

## INDETERMINACY IN RAWLS' ARGUMENT FOR THE DIFFERENCE PRINCIPLE

The contractarian theory elaborated by John Rawls in *A Theory of Justice* exploits the difference principle in a great many ways. Rawls argues that, when used as part of a set of guiding principles for structuring the basic institutions of society, it simplifies the problem of interpersonal comparisons (91-4)<sup>1</sup>, helps compensate for the arbitrariness of natural endowments (101-3), promotes a harmony of interests between citizens (104-5), reintroduces the principle of fraternity to democratic society (105-6), and, what is critical to his contractarian theory, it is an essential part of the principles of justice which would be chosen by free, equal, and rational persons in the original position.

Given the complexity of Rawls' contractarian argument, it is evident that his argument for the difference principle may be criticized in a number of ways, and this fact is reflected in the literature. Some, for instance, have attacked Rawls for incorporating features in his original position which seem designed to favor his principles over their rivals and which have little else to recommend them.<sup>2</sup> Others have argued that, given some plausible feature of human nature or economics which Rawls did not include in his choice situation, parties in the original position would have chosen differently.<sup>3</sup> Still others have complained that the difference principle leads to unacceptable consequences.<sup>4</sup> Such lines of criticism attack the acceptability of either Rawls' premises (the conditions of the original position) or his conclusion (the principles which Rawls claims parties in the original position would unanimously choose). My project here is to take a somewhat different tack: that Rawls' own premises provide grounds for serious doubt about what

---

<sup>1</sup> Parenthetical page references in the body of the text are to Rawls (1971).

<sup>2</sup> E.g., Nagel (1989), p. 11-12, and Hare (1989), pp. 95ff.

<sup>3</sup> E.g., Barber (1989), pp. 296-298, Miller (1989), p. 214, and Brandt (1979), p. 239.

<sup>4</sup> Arrow (1973), p. 251; Griffin (1986), pp. 170-1.

the parties would choose.

I will develop this idea along the following line. After briefly summarizing Rawls' general argument and his argument for the difference principle, I will argue along with Kavka, Nagel, and others that there are internal flaws in Rawls' argument for the adoption of the difference principle, particularly in the argument for the maximin rule which Rawls exploits to choose his principles of justice. These flaws make the choice of Rawls' two principles of justice more doubtful than Rawls makes it appear; and given Rawls' specification of the original position and the parties in it, a different, mixed conception might well be preferable. However the proof of even this is beyond the scope of this paper, as I will try to make clear.

## **II. RAWLS' ARGUMENT FOR THE DIFFERENCE PRINCIPLE**

As noted above, Rawls believes the difference principle has a number of important virtues. But in a contract theory the point isn't whether the proposed principles are appealing; it is whether they would be preferred by agents with stipulated qualities under the conditions specified for their choice, i.e., by the parties in the original position. The main points of Rawls' argument that his principles of justice are "the unique solution to the problem set by the original position" (119) are as follows.

In the original position (OP), the parties are to choose from a list of alternative principles which comprises the classical and average principles of utility, a variety of intuitionistic conceptions, and some hybrids, called "mixed conceptions" (124). The chosen principles will order the basic structure of the society in which they will all live. The agreement will be final, in two important senses. It is a feature of the OP that the agreement is made in perpetuity; there is no going back (147, 176). And once the agreement is made, there will be no recourse to higher principles for settling matters (135). Whatever is agreed to will be "the final court of appeal in practical reasoning" for keeps.

The veil of ignorance behind which the parties are situated is designed to prevent arbitrary contingencies from biasing the outcome (141). Behind it, no one has any knowledge of

their place in society, natural endowments, or even what their individual conception of the good is. While they “are presumed to know whatever general facts affect the choice of the principles of justice” (137), they do not know the economic or political condition of their society, or its level of civilization or culture (137, 155). The OP’s constraints thus create a condition of equality for all parties. There is no opportunity for bargaining, no leverage for threats, no way to gain advantage by adopting principles favorable to oneself, one’s friends, the place in society one occupies, or one’s life plans (137, 139-40).

Each party is assumed to be interested in advancing their conception of the good, and it is assumed that the primary social goods of rights and liberties, opportunities and powers, and income and wealth<sup>5</sup> will suffice for whatever purposes one turns out to have in life (92, 142).

Once the parties are assembled behind the veil, Rawls notes that the obvious first step in the decision process is to agree on complete equality in the distribution of all social goods (150-1). But given that allowing inequalities in the basic structure could create incentives which boost productivity, it is possible that all parties could be better off than they would be at the benchmark of equality if some kind of inequality were permitted.

Given certain circumstances present in the OP (which will be examined below), Rawls argues that the parties ought to employ the maximin rule for choosing among the alternatives, according to which one ranks the available options according to their worst possible outcomes, choosing the “best worst” of them (152-6).

According to Rawls, since the alternative conceptions on the list possibly allow for worse outcomes (156), the rational choice would be his principles of justice (302; hereafter abbreviated RPJ):

1. 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.
2. Social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged [as determined from the standpoint of a

---

<sup>5</sup> Following Rawls, for simplicity I omit here the primary good of self-respect.

representative of that group; this is the difference principle], consistent with the just savings principle, and (b) attached to offices and positions open to all under conditions of fair equality of opportunity.

Under this conception, the proceeds of the increased productivity encouraged by inequality are harnessed for the good of the less advantaged, leaving them better off than they would have been under complete equality.

### III. ADOPTING THE MAXIMIN RULE

My line of criticism requires a closer examination of the steps Rawls takes in his argument for adopting the maximin rule for choosing among the alternative principles on the list.

There is an ambiguity in Rawls' text we should attend to at the outset. Rawls does not explicitly appeal to the maximin rule in his arguments against mixed conceptions. He only appeals to it in his comparison of RPJ and utilitarianism (§ 26-28; see 1st sentence of § 26, p. 150), with the exception of p. 175, where he appeals to it as part of an "heuristic schema." Thus it might be inferred that he only intended it be used in the comparison of RPJ and utilitarianism. But Rawls states (155) that "the original position has been defined so that it is a situation in which the maximin rule applies"—because (153) it "manifests these features to the fullest possible degree, carrying them to the limit, so to speak"—without qualification. I take this to mean it applies to the original position as a whole, including the full list of alternatives (cf. 124, 146). Hence, I will assume the maximin rule is Rawls' guiding rule regardless of which comparison is being made. Later in the paper, though, this will make little difference.

Rawls does not recommend maximin as a general rule for choosing among alternatives. Given certain kinds of information (such as the probabilities associated with the alternatives one faces), one could often do better for oneself by using some other rule, as Rawls knows.<sup>6</sup> Rather, he claims that adopting the maximin rule is rational when three features obtain in a choice

---

<sup>6</sup> Rawls (1974b), p. 649; also Rawls (1971), pp. 153, 172. Rawls' cited source on the maximin rule, Fellner (1965), p. 142, is emphatic on this point; see note below.

situation (153-6):<sup>7</sup>

- F1: There must be some reason for sharply discounting estimates of the likelihoods of the possible circumstances.
- F2: The person choosing has a conception of the good such that he cares very little, if anything, for what he might gain above the minimum stipend that he can, in fact, be sure of by following the maximin rule.
- F3: The rejected alternatives have outcomes one can hardly accept.

When these conditions obtain, Rawls says, one should apply the maximin rule, ranking the alternatives according to their worst possible eventualities, and choose the best worst. The conception of justice chosen from the list, then, will assure the highest minimum index of primary goods, and the gains above this minimum offered by alternative conceptions will be of little interest.<sup>8</sup>

The next step for Rawls is to show that the OP is such that the maximin rule applies there (155-6, 175ff). F1 is present by virtue of the veil of ignorance. Behind it, the parties do not know who they are, what place in society they occupy, or how many places there may be to occupy; they likewise do not know what indexes of primary goods might be associated with any social position, or how much total wealth the society has or what form it takes. Given that they

---

<sup>7</sup> Cf. Fellner (1965), p. 142. Rawls (p. 154n) cites Fellner as his source, but where Rawls lists three circumstances Fellner only gives two, neither of which precisely matches any of Rawls', and only one of which seems (on Fellner's view) directly pertinent to Rawls' choice situation. Fellner states that maximin may be the rule of choice for

“one of two reasons. One of these is that even when an attempt is made to measure the utility of wealth to the decision maker in a standard process, it may turn out that the utility of the goods he can obtain for the first units...of money gain [i.e., the best worst available gain] is infinitely greater than the further utility embodied in the goods obtainable for additional units of money gain; the second possible reason is that in the particular process to which Matrix 1 [a matrix similar to Rawls' gain-and-loss table on p. 153] relates, he slants downward to an extreme extent the subjective probabilities of all events which would yield him more than the minimum guaranteed pay-off. Only the second of these two reasons can be used for justifying the maximin-of-gain rule if Matrix 1 is interpreted as expressing pay-offs in *utility* (rather than in money), as is sometimes the case.” (Fellner's italics.)

<sup>8</sup> There is another ambiguity in Rawls' text which could have significant implications. At some points, Rawls writes that the basis of expectations under utilitarianism is not primary goods, but “total satisfactions enjoyed and foreseen” (p. 162; also cf. pp. 174-5). This could be taken to mean that the parties in the original position are to weigh the difference principle according to its prospects of primary goods and utilitarianism according to its expectations of satisfaction. But in his specification of the original position, Rawls states that the parties accept primary goods as the basis of choice (p. 155). I will therefore assume that my statement in the text above is correct.

also do not know what their individual conceptions of the good are, there is no way to ascertain the best way to further it. Hence there is no basis (aside from what might be obtained from general knowledge, which would not be much) for employing probabilistic reasoning in the OP.

Rawls argues that F2 is present because, since RPJ “provide a workable theory of social justice, and...they are compatible with reasonable demands of efficiency...this conception guarantees a satisfactory minimum.” (156). Given the value Rawls feels the parties would attach to liberty over other primary goods, they would not wish to try for gains in other goods at the expense of liberty (156, 176-7, 542-3), and “any further advantages that might be won by the principle of utility, or whatever, are highly problematical, whereas the hardship if things turn out badly are [sic] intolerable” (175). Further, it should be noted that the contract is assumed to be made under conditions of moderate scarcity, where “[n]atural and other resources are not so abundant that schemes of cooperation become superfluous, nor are conditions so harsh that fruitful ventures must inevitably break down” (127). Thus the parties know that their society will not, as a whole, be desperately poor.

According to Rawls, F3 is present because the alternative conceptions on the list present the possibility of far worse outcomes than RPJ. For example, the principle of average utility allows, at least theoretically, the possibility that one’s welfare may be sacrificed for the good of others, perhaps to an extreme degree.<sup>9</sup> Considering the finality of this decision and its consequences for the parties and their descendants, this would render such options unacceptable.

The presence of all three features in the OP would lead to a state of mind we could describe as highly risk-averse,<sup>10</sup> in which it is appropriate to employ the maximin rule, which Rawls contends would select his two principles of justice.

This argument for the adoption of the maximin rule has provoked a good deal of criticism, to which I intend to add my voice. While critics (notably utilitarians) have found fault

---

<sup>9</sup> This point is made more succinctly in Rawls (1974b), pp. 640-1 than in *A Theory of Justice*. Cf. p. 156 of Rawls (1971).

<sup>10</sup> Rawls (1974b), p. 649.

with Rawls on all three of the above claims concerning the OP, I wish to focus here on the claims that the OP has F2, the assurance of an acceptable minimum above which the parties would have little interest in aspiring, and F3, that the alternatives present possibly intolerable prospects.

Rawls' claim that "the original position manifests these features to the fullest possible degree..."

(p.153) rests on these contentions concerning the OP:

- C1: RPJ provides a satisfactory minimum, by virtue of providing a workable and efficient theory of social justice and by guaranteeing the priority of liberty. (Supports F2.)
- C2: The alternative conceptions present the possibility of intolerable outcomes. (Supports F3.)
- C3: The possible benefits offered by alternative conceptions are highly problematical. (Supports F2.)

Rawls, however, himself gives us reasons to believe that all of these contentions are false or at least highly questionable.

#### **IV. C1: SATISFACTORY MINIMUM**

If Rawls has indeed established that his principles provide a viable theory of social justice (a question I will not take up here), that is an important achievement. But it is insufficient to guarantee a satisfactory outcome to the parties in the OP (POPs).

One reason why this is so, raised by Kavka,<sup>11</sup> regards Rawls' proposed standard for measuring the welfare of the worst-off group, namely the point of view of a representative of that group. This point of view is vaguely defined, in two ways. First, the definition of the worst-off group is left open: it is perhaps a chosen social position, such as unskilled worker, or perhaps the group of all persons with less than half the median income and wealth (98); in either case, there is a significant spread of possible indexes of primary goods within the worst-off group. Second, the standing of the representative of that group is not defined at all. Rawls admits that this element of his theory is somewhat ad hoc, but that is not really the problem. It is rather that all this vagueness leaves us with a dilemma. For if, as Kavka points out, the representative is

---

<sup>11</sup> Kavka (1986), p. 197.

defined as an average member of the group, there will by definition be persons living below this level, i.e., possibly below an acceptable level even if the “minimum” defined by RPJ is acceptable. Each party in the OP knows that he might be one of them. But if, on the other hand, one attempts to avoid this problem by defining the representative as the lowest of the low, the difference principle would then dictate that inequalities could only be allowed when the prospects of the very worst-off persons in society—e.g., the severely physically and mentally handicapped—are improved. Considering the enormous cost of improving their prospects even a little, this would be unreasonable: it could cripple the rest of society.<sup>12</sup>

A point could be raised against this objection, I think. If expectations are counted in terms of one’s index of primary goods, then wouldn’t the prospects of the severely crippled or vegetative be improved simply by depositing money in their bank accounts or putting them in prettier hospital rooms? If so, then the cost to society of improving their prospects would not necessarily be prohibitive. According to this idea, society might not be obligated to improve conditions for the disabled or ailing by expending resources specifically to overcome their disabilities or cure their ailments. Instead, society could discharge its obligation to them by improving their conditions in other ways, as suggested above, which would presumably cost far less.

But at least three considerations weigh against this idea. First, if Rawls adopted it to get around the objection to his “representative” definition, it would be at the cost of something obviously important to him, namely the idea that organizing society according to the difference principle will really make things better for everybody in a way that would make us all happier. Second, this idea may not count as a solution if we calculate improvements in terms of long-term plans of life as seen from the representative’s point of view, as Rawls intends (92-3, 129). If we suppose, quite reasonably, that the disabled and ailing would most want to be cured or at least to have their conditions ameliorated enough for them to lead productive, satisfying lives, this

---

<sup>12</sup> Kavka (1986), pp. 187, 210. Arrow (1973), p. 251, makes a similar point.



solution might be disallowed.<sup>13</sup> Third, while such measures may improve the lot of the very worst off, it is unclear how their prospects could be maximized by these means, as the difference principle requires.

Rawls avoids the problem of the very worst off by stipulating it away: he assumes that everyone has “normal physical needs so that the problem of special health care does not arise.”<sup>14</sup> This spares him from having to contend with the second problem raised by Kavka.

But we are still left, though, with the problem of breadth of positions. If Rawls wants us to believe that even with this breadth the POPs could assure themselves of a satisfactory outcome, he would have to show that the minimum to be had by the representative of the worst-off group is high enough that even those somewhat below it would be well off. But this he cannot do, however narrow the spread of indexes of primary goods within the group.

Rawls defines the primary social goods which form the parties’ basis of expectations as things we would want more of rather than less; rights and liberties, opportunities and powers, and income and wealth are assumed to be the means of pursuing whatever their conception of the good is (92, 95, 128; 142, 143). Further, Rawls assumes that each party will “try to advance their conception of the good as best they can, and...in attempting to do this they are not bound by prior moral ties to each other” (128). And behind the veil of ignorance, the parties won’t know what their conception of the good will be (137)—or, hence, how expensive it will be to attain and maintain. Given that the veil also conceals any knowledge of their society’s stage of development or available resources, they have no way of knowing, in absolute terms, what the guaranteed social minimum offered by Rawls’ principles (or any other conception of justice) will be.<sup>15</sup> All things considered, parties concerned to advance an unknown conception of the good would have to be aware of the possibility of a great shortfall between their wants in society and

---

<sup>13</sup> But given that Rawls rejects the idea of judging prospects in terms of expected satisfaction (p. 94), this is uncertain.

<sup>14</sup> Rawls (1974b), pp. 639-40; also cf. Rawls (1974a), p. 142.

<sup>15</sup> This last point is made by a number of critics, including Nagel (1989), p. 12, and Hare (1989), p. 105.

what society may offer them once the veil is lifted. Even if we assume, as Rawls does, that “the members of society are rational persons able to adjust their conceptions of the good to their situations” (94), this only applies to their ability to adjust their ends to the *kinds* of means available. It is too much to expect that they should also adjust to the *amount* of means available, no matter how small. Given the range of possibilities afforded by the assumption of moderate scarcity, the means could be quite meager. Thus, even though RPJ would guarantee liberty and a greater share of other primary goods than an egalitarian system would offer, there is nothing to show that this will provide enough riches that one would be nearly indifferent to having more, as F2 requires. Indeed, if Rawls believes each man’s preference would be for a first-person dictatorship (119), it is hard to see how he could think that anyone could ever be so happy with the outcome of a social contract, since compromise is inevitable.

It should not be forgotten that another part of C1 relates to the priority of liberty.<sup>16</sup> Rawls argues that “if the persons in the original position assume that their basic liberties can be effectively exercised, they will not exchange a lesser liberty for an improvement in their economic well-being...” (542). If all other values pale in comparison to liberty, then the guarantee of maximum equal liberty afforded by Rawls’ First Principle of justice would go a long way towards providing an acceptable minimum. But is liberty a high enough priority to assure an acceptable outcome almost single-handed? To believe this, we must imagine that one would be happy with an equal share of liberty but not much else beyond the subsistence one could be reasonably assured of under moderate scarcity. Again, given the parties’ desire to further an unknown conception of the good under perhaps meager circumstances, this is unrealistic.

It might also seem that giving liberty lexical priority, as Rawls does, is too strong. That

---

<sup>16</sup> At least that is how I interpret Rawls on p. 156, 1st open paragraph. Some of the literature treats Rawls as if the point of the original position is to choose the difference principle. But while the difference principle is the centerpiece of Rawls’ theory of distributive justice, it does not operate alone (cf. p. 158) but rather in conjunction with the other principles. Arrow (1973), p. 250, thinks that the maximin rule is applied *separately* to liberty and to “nonpriority goods”, which seems to imply that the minimum standard of living we are to weigh in the original position is *only* that supplied by the difference principle. If Arrow is right, Rawls’ argument for the difference principle is weaker than I am here construing it to be.

the parties should know they value liberty so highly seems to assume that they do have some conception of their good. Else, why would they know that liberty was the most important facilitator of it? If they truly had no idea what their conception of the good would be, why not endorse a more flexible set of principles which would allow them to trade off liberty for other primary goods if, after the veil is lifted, they find that this would better suit them? A possible (and I think persuasive) answer is that liberty is not the sort of thing that, once given up, can easily be gotten back if one changes one's mind. But Rawls' answer to this is to stipulate that the parties

regard themselves as having a higher-order interest in how their other interests, even fundamental ones, are regulated and shaped by social institutions. They think of themselves as beings who can choose and revise their final ends and who must preserve their liberty in these matters.<sup>17</sup>

Clearly the parties are assumed to have some conception of their good, then. But Rawls rests this assumption on a conception of human nature which runs deep in his theory, and which I do not wish to contest here. It suffices that we have established that C1 is not a feature of Rawls' OP.

But this does not mean, as Nagel apparently believes,<sup>18</sup> that we have disposed of F2. For if C3 holds, the parties have nothing better to hope for from other conceptions; hence F2 would remain intact even if the parties are not particularly happy about RPJ. And if they are not—if the parties know that RPJ may not provide them with satisfactory means for fulfilling their conceptions of the good—it is reasonable to suppose they would be inclined to choose an alternative if it promised them a prospect of more. But they might not do so if the alternative were too risky, i.e., if C2 holds.

## V. C2: INTOLERABLE ALTERNATIVES

The argument against C1 implies that none of the available alternatives on Rawls' list,

---

<sup>17</sup> Rawls (1974a), p. 143.

<sup>18</sup> Nagel (1989), p. 12.

including RPJ, can assure the POPs a satisfactory outcome. No one can guarantee that the parties will have the means to succeed at their plans of life whatever they turn out to be. If this meant that every alternative presented the possibility of intolerable outcomes, the parties would be unable to proceed, since they are constrained against entering into agreements that may have unacceptable consequences (176).

But I think we may draw a useful distinction between dissatisfaction—even severe dissatisfaction—and intolerability. And it seems to me the parties are in a position to assure themselves of a tolerable life. I would like to suggest an interpretation of the choice situation which shows how this can be so. The benefit of this will be to provide a standard which at least some conceptions can meet, but by which others may be ruled out.

Some factors influencing the availability of primary goods are out of the parties' control (as well as unknown behind the veil of ignorance). The availability of income and wealth is strongly influenced by a society's material resources and level of cultural and economic development, as is the number and kinds of social positions which the parties may occupy. A primitive agrarian society, for instance, has less wealth and power, and fewer positions than, say, the State of California. Although the parties can agree on social structures which have greater or lesser potential to affect these variables, they generally only change over generations; the parties cannot conjure them up for themselves by agreeing on this or that set of principles.

But rights and liberties are primary goods the parties can give themselves. And regardless of the number and kinds of available social positions, the parties can agree on a scheme controlling access to them. Likewise, while the total available wealth and income is out of their hands, the parties can agree on the system for its distribution. And these things are ultimately what the parties in Rawls' OP are trying to agree on. The assumption of moderate scarcity assures each person of a chance at avoiding destitution; the rest is up to the social structure they create. Thus the only kinds of disasters they are in a position to avert are self-caused: allowing one person or group so much of the available wealth that others are left homeless or starving; allowing access to social positions based on arbitrary, inefficient, or reprehensible criteria; or

allowing some the liberty to pursue their life plans while denying it to others, whose life plans are thereby crushed. A tolerable conception of justice, perhaps, is simply one which avoids such outcomes; it avoids hardships that are mainly the result of how the structure of society treats people. When judging whether a given conception of justice allows an intolerable outcome, then, we should restrict our consideration to the factors that are under the parties' control.

Given the veil of ignorance, the parties cannot know whether the principles they choose—and hence the social structure these create—will give them wealth; that largely depends on the circumstances of the society. However, the safety afforded by a workable, efficient (cf. 156) and fair conception of justice is extremely valuable. A guarantee of maximum equal liberty, fair equality of opportunity, and a distribution of the other primary goods such that the worst-off<sup>19</sup> benefit from whatever inequalities are allowed, does indeed

in large part sidestep the uncertainties of the original position. [If the parties choose RPJ they] can guarantee the protection of their liberties and a reasonably satisfactory standard of life as the conditions of their society permit. (169)

If no alternative to RPJ assures a tolerable outcome, C2 is true, and so is F3. And if F3 is true, applying the maximin rule yields a clear choice. Not necessarily a happy, abundant choice, but one with a very attractive level of safety.

The question now is, is RPJ uniquely tolerable?

Rawls points out that while it might be true that the principle of utility (classical or average) may provide an acceptable outcome for everyone under suitable circumstances, it does not *rule out* harsh sacrifices, such as deprivation of liberty or other primary goods, simply for the sake of others' welfare (156, 160-1, 169-71, 176ff)—i.e., hardships which are not a necessity imposed by fortune but rather by society itself. Since the parties behind the veil of ignorance cannot know whether favorable circumstances obtain, choosing the principle of utility as the sole governing principle presents a dire risk. A person in the OP could not tell until too late whether

---

<sup>19</sup> And everyone else, if we assume chain-connectedness and close-knitness, or adopt the lexical interpretation of the difference principle. Cf. Rawls (1971), pp. 80-83.

she would end up an empress, a clerk, or perhaps a slave. She could not know whether she would find herself in a society which would distribute whatever primary goods are available in a way she could accept regardless of her position. It must be emphasized that here “acceptable” does not mean “adequate to fulfill any plan of life she might have,” because on the view I am suggesting the parties are not in a position to assure themselves of this. Rather, “acceptable” means “in a way which does not leave her, from her point of view, unduly burdened, or cheated out of prospects made available to others, as a matter of social policy.”

Utilitarianism, then, fails the tolerability test. But of the alternatives on the list (124) is RPJ the only one which provides these protections? A glance at the list shows that we can choose a set of principles from the selection of mixed conceptions which provide most of (I will later contend perhaps more than) the guarantees of RPJ, namely:

1. The principle of greatest equal liberty.
2. The principle of average utility, constrained by
  - (a) that a certain social minimum be maintained, and by
  - (b) equality of fair opportunity.

Or, A1 + B2(a) + B3 from Rawls’ list. On this conception (which I will simply call the mix), the primary goods of rights, liberties, opportunities, and powers are apportioned as in RPJ, and (again assuming moderate scarcity) unlivably low levels of the other goods are averted by a social minimum. The only difference between the mix and RPJ is that the difference principle is replaced in the mix by a social minimum constraining the principle of average utility. To make the comparison as close as possible, I will assume lexical ordering: 1 over 2, 2(b) over 2(a), and 2(a) over the principle of average utility. So if RPJ is to win in this comparison, it will be on the strength of the difference principle alone, since all the other elements of RPJ are also in the mix.<sup>20</sup> Both alternatives allow the protections lauded by Rawls in the quote above.

It will be objected that the mix suffers from the same flaw pointed out earlier by Kavka.

---

<sup>20</sup> Rawls compared RPJ with mixed conceptions both in (1971) and (1974a), pp. 143-145, but not with the mix I have here created.

Wherever the social minimum is set, due to the breadth of social positions one can expect to possibly end up below the minimum, i.e., with an unacceptable outcome. I wish I could answer this with a scheme for establishing the minimum which would eliminate the difficulty altogether. But the best I can do in the way of meeting the objection head-on is to note that if the mix suffers from it so does RPJ; and if the parties cannot choose an outcome which might be unacceptable, then RPJ and the mix fall together into the dustbin, and so does every other conception until the dilemma outlined earlier can be solved.

We could try to salvage one of the options by employing the strategy of disaster avoidance outlined by Kavka, which maximizes the chances of avoiding all disastrous outcomes rather than trying to “play it safe”, in circumstances where there is no “safe” to play. The disaster avoidance rule applies when (a) all alternatives have disaster potential, and (b) the alternative with smallest disaster potential risks a disaster of greater magnitude than does some other alternative, and recommends choosing the alternative with the lowest probability of disaster.<sup>21</sup>

Now condition (a) applies to both RPJ and the mix by virtue of the breadth-of-positions problem. Condition (b), however, is impossible to meet. We may reasonably suppose that the mix represents a risk of greater magnitude. This is because the welfare of the people living at the mix’s social minimum—and hence that of the people below it—will be lower, other things being equal, than the worst-off group under the difference principle. I will explain this further in the next section, but suffice it to say that the reason for this is that under the mix the better off are enjoying the goods which the difference principle would have given the worst-off under RPJ. Given this, the disaster-avoidance rule would recommend the mix if, under it, the probability of disaster were lesser than under RPJ.

We cannot discern, however, whether the probability of disaster (i.e., what one’s potential is for ending up below the social minimum) is lesser under the mix. The veil of ignorance prevents the parties from knowing how probable it is they will be in any given social position

---

<sup>21</sup> Kavka (1986), p. 205.

once the veil is lifted. They are also denied any knowledge of their personal characteristics, and hence cannot form any idea of what position they may be able to attain in society, unlike Hobbesian contractors.<sup>22</sup> The POPs are completely in the dark and thus unable to apply the disaster-avoidance rule. And even if they could, the breadth-of-positions problem would still haunt them<sup>23</sup>

There is another solution which, while not entirely satisfactory, will at least allow us to proceed. This is to reiterate the idea developed at the beginning of this section that it is beyond the capability of a conception of justice to guarantee any particular standard of living. Its purpose is to assure an acceptable level of the goods we can control (e.g., liberty) and an acceptable distribution of the rest. And if we can again exploit the condition of moderate scarcity, we may say that no one will end up too badly off unless their chosen structure of society makes it so. Given moderate scarcity, then, an acceptable social minimum can be defined which will be livable, such as that suggested by Kavka:

...that level of economic support which would allow most all individuals (1) to live out their normal physiological life spans, save for the effects of accident, homicide, and unpreventable or incurable disease, and (2) to participate in the normal economic and social activities of members of the society.<sup>24</sup>

And while a person would tend to prefer the status of the worst-off group under the difference principle to the guaranteed social minimum under the mix, there are no grounds for calling the latter intolerable.

Thus the case for C2 is eliminated, and F3 along with it. Employing the maximin rule of choice, then, will not single out a uniquely tolerable alternative.

The question now is whether one of the tolerable alternatives is clearly superior. If, as

---

<sup>22</sup> Kavka (1986), p. 205.

<sup>23</sup> My thanks to Eric Cave for this point.

<sup>24</sup> Kavka (1986), p. 213-14. Of course this is a vague definition but it is no more vague than Rawls' definition of the worst-off group (p. 98), so it will serve for comparison.



Rawls believes, the possibilities of additional gain afforded by alternatives to RPJ are dubious on theoretical as well as practical grounds (the unknown level of scarcity prevents any sense of what real possibilities of absolute or relative gain there are), the safety of a workable and efficient conception of justice is, comparatively, extremely valuable. It is not to be passed over.

This leaves us in a tricky situation. Recall that C1 and C3 each provide independent support for F2. Hence, demonstrating that the parties know they mayn't be satisfied (in the sense of having the means to fulfill their life plans) with RPJ's minimum does not defeat F2. If it is true (C3) that the alternatives offer only dubious prospects of better, then showing, as we did above, that more than one alternative averts an intolerable outcome still does not provide enough grounds for the parties not to choose RPJ. Now, if we consider that the argument against C1 also cuts against the mix—given an unknown conception of the good, one can't be sure the mix will provide enough primary goods—C3 gains paramount importance. Even if an alternative to RPJ provides a tolerable minimum, the alternative will not be preferable to RPJ unless it promises real benefits over RPJ.

We must be careful not to argue against C3 in such a way as to strengthen C1, e.g., by arguing that there is some way in which the parties, under the mix, could somehow adjust to their circumstances. If they could do it for the mix, they might do it for RPJ;<sup>25</sup> and while, if they could do it for both, we would have established another alternative's tolerability, we would have done little to establish that the alternative was preferable. The considerations raised in the argument against C3 should count only against RPJ and/or in favor of the mix.

Given our argument in § IV for the parties' potential dissatisfaction with their society, and the considerations raised at the beginning of this section regarding primary goods that are out of the parties' control, the best that could possibly be said for any alternative is that it provides a viable prospect of a greater share of primary goods than RPJ does within a tolerable conception

---

<sup>25</sup> For instance, it could be argued that the parties would know that their conceptions of the good will be molded by their actual circumstances in life (cf. Kavka (1986), p. 213), and hence that the shortfall between expectations and available goods would probably not be devastating. This consideration works equally well for the supporter of RPJ.

of justice.

I must admit that it is beyond my ability—as well as beyond the reasonable scope of a paper such as this—to prove that a social structure based on the mix is at least as workable as one under RPJ. My task in the next section, then, is to give reasons why the POPs might prefer the mix to RPJ—that is, prefer the strongly constrained principle of average utility to the difference principle.

I will say, however, that I think it is doubtful that any conception of justice can be guaranteed to be tolerable even in the restricted sense. Any conception is at best only possibly tolerable. The parties do not have the luxury of a sure thing against which to weigh the alternatives. Given the parties' preference for more primary goods than less, the uncertainty of attaining a tolerable conception of justice would weigh in favor of the mix, if it offers better prospects.

### VI. C3: PROBLEMATIC BENEFITS

As implied above in the argument against C1, the POPs would be inclined, perhaps strongly, to accept an alternative which provided the prospect of more primary goods than was available under RPJ, if it was not too risky. After all, even in a tolerable social system (as defined at the beginning of § V above), more primary goods are still better than less. We have just established that the mix provides an alternative which avoids risking disaster at least as effectively as RPJ. Now, does the mix provide the better high-end prospects?

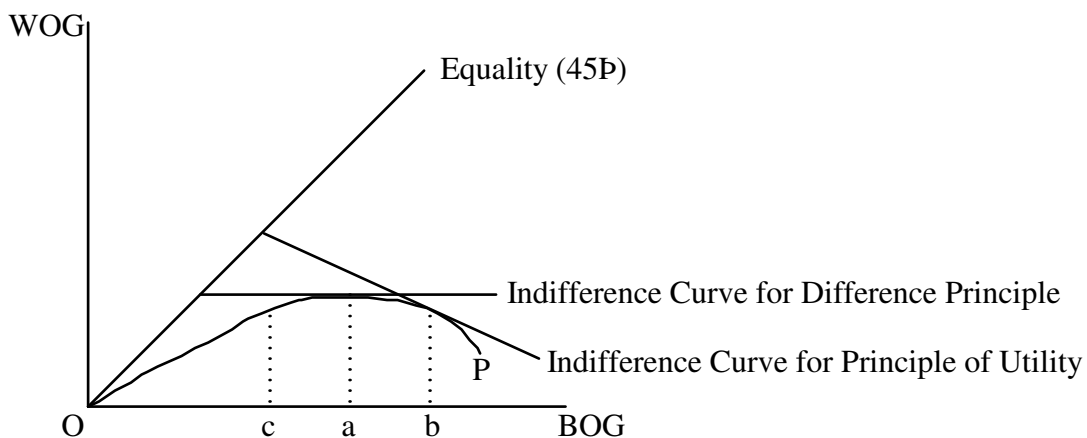


Figure 1

Consider Figure 1, which is adapted from Rawls' figures (76-77). WOG and BOG represent the worst-off group and best-off group, respectively. Curve OP represents the contribution to the WOG's expectations made by the greater expectations of the BOG. Point 'a' on curve OP represents the "perfectly just scheme" according to the difference principle (cf. 78). When there is a distribution beneficial to both the BOG and the WOG, the difference principle selects this point, because to the right of 'a' the BOG benefits more while the WOG's benefits decrease, and to the left of 'a' the WOG's benefit is less than maximum. Thus the BOG's gain is held in check by the need to maximize the WOG's prospects.

The principle of utility, however, allows the BOG's benefit to increase to point 'b', representing higher prospects for the BOG than the difference principle would allow. Note that point 'b' is on a level with 'c'; it thus allows the WOG a benefit that is still "just...but not the best just" to Rawls (79).<sup>26</sup> Technically, the principle of utility could allow the BOG to benefit at the WOG's expense, that is, would allow the WOG's prospects to fall below point 'O' while the BOG's rose, if that state of affairs improved the average of utility for all. Rawls could complain that if this happens, the parties will feel the strains of commitment (176ff). Considering the greater overall gain in utility to the northeast, though, the principle of utility would discourage this. But even if circumstances were such that following the principle of utility dictated a southeast point, the strain would be mitigated powerfully by (i) the guaranteed social minimum and (ii) the prospect of fair equality of opportunity guaranteed by the mix, and by (iii) the fact that no one's liberty would be allowed to drop at all, except under extreme circumstances in which the same would happen under RPJ (542).

Rawls rightly notes (78) that, other things equal, utilitarianism allows larger inequalities between groups. But if the POPs desire more primary goods rather than less, this is just what

---

<sup>26</sup> For "c" to be just requires a non-lexical interpretation of the difference principle, of course.

they may want. They might well want to run the risk of getting a merely livable social minimum if society offers them the prospect of possibly getting more than the difference principle allows, either by landing in a privileged position once the veil is lifted or by attaining it through fair equality of opportunity. Against the objection that the decision is final and the parties' descendants might reproach them for taking this risk, Hare's reply can be given: "might not their descendants be just as likely to say 'Nothing venture, nothing win'?"<sup>27</sup> The finality of the decision, then, cuts both ways. Given the nature of the primary goods they could amass by taking the risk—income, wealth—it would be rational to want more of them even in moderate scarcity, i.e., even when there isn't much to be had. This desire for more need have nothing to do with a sense of desert (cf. 103-104). It is not necessary that the parties feel they deserve what they can get in order for them rationally to want to have it.

Rawls might try to object that the mix's social minimum may be set in such a way that things operate as if the difference principle were in effect (316-7, 320). This state of affairs is not ruled out *per se* by the mix. But on my view the parties would not allow that to happen if it would defeat their purpose in adopting the principle of utility, namely, to allow the wealth the difference principle would have amassed at the bottom of society to instead benefit the better off, while the rest are assured of at least a decent living. And even if the distribution of wealth under the mix turned out to be the same as it would have been under the difference principle, it would be contingent on particular circumstances that make this the optimal arrangement. If circumstances change, the principle of utility allows the flexibility to adapt to a more optimal distribution.

The difference principle, on the other hand, maximizes the prospects of the WOG whether they want it, or need it, or not. Now, the POPs might find this feature of the difference principle attractive, given their motivation to secure the means to support an unknown life plan and the fact that they may turn out to be in the WOG. But once the veil is lifted and plans of life

---

<sup>27</sup> Hare (1989), p. 103.

are known, why insist on maximizing the prospects of the WOG even when the available wealth and income might be put to better use advancing the life plans of those whose plans require more? I am aware—as would be the parties—that the idea suggested by the principle of utility is a troublesome one in this regard. For if resources are allocated according to life plans, lazy but greedy persons could have leverage to better themselves by exaggerating the cost of their plans. But the difference principle's alternative is troubling, too, for it means that those who work to attain more will have more of it taken away from them than under the mix, and it will be given to people who may have less use for it. Granted, this may be an unrealistic scenario, but is it any more unrealistic than the scenario in which (cf. my Figure 1) the WOG's prospects are allowed to fall below point 'O', when average utility will rise far faster if it does not? In the name of safety, the difference principle could commit the parties to a policy of waste. Rawls complains that the principle of utility requires a greater identification with the interests of others (177f). But, as has been mentioned, the magnitude of sacrifice is minimized by the mix's constraints; and it must be wondered whether so much more benevolence is required to allow those above one's social level to prosper at one's expense than to do the same for those beneath, as the difference principle asks.<sup>28</sup>

Rawls has enumerated other problems with utilitarianism even when it is constrained in a mixed conception.<sup>29</sup> He notes that, compared to the difference principle, the principle of utility requires more information for its application because it is more difficult to ascertain how to maximize utility among all the different social groups than for just one, the WOG, as the difference principle dictates. The principle of utility's complexity in application, moreover,

---

<sup>28</sup> Cf. Arrow (1973), p. 257, complains that the difference principle does not escape treating people as means, because it "certainly seems to imply that the better off should sacrifice for the less well off, if that will in fact help." Nagel (1989), p. 13, makes a similar point: where utilitarianism may sacrifice the prospects of the WOG, the difference principle sacrifices those of the BOG, and it seems arbitrary to care more about one than the other. Nagel ignores, however, that when the BOG's prospects are lessened they will still probably have enough to eat. This is less likely true of the WOG. Such a state of affairs could obtain when the WOG and BOG are not far apart in wealth, i.e., when the rule of declining marginal utility of income does not strongly apply. This underscores the importance of the guaranteed social minimum offered by the mix.

<sup>29</sup> Rawls (1974a), pp. 143-5.

makes it more difficult to have confidence that it is being properly applied. But are these considerations decisive? Even Rawls does not think so.<sup>30</sup> Rather than argue that the costs of attempting to maximize utility are too high, he retreats to the idea that the parties will be inclined towards the difference principle due to the particular nature he assumes them to have. He feels that their aspirations as free and equal persons will point them to the difference principle. They will be drawn to it, he says, because it allows them to view the natural distribution of abilities (natural endowments such as strength and intelligence) as a collective asset to be managed for everyone's advantage, and this because "those able to gain from their good fortune do so in ways agreeable to those less favored."<sup>31</sup> This, however, is not a decisive consideration either. The parties must weigh the niceties of the sort of "personality" they would like society to have (e.g., the considerations raised in the first paragraph of this paper's introduction) against their desire for more rather than less primary goods. It is a delicate question whether RPJ could succeed at creating such fellow-feeling; why wouldn't many find its "one for all and all for the WOG" attitude rather onerous? It may be less to ask of people that they be willing to maximize the prospects of the worst-off rather than some other group, but even so it requires a generosity that is at odds with the need to see to it that one (and one's descendants) are as well off as possible. Not to opt for the prospect of a greater share of primary goods seems to imply not only an exaggerated fear of envy, but a sort of benevolence that is not a part of the OP. I am not contesting Rawls' statement (142) that "In choosing between principles each tries as best he can to advance his interests." I am claiming that Rawls has not made clear enough where those interests lie and hence what the unique outcome of the choice situation will be.

The point is not simply whether the utilitarian system, as worked out by its theorists in the literature, is troublesome, or whether it might be chosen through some questionable decision process such as the principle of insufficient reason (168ff) or by deciding as if we were in

---

<sup>30</sup> Rawls (1974a), p. 144.

<sup>31</sup> Rawls (1974a), p. 145. Obviously the importance of this consideration to the parties varies depending on whether they agree with Rawls' stance on natural endowments, which he considers a "fixed point." Rawls (1971), p. 104.

another's place (173-4). Even if it is, there is still the fact that not to choose the principle of utility is to say that at some points one will choose to have less rather than more of what one really wants more of.<sup>32</sup> I grant that the parties may be unwilling to make utilitarian tradeoffs in the distribution of their liberty, opportunities, and minimum living conditions, and hence in the primary goods of rights, liberties, powers, opportunities, and minimum income and wealth. But are they willing to forgo all such tradeoffs, to assume at the outset that they will in perpetuity, other things equal, settle for less rather than more wealth and income overall, and limit individuals' ability to amass more wealth because the basic structure demands the maximization of the WOG's prospects? Given the tension created by moderate but unknown scarcity and their motivation to have more rather than less primary goods, this seems quite possibly too much to ask.

I do not see, however, that any of these considerations, or all of them, are decisive in favor of either the mix or RPJ. I have been arguing that, *other things being equal*, the mix's distribution of wealth may be more attractive to the parties than RPJ's. But how are we to know whether things are equal? Of course we cannot know whether the material circumstances of the society will be more or less favorable; the veil prevents this. And neither can we know whether the circumstances of the society will turn out to be such as to favor the mix or RPJ, if either. But is it possible to determine, behind the veil of ignorance, whether either the mix or RPJ will provide a better basic structure for developing whatever resources the society turns out to have available? If one conception were to prove generally more *productive* than the other, that would be an influential factor.

Utilitarians would be quick to claim that even the constrained version of utilitarianism provided by the mix would have a clear advantage. But as noted above, their critics could easily

---

<sup>32</sup> Scheffler (1988), p. 9, makes this point regarding consequentialist vs. agent-centered moralities. I am adapting it here to apply to maximin vs. maximizing social structures. Strictly speaking, the problem is that to choose the difference principle is to choose an unalterable principle which will determine society's basic structure, such that the overall level of primary goods will be tied to distributions which maximize a certain group's prospects, even when to do otherwise would raise prospects for society generally.

point to the lengths utilitarianism must go to make its calculations, and ask whether this would undermine its advantages. This is obviously not a question answerable here, but I could point out that the mix, like RPJ, is only to be used to set up a basic structure for society (7), and a utility-maximizer would strongly wish to avoid creating a basic structure that required a great deal of effort to do its work. The maximizer's ability to avoid this, however, is contingent on his solving a number of formidable theoretical problems.

What we have, then, is a number of significant factors; but not, as far as I can see, enough to make up the parties' minds. The version of utilitarianism contained in the mix is subject to some of the problems of utilitarianism generally, but there is nothing in Rawls' definition of the OP to stop the parties from deciding not to be compelled by this. On the other hand, RPJ provides attractive simplicity and safety. If Rawls wants to show that parties will clearly choose one and only one of the available options, then, he will have to alter his OP. The fact that the conditions of the OP will lead the parties to be highly risk-averse is insufficient grounds for choice, because it is not clear which risk is greater: a safer but perhaps poorer society, or a maximizing but complicated one; and given F3's downfall, risk aversion would be eased anyway. But in altering the OP Rawls runs the risk of adding in assumptions which are more controversial than the ones already there, thus undermining the contract theory's justificatory power.



## BIBLIOGRAPHY

- Arrow, Kenneth (1973). "Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice," *The Journal of Philosophy* **70:9** (May), pp. 245-263.
- Barber, Benjamin (1989). "Justifying Justice: Problems of Psychology, Politics and Measurement in Rawls," in Norman Daniels, ed., *Reading Rawls: Critical Studies on Rawls' 'A Theory of Justice'*, Stanford University Press, Stanford, California, pp. 292-318.
- Brandt, Richard B. (1979). *A Theory of the Good and the Right*, Clarendon Press, Oxford.
- Daniels, Norman (1979). "Wide Reflective Equilibrium and Theory Acceptance in Ethics," *The Journal of Philosophy* **76:5** (May), pp. 256-282.
- Fellner, William (1965). *Probability and Profit: A Study of Economic Behavior Along Bayesian Lines*, Richard D. Irwin, Inc., Homewood, Illinois.
- Griffin, James (1986). *Well-Being: Its Meaning, Measurement and Moral Importance*, Clarendon Press, Oxford.
- Hare, R.M. (1989). "Rawls' Theory of Justice," in Norman Daniels, ed., *Reading Rawls: Critical Studies on Rawls' 'A Theory of Justice'*, Stanford University Press, Stanford, California, pp. 81-107.
- Kavka, Gregory S. (1986). *Hobbesian Moral and Political Theory*, Princeton University Press.
- Miller, Richard (1989). "Rawls and Marxism," in Norman Daniels, ed., *Reading Rawls: Critical Studies on Rawls' 'A Theory of Justice'*, Stanford University Press, Stanford, California, pp. 206-230.

Nagel, Thomas (1989). "Rawls on Justice," in Norman Daniels, ed., *Reading Rawls: Critical Studies on Rawls' 'A Theory of Justice'*, Stanford University Press, Stanford, California, pp. 1-15.

Rawls, John (1971). *A Theory of Justice*, Harvard University Press, Cambridge, Massachusetts.

Rawls, John (1974a). "Some Reasons for the Maximin Criterion," *American Economic Quarterly* **64**: 2 (May), pp. 141-146.

Rawls, John (1974b). "Reply to Alexander and Musgrave," *Quarterly Journal of Economics* **88**:4 (November), pp. 633-655.

Scheffler, Samuel, ed. (1988). *Consequentialism and its Critics*, Oxford University Press.