek the maximization of the primary goods they control. They are prevented from making any rational decisions. Consider now the first horn of the dilemma: the veil of ignorance is abandoned. In this case, the hypothetical contractors will have reasonably detailed knowledge about their particular (and morally arbitrary) characteristics. But this means that they cease to be interchangeable. The same argument will not do for all of them. The process of reaching an

⁴⁴ Dworkin makes this objection to Scanlon's 'reasonable rejection test'. See R. Dworkin, "Foundations of Liberal Equality," *The Tanner Lectures on Human Value IX* (Salt Lake City, 1990), 28ff.

agreement will be transformed into a bargaining deal, and the agreement itself will consist in a compromise. And, worse still, its fairness will become questionable. Each individual will bring to the bargaining table his or her resources, both personal and impersonal, and the arbitrary distribution of the bargaining resources will have a decisive impact on the agreement. The contract, established in this manner, will not confer moral authority on the principles agreed upon.⁴⁵

Is there a way out of the dilemma? One possible way which suggests itself is to look for a better list of primary goods that would prove to be really neutral in relation to the preferences the veil of ignorance conceals. However, this type of attempt to overcome the dilemma is not particularly promising, because criticism of the neutrality assumption rests not on specific features of the Rawlsian primary goods, but on general properties rational individuals form, revise and pursue about conceptions of the good.

Instead of changing the list of primary goods, another way of solving the dilemma could involve not allowing for any specific information about their diversity. In other words, an even more thickly woven veil of ignorance might, perhaps, patch up the neutrality assumption's shortcomings. Suppose we abandon any specification of primary goods and define the resource which a rational person is supposed to want as something "infinitely divisible and malleable, capable of transformation into any physical object a person may desire," such as Bruce Ackerman's manna.⁴⁶ Once we stipulate this unique resource, the neutrality problem is by definition resolved, because manna is supposed to fit equally well the pursuit of any imaginable conception of the good. But this solution raises at least as many problems as it solves. First, in ordinary life, individuals do not hold their resources in the form of one homogenous asset. They each control bundles of different kinds of goods, which are neither infinitely malleable, nor infinitely divisible, nor capable of unlimited transformation into each other. So if the distribution of resources is to be equal (unless inequality is justified by its positive impact on the well-being of the worst-off), we still have to tell which particular distributions satisfy this requirement of equality. The assumption of manna, as a perfectly abstract and homogenous resource, does not help in solving this further problem. If there is no way to find an independent, and therefore neutral, metric for equality, then the results reached by suppressing even that minimal information Rawls still wanted to allow would be of no interest. If, however, such a metric

⁴⁵ See J. Buchanan, *The Limits of Liberty: Between Anarchy and Leviathan* (Chicago: University of Chicago Press, 1975).

⁴⁶ See B. Ackerman, *Social Justice in a Liberal State* (New Haven: Yale University Press, 1980), 31. The only information Ackerman permits is that manna is a scarce resource: there won't be enough of it to satisfy the total demands of society. In other respects, however, he does not want to limit the information available to the parties engaged in the discussion about the principle of justice; he thinks it sufficient to impose conversational constraints upon arguments which are held publicly.

can be constructed, then the imposition of further informational constraints is superfluous. Secondly, and even more damagingly, the assumption of manna excludes the lexicographical priority of basic liberties over other values. Suppose, for example, that there is one single primary good, manna. It will be uniquely rational for any individual, behind the veil of ignorance, to consent to that set of principles under which the amount of manna available to him is maximal. Lexicographic ordering of the particular forms into which manna can be transformed is, by definition, incompatible with this norm. To rank a particular value (e.g. freedom of speech) as lexicographically prior to another value (e.g. financial wealth) means that the more highly ranked value can never be traded for a lower one, not even in the case when the trade-off would secure more manna than its alternative. The maximizing of manna, however, requires choosing precisely that alternative which is conducive to procuring the largest possible amount of this uniquely primary good. Thus either the hypothetical choosers have no other good to pursue than manna, and so the theory does not yield the priority of basic liberties over any other good, and the priority of fair equality of opportunity over the difference principle; or the claim of lexicographic ordering is preserved, and so the theory does not yield any determinate principles at all (because there is no way for the hypothetical contractors to make rational choices between more specified and diverse primary goods). In other words, there is no way for the hypothetical contract to at once maximize one single resource and to yield a distinctly liberal theory of justice—and there is no way for it to maximize a diversity of specified primary goods and to yield a determinate theory of justice, liberal or not.

There is, of course, a third way to salvage the device of primary goods. Rawls contends, we should remember, that we should not make any restrictive assumption concerning conceptions of the good people hold in their ordinary lives. But, as I tried to show, his stipulation that it is always rational, behind the veil of ignorance, to maximize one's expectations of primary goods implies a strong postulate about real-world conceptions of the good: namely, people are not supposed to hold any preference in their ordinary lives which might conflict with the desire to have more rather than less of (all or some) the primary goods, and which, in cases of conflict, would override that preference. One could try to make this implied assumption explicit, and thereby marshal empirical evidence in its support. In fact, this is the strategy Rawls adopts in his later work. He starts from the observation that modern societies coordinated by the institutions of constitutional democracy are pluralistic insofar as their members hold conflicting ethical, religious, epistemic and metaphysical conceptions.⁴⁷ At any given point in time, some of these conceptions encounter disagreements which cannot, in the short run, be rationally resolved. Should the justification of the principles of justice depend on such controversial ideas, we

⁴⁷ See *Political Liberalism*, 8f., 36, 58, 64, etc.

would be compelled to recognize that no rational consensus on such principles is possible (in a pluralistic society). This problem is more general than the one we identified about primary goods, but the latter clearly falls within the scope of the former.

Rawls makes three postulates of a factual nature in order to overcome this difficulty. First, he asserts that principles of justice are latent in the political institutions of a society, and denote those ideas which explain and justify these institutions. These principles are, therefore, part of a *political conception* which can be constructed through the method of reflective equilibrium, as applied to a limited set of moral intuitions generated by the workings of the institutional order. They are implicit in the public culture of that order as 'freestanding views', i.e. as views which do not presuppose ethical, religious or philosophical ideas of a more comprehensive sort for their justification.⁴⁸ Call this the postulate of *political* intuitions. Secondly, Rawls contends that the institutions of constitutional democracy are capable of generating the consensus necessary for a shared political conception to emerge. That is to say, he maintains that their workings bring the overwhelming majority of the citizenry to roughly the same political intuitions.⁴⁹ Call this the assumption of *shared* political intuitions. Yet a problem remains. A political conception of justice, worked out on the basis of shared political intuitions, might conflict with the ethical, religious or philosophical views some people hold. The requirement of consistency demands that such conflicts be resolved, and so we are back to disagreements between what Rawls calls *comprehensive conceptions*. To settle this problem he makes a third factual postulate. He states that constitutional democracy is capable of eliminating possible conflicts between the political conception justifying it and the comprehensive conceptions its citizens hold. Constitutional democracy does this, according to him, by recourse to the consistency requirement. Rational individuals must seek consistency among their beliefs, and this search presses for change in their ideas. Political convictions, however, are particularly resistant to change, or at least with respect to that type of change which is not supported by the pressure of the institutional order. As a result, the other views must yield. In the long run, so the assumption goes, constitutional democracy eliminates those comprehensive views which conflict with the political ideas which are latent in its institutional order and, by virtue of this, restricts the scope of the comprehensive conceptions held by its citizens to those which are compatible with its political conception of justice. In other words, only those comprehensive conceptions which are consistent with this particular conception of justice can survive the selective pressure of the workings of constitutional democracy. And so, the pluralism of comprehensive conceptions tends to become what Rawls calls *reasonable pluralism*, a pluralism of views which, though they might

⁴⁸ *Ibid.*, 10f., 97.

⁴⁹ *Ibid.*, 158f., 164f.

irresolvably conflict with each other, nevertheless concur about political ideas of freedom, equality and toleration.⁵⁰ He calls this the fact of an *overlapping consensus*. Once this fact is created, the political conception of justice can enjoy a secondary reinforcement from various comprehensive conceptions: after making an agreement, on the basis of their shared political intuitions regarding common principles of justice, the citizens of a constitutional democracy can each arrive at the same result by starting from his particular comprehensive ethical, religious or philosophical views.⁵¹ Let us label this the postulate of the overlapping consensus.

The third factual postulate Rawls makes is of particular relevance for us. Clearly, the stipulation that it is rational for the hypothetical choosers to maximize their expectations of primary goods, whatever else they may desire, falls within the scope of this postulate. Given this, Rawls maintains, in his later work, his commitment to the idea of primary goods, and we can see these as resources which, together with their ordering, are endorsed by constitutional democracy—as belonging to its political conception. And the particular conceptions of the good life people hold can be seen as belonging to the field of their comprehensive views. The third postulate secures the compatibility between the two realms of ideas, which entails, for our specific aims, the requisite limitation of the scope of preferences the citizens of a constitutional democracy hold.⁵² If this postulate is warranted empirically, then our problem is resolved.

But is this a satisfactory solution? One can attack it in two different ways. One way is to show that the empirical assumptions Rawls makes are false, or are at least implausible. For example, we can challenge the claim that the political views supported by the institutional order are always more resistant to change than ethical or religious convictions, which are incompatible with them. Another type of objection is the following. Suppose all the empirical claims made by the later Rawls are true. But even if we concede this much, the reliance on them makes the theory of justice dependent, for its validity, on contingent factual evidence; and this dependence alters dramatically the theory's status. *Theory* claimed to offer universal standards of justice against which existing social systems can be assessed and ranked:

[T]he two principles of justice are not contingent upon existing desires or present social conditions. Thus we are able to derive a conception of a just basic structure, and an ideal of the person compatible with it, that can serve as a standard for appraising institutions and for guiding the overall direction of social change.⁵³

⁵⁰ *Ibid.*, 15, 39f., 58f., 78.

⁵¹ *Ibid.*, 170f.

⁵² *Ibid.*, 75f., 187f.

⁵³ *Theory*, 263.

In *Political Liberalism*, Rawls explicitly abandons this claim. There, he describes justice as fairness as a theory “addressed [...] to citizens in a constitutional regime,”⁵⁴ and not to any person willing to use his rational powers. Such a theory cannot be the standard bearer for appraising alternative institutional orders. It cannot, nor should it, claim to assess the moral appropriateness of constitutional democracy against rival political and social systems, nor can it make a case for the superiority of constitutional democracy. It must reduce its trajectory to that of settling disputes between citizens in a constitutional regime, between citizens who share the same political intuitions of freedom and equality and hold reasonable comprehensive conceptions. This is too high a price to pay for the solution Rawls offers to the problem of pluralism, as many critics have argued.

The argument I have presented in this article was based on the claim that the veil of ignorance device requires, but cannot secure, the neutrality of primary goods. But what is wrong with the neutrality requirement? This is an important question because, after *Theory*, many liberal thinkers identified neutrality as one of the foundational values that allow us to flesh out and interpret the root conviction which defines liberal political theory. Critics argue that neutrality is an incoherent principle, and therefore that either liberalism itself must be abandoned, or that it must be reconstructed in such a way as to be rid of the dubious idea of neutrality. I think neutrality, when introduced as a foundational principle, is indeed incoherent, but that incoherence is not a necessary feature of a more modest neutrality principle which enters the political argument at a later stage, when much is already settled by the operation of other, more basic principles. The intuition that liberalism needs a concept of neutrality—which is approximate to, but not identical with the concept of toleration—as part of the normative constraints a liberal state is required to abide by, is sound, or so I believe. This, however, would be the subject of another paper.

⁵⁴ *Political Liberalism*, 369.