Rational and moral action : a critical survey of rational choice theory

de Jonge, J.P.R.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER XXII

FAIRNESS AND PRIMARY GOODS

1. Introduction:

For Rawls questions of justice are questions about how the benefits and burdens in society as a system of social cooperation are to be shared. These questions must be solved in "reasonable terms", indicating that social cooperation is founded on the idea that economic institutions are arrangements for mutual advantage in which parties cooperate on a footing of equality. The justice of social and political institutions is measured by their tendency to counteract the material inequalities deriving from birth, talent and circumstances. The basic arguments for his principles of justice are that they best systematize our considered judgements when we consider them in "reflective equilibrium", and that they would be chosen from a perspective which embodies constraints that seem morally compelling: the "original position".
Rawls himself begins his theory of the good with the idea of goodness as rationality as the starting point of a sequence of different ideas of the good. This idea allows him to say that things are good if they have properties that make it rational for people to want them, given their plan of life. Starting with that idea we next get the primary goods; once we have these, the argument from the original position can proceed, and so we arrive at his two principles of justice, which are then used to specify permissible (comprehensive) conceptions of the good. Once the two principles of justice are accepted, the political virtues essential to sustain a just basic structure can be identified. (Rawls, 1988) Thus, given a formal definition of goodness, social unity is founded on the publicly accepted political conception of justice to regulate the basic structure of society. The concept of justice is independent from and prior to the concept of goodness in the sense that its principles limit the conception of the good that is permissible. (see Rawls, 1985)

The parties in the original position have to decide under conditions of ignorance. They must be viewed as being within an initial situation of equality with regard to the primary goods and their task must be seen as determining the principles for the distribution of these goods in cases where this situation of equality is not in force.

I will start my exposition of Rawls’s theory with presenting its Kantian roots. Thereafter, I describe the original position in some detail, before I turn to the principles of justice as proposed by Rawls. Of course, the question is raised why people in real life should accept principles agreed to under a veil of ignorance. This question will be discussed as the problem of stability in section 5. The next two sections will be devoted to the criticism on Rawls’s theory of justice. A central place is occupied by the comparison of alternative conceptions of the original position.
2. Kantian Constructivism

Rawls's theory of justice "is a theory of the moral sentiments (...) setting out the principles governing our moral powers, or more specifically, our sense of justice." (Rawls, 1971, 51) His theory is grounded on a Kantian perspective of morality.

Kant conceived of moral judgments as essentially to be reason-based. From which follows that moral judgments should be universal. An ethical system, such as utilitarianism, which judges institutions and actions on the basis of how much pleasure they generate or how many desires they satisfy, fails to consider the most important aspect of human beings, namely, their rational nature. "Kant's account of what we ought to do is to be understood in terms of what we would do, if reason completely determines the will." (Darwall, 1980, 319)

An essential point for any conception of justice is that it only fulfills its social role provided that citizens, by affirming it, are led to a sufficient convergence of opinion. "Thus a conception of justice should be framed to meet the practical requirements of social life and to yield a public basis in the light of which citizens can justify to one another their common institutions." (Rawls, 1980, 561) Principles of justice are to be applied to the design or criticism of the major institutions of society that define people's rights and duties. Institutions are defined by their constitutive rules, which create offices and roles, regulate procedures, and influence the overall distribution of benefits and burdens. The principles of justice apply to these rules and only indirectly to the acts and to the states of affairs that fall under the rules. A Kantian concept of justice tries to dispel the conflict between different understandings of freedom and equality, by asking which conception free and equal moral persons would agree to. "The two basic model-conceptions of justice as fairness are those of a well-ordered society and of a moral person. (...) The 'original position' is a third and mediating conception. (Rawls, 1980, 520)

What are the features of a well-ordered society? First, as we just saw, it is a society in which everyone accepts, the same
public principles of rights and justice. Justice as fairness takes as its fundamental idea that of society as a fair system of cooperation over time, from one generation to another. To the citizens, as participants in a fair system of social cooperation, two moral powers are ascribed: a capacity for a sense of justice and a capacity for a conception of the good. Related to these two capacities are a disposition to be reasonable and a disposition to be rational. The reasonable subordinates the rational because its principles limit the final ends that can be pursued. "In justice as fairness the Reasonable frames the Rational (...)." (Rawls, 1980, 532). This means that the principles of justice are lexically prior to claims of the good.  

The principles of justice and the rights and liberties they define cannot be overridden by considerations of efficiency and a greater net balance of well-being. (Rawls, 1985) In a well-ordered society there is a public understanding of the appropriate claims for citizens to make. How is such public understanding imaginable? The notion of primary goods addresses this moral and practical problem. It rests on the idea that particular conceptions of the good, how different they may be, require for their advancement roughly the same primary goods. Rawls assumes that each citizen is presented

---

340 Citizens recognize one another as having the moral power to have a conception of the good. The weight of their claims is not given by the strength and psychological intensity of their wants and desires, but by their needs and requirements as citizens. (Rawls, 1985, 244) Everyone is equally capable of understanding and complying with the public conception of justice. Therefore, all are capable of honouring the principles of justice and of being full participants in social cooperation throughout their life.

341 Reasonableness is a matter of being prepared to propose, to listen to, and to abide by proposals that constrain the pursuit of ends if others will accept the same constraints. Reasonableness is a matter of willingness to accept terms 'which all can accept'.

342 Thus, it was an error to describe his theory in *A Theory of Justice* as part of the [economic] theory of rational choice (Rawls, 1985, 237).
by a party in the original position, who decides between alternative principles of justice according to how securely these principles provide all citizens with primary goods. The share of primary goods that citizens receive is not intended as a measure of their psychological well-being. "The idea of restricting appropriate claims to claims to primary goods is analogous to taking certain needs alone as relevant in questions of justice." (Rawls, 1982, 172)

3. The original position

In the derivation of the principles of justice Rawls makes use of the same special hypothetical situation as Harsanyi, "the original position". People have to agree over principles that will shape the basic institutions of society while behind a "veil of ignorance" that deprives them of the knowledge of the role they will occupy in society, their good or ill fortune in the distribution of natural talents and abilities, their final aims and interests, and of their particular psychological makeup (their aversion to risk or liability to optimism or pessimism). Neither do they have knowledge over the present state of society, the stock of natural resources or the level of technology. That is, they do not know the economic or political situation, or the level of civilization and culture. The veil of ignorance also rules out any knowledge about entitlements to certain sorts of goods and claims they have toward other individuals. Thus they have to decide in a situation of radical uncertainty. It is taken for granted, however, that they know the general facts about human society. They understand political affairs and the economy; they know the basis of social organization and the basic elements of human psychology. "While the original position

---

The whole point of Rawls's approach is that it leaves people to form, revise and pursue their own conception of the good. Would the parties in the original position know their conception of the good, they would have no incentive for building into the principles of justice a preference for each particular conception of the good.
as a whole represents the two moral powers, the capacity to be reasonable and the capacity to be rational, the parties in the original position, as rationally autonomous representatives of the citizens in society, represent only the rational. 

([..] The reasonable, the capacity to honour fair terms of social cooperation, is guaranteed by the various restrictions to which the parties are subject in the original position and by the conditions imposed on their agreement." (Rawls, 1993, 305) “Indeed, the rational autonomy of the parties is merely that of artificial agents who inhabit a construction designed to model the full conception of the person as both reasonable and rational” (Rawls, ib., 306) The parties in the original position are presented with a short list of traditional conceptions of justice and are required to agree unanimously to one conception that they consider is the best among those enumerated. These principles must bear on the distribution of primary goods. Primary goods are singled out by asking which things are generally necessary as social conditions and all-purpose means to enable agents to exercise their moral powers. The idea is that persons require for their advancement roughly the same basic rights, liberties, opportunities, and the same all-purpose means as wealth and income. 344

These goods are things citizens need as free and equal persons, and claims to these goods are counted as appropriate claims. The basic list of primary goods has five headings; a) basic rights and liberties; b) freedom of movement and freedom of choice of occupation; c) powers and prerogatives of offices and opportunities; d) income and wealth; and e) the social basis for self-respect. People should enjoy the same liberties and equality of opportunity [(a) and (b)]. The only permissible difference among citizens is their share of the primary goods under c), d) and e). (Rawls, ib., 308)

344 Primary goods are defined as things that: “from the standpoint of the original position, it is rational for the parties to want whatever their final ends (which are unknown to them). They serve as generalized means ([..]), for realising all, or most all, rational systems of aims.” (Rawls, 1993, 278)
After the principles of justice are selected, they have to be implemented in society. To this end Rawls has developed a sequence of four stages, of which the original position is the first. The following stages are the making of the constitution, the legislation and the adjudication of particular cases. At each stage the veil of ignorance is progressively lifted so that the general knowledge required at that stage is available. At each stage the options are subject to constraints imposed at the earlier stage.

Crucial questions concerning the appraisal of Rawls’s approach are: 1e who are to be included among the parties in the original positions 2e what are they allowed to know and 3e how can we be sure that the actual living people accept the principles agreed to in a completely hypothetical setting? Rawls has restricted the membership to the parties to the members of one generation in world’s history. (Rawls, 1971, 292) 345 It is evident that the parties in the original position are to take a decision behind a thick veil of ignorance. This characteristic of the original position has two aspects. On the one hand the parties are initially in a situation of complete equality with regard to the primary goods and their task must be seen as determining the principles for the distribution of these goods in situations where equality is not in force. On the other hand, the parties find themselves in a situation of extreme uncertainty. Obviously this will influence their decision. This aspect of uncertainty has provoked many discussions, as we will become aware off in the sections to come. The third question will be discussed in section 5, after I have discussed Rawls’s principles of justice and their justification.

345 In *A Theory of Justice* Rawls addressed the question of savings between generations. The problem is that his difference principle would not support such savings, assuming that next generations are better off. (Arrow, 1977) By restricting the membership to one generation he sidestepped this problem. (see Rawls, 1974)
4. Justice as fairness

Rawls argues that in the special circumstances of deciding behind a veil of ignorance it is rational first to seek to safeguard basic political and personal liberties and second to minimize the costs of winding up at the bottom. Therefore the parties in the original position will judge basic social and economic institutions according to the capacity to secure basic liberties and to the tendency to promote the interests of the least well-off social group 346. The most well-known and most disputed principle is the second principle of the special conception, called the "difference principle" (especially the second part of it).

What are Rawls' arguments for this difference principle? Rawls uses two different kinds of considerations. He derives his arguments from an analogy between this principle and the maximin rule for choice under uncertainty. The principle is, moreover, based on considerations of fairness and stability. "The maximin rule is not a suitable guide for choice under uncertainty in general, but is attractive in situations marked by certain special features." (Rawls, 1971, 153) Firstly,

346 Rawls, in fact, formulated two conceptions of his principles of justice. A general conception, that says:
"All social primary goods (...) are to be distributed equally unless an unequal distribution of any of all these goods is to the advantaged of the least favoured." (Rawls, 1971, 303)
This general conception is held to be applicable in relatively frugal conditions.
The special conception is formulated in two principles:
1. "Each person has an equal right to a fully adequate scheme of equal basic rights and liberties, which scheme is compatible with a similar scheme for all; and in this scheme the equal political liberties, and only those liberties, are to be guaranteed their fair value."
2. "Social and economic inequalities are to satisfy two conditions: first, they are to be attached to positions and offices open to all under conditions of fair equality of opportunity; and second, they are to be to the greatest benefit of the least advantaged members of society." (Rawls, 1993, 5/6; see also Rawls, 1971, 302/3)
The first principle is lexical prior to the second. The two principles together regulate the basic institutions that realize these values.
the situation is such that knowledge of likelihoods is impossible, or at best extremely insecure. Secondly, it is not worthwhile to take chances for further advantages above the minimum. Thirdly, the rejected alternatives have outcomes that are hardly acceptable. The paradigm situation for following the maximin rule is when all these features are realized to the highest extent. The original position has been defined in such a way that the maximin rule applies. (Rawls, 1971, 154) One cannot properly assign probabilities to the options available in the original position. The information to derive objective probabilities is lacking and Rawls rejects the logicist view of probabilities. He draws attention to the well-known paradoxes that arise when attempts are made to apply Laplace's principle of insufficient reason. (see the preceding chapter)

Rawls also rejects the use of subjective probabilities. He thinks this is not appropriate for 'large world' decision problems. Thus the original position is characterized by

---

347 The maximin rule was popular in the first part of the twentieth century, but was abandoned with the rise of modern, Bayesian, decision theory and replaced by the principle of expected utility. It has to be added that simultaneously the concept of pure uncertainty was replaced by the concept of calculable risk in the guise of participating in a lottery. In the vNM utility functions uncertainty is replaced by risk combined with a concept of risk aversion. The question of risk aversion, therefore, is a dominant theme in the discussion about Rawls's difference principle. It is relevant in this context to notice that Von Neumann and Morgenstern explained that the solution concept of a two person zero sum game is the minimax criterion: every player behaving rationally receives at least this amount when the other player acts rationally (it is a Nash solution). The minimax is played after the elimination of dominated strategies. The reason for choosing the minimax strategy is that it guarantees the security level. It secures the best of all the worst outcomes, while avoiding the risk of being found out by a rival. There is a close resemblance with the maximin rule. This resemblance remained apparently unknown for many critics of Rawls, as game theory remained largely unknown to many economists until the late 1970s.

348 In a way Savage supports this point of view, because he insisted that it is only sensible to use subjective probabilities in the context of what he called a 'small world'. As Savage puts it, a small world is one in which it
radical uncertainty and great risk. The implication of complete ignorance requires the parties in the original position to consider the worst eventualities and to regard them as real possibilities. Another consideration for the parties involved in the original position is that they must select only those principles about which they are sure that they can always be honoured; principles moreover that shall be used to arbitrate conflict and to assess as a final court of appeal. This makes the parties involved reluctant to gamble on principles. Therefore they have to negotiate a final result that is fair to all reasonable persons.

When the parties consider both themselves and their companions as free and morally equal, an equal division is the most obvious option. The reference point for the division of the primary goods is, therefore, an equal division. But the size of the societal product is not fixed; rather it depends on individual incentives. It is the incentive problem that moves Rawls away from strict equality. Inequality in the division of primary goods is therefore allowed if it increases the size of societal product. Rawls claims that the difference principle is the only distributive rule that meets the reciprocity condition: those who are better off at any point are not better off to the detriment of those who are worse off at that point and they is always possible to "look before you leap". A decision maker can then in advance take into account the impact that all conceivable future pieces of information might have on the underlying model that determines his subjective beliefs. Any mistake built into his original model that might be revealed in the future will then already have been corrected, so that no possibility remains of any unpleasant surprises. (Binmore, 1994, 305) Rawls, by the way, emphasizes that probabilities, if they are to be grounds of rational choice, must have an objective basis. (see Rawls, 1971,172)

Despite this analogy, Rawls emphasizes that the difference principle and the maximin rule for choice under uncertainty are two very different things. (Rawls, 1974, 141) The only purpose of the analogy is to argue that there is no sense in assigning probabilities to uncertain events or states. In spite of this, most critics have tended to equate the difference principle with the maximin rule. That is, they accept the maximin rule as an explanation for the choice of the difference principle.
should choose the outcome that maximizes the outcome of the worst off. It satisfies the "strain of commitment"; the requirement that parties must decline a principle if they have reason to doubt that they will be able to honour the consequences of its consistent application. Rawls argues that the greater the reciprocity that members of a society feel with one another, the more likely they are to gain allegiance to principles and institutions that embody that feeling and thus the more stable a society guided by these principles is likely to be. There exists in "justice as fairness" a strong link between the question of fairness and the question of stability. The problem of stability will be further discussed in the next section.

To conclude this section on Rawls's principles I will mention a critical remark on the ‘index-number problem’. The difference principle assumes that we are able to pick out the least advantaged representative group or individual. This requires statements of the kind “individual A is worse off than individual B”. Thus we have to be able to rank the individuals on the basis of the amount of primary goods they possess. But the primary goods are different of kind and therefore there is an index-number problem in commensurating these different primary goods. (Arrow, 1973, 254)

How can we aggregate a person's holdings of various primary goods into an overall measure of the size of his share of primary goods? Rawls has not really offered a solution to this problem. 351

---

350 Rawls believes that the strain of commitment is less when the difference principle is selected instead of the average principle, because the less advantaged will adopt it sooner than the average principle. Buchanan even argued that, given the threat-potential of the relatively disadvantaged, it is rational to select the difference principle. In fact, this stability argument far outweighs the role attributed to risk-aversion. (Buchanan, 1976)

351 When primary goods are defined as concrete goods such as food, shelter, security, etc., they will be differentially useful to individuals depending on their fundamental values. There are theorists who argue that the discussion about the size of the share of primary goods neglects that people who have the same amount of primary goods may differ in
5. The problem of stability

The Rawlsian original position is one in which individuals are stripped of all their particular interests, desires and special knowledge in order that they can be seen to be in a situation in which all are equally informed about the general nature of society without any knowledge of its actual state and their own position within society. What the principles of justice must regulate is the inequalities in life prospects between citizens that arise from social starting positions, natural advantages and historical contingencies. How should this hypothetical construction claim (ethical) authority? Why would people in real life commit themselves to the principles of justice obtained in a hypothetical situation? Gauthier, for instance, believes that redress of the accidents of natural and social endowment would not be a concern of self-interested persons. (Gauthier, 1974, 16) Thus, they would never agree to redistribution. They would, for instance, argue that people not involved in the actual production of primary goods should be excluded from the benefits of the cooperative efforts. In a reaction, Rawls notices that there is no place for the idea of an individual’s contribution to society that parallels that of an individual’s contribution to associations within society. Membership in society, which he conceives as a cooperative enterprise, is given. (Rawls, 1993)

The question of the stability of the agreed principles requires that persons possess a ‘sense of justice’, an effective desire to

---

well-being. The question whether a distribution is fair should, therefore, depend on the outcome in terms of well-being that individuals reach under that distribution. (see Arneson, 1990b) But the consequence of this kind of reasoning is that people with expensive preferences should be entitled to receive more of the primary goods than people with modest preferences. (see also chapter XXIV) I think that the index-number problem is slightly exaggerated. Primary goods are all-means goods, goods that every agent wants. The most relevant are income and wealth. In terms of these primary goods it is not difficult to establish who belong to the worst-off.
act as the principles of justice require. Since Rawls's principles give maximal expression to the interests of the less advantaged, it is not unlikely that they will support the principles. But what about those who are well off? What serves to engender in them the requisite sense of justice? Crucial to make this link is Rawls's notion of a reflective equilibrium. According to Rawls each individual has some intuitive notions of ethical justice and also an image of a hypothetical original position from which an ethical principle may be derived by contractarian processes. These intuitions will play a part in the deliberations in the original position and equally the principles deriving from the hypothetical position may cause the individuals to revise their intuitions. Thus the method of reflective equilibrium is an attempt to produce coherence in a triple set of beliefs held by a person, namely; "a) a set of considered moral judgments, b) a set of moral principles and c) a set of relevant back-ground theories [for instance: theories of the person; theories of procedural justice; general social theory and so on]. (..) We can imagine an agent working back and forth, making adjustments to his considered judgments, his moral principles and his background theories." (Daniels, 1979, 258/9) Only when this process of redesign and revision is stationary, so that the principles agreed are consistent with his moral views and his general knowledge, an individual is said to be in a state of reflective equilibrium. This idea of a reflective equilibrium as a kind of mental state is a device to ground the hypothetical contract in the ethical intuitions of individuals, so that it can be argued that this contract binds real individuals. This device supports a "coherence" theory of morality.

In his *A Theory of Justice* Rawls has solved the congruence between the requirements of justice and of the good via the Kantian interpretation of justice. According to this point of view acting justly is something we want to do as free and equally rational beings. However, a modern democratic society is characterized by a pluralism of incompatible, yet reasonable, comprehensive doctrines. In Political Liberalism (1993), Rawls therefore has reformulated the idea of a well-
ordered society so as to take account of this pluralism. The conception of justice should, so far as possible, be independent of controversial philosophical and religious doctrines. Thus, to formulate such a conception, the public conception of justice must be political, not metaphysical. As a consequence, the Kantian interpretation can, according to Barry, no longer be regarded as a straightforward solution for the problem of congruence (and stability). The reason being that the Kantian interpretation is just one of the reasonable comprehensive views Rawls takes into consideration. The stability requirement now demands an "overlapping consensus" among these comprehensive views. Rawls's first principle is at least tolerably established among the most of the major reasonable comprehensive views. "It is when we come to the second principle that the strategy of an overlapping consensus collapses." (Barry, 1989, 911) Barry suggests abandoning the idea that a consensus is to be reached among doctrines, and focusing instead on the need for an overlapping consensus among reasonable people (with different comprehensive ideas). 352 To make liberal democratic policies secure, we only need persons who combine "scepticism on conceptions of the good or "comprehensive views" with strong conclusions about the requirements of justice." (Barry, 1995, 915) All that can, and needs to be justified to reasonable others, are the principles of justice and the basic structure of a just society.

6. Rawls and his critics 353

The idea of the original position is to set up a fair procedure

352 Barry distinguishes the epistemological claims of comprehensive views from their content. With regard to the epistemological claims he makes a further distinction between dogmatist and sceptical views. Only persons representing sceptical views can be part of an overlapping consensus. (Barry, 1995)

353 In this section I will present a heterogeneous collection of critical remarks about the whole of Rawls's enterprise. This will be followed by an exchange with utilitarianism.
that guarantees that any principle agreed to will be just. When there is no independent criterion of justice, we must turn to fair procedures. When the procedure is fair and properly followed, then the outcome can be regarded as fair, or not unfair. As Rawls formulates it, the original position incorporates pure procedural justice. "This means that the outcome of the original position defines (..) the appropriate principles of justice." (Rawls, 1980, 523) But both the procedure as the resulting principles of justice (especially the difference principle) came under attack.

Rawls calls his theory a contractarian theory, but some critics wondered whether his principles of justice have a contractual basis at all and even Rawls sometimes seems to deny this. Hampton argues that the veil of ignorance ensures that the differences among the parties are unknown to them, and since everyone is equally rational and similarly situated, each is convinced by the same arguments. Faced with identical information and reasoning in an identical way, the parties can only arrive at identical conclusions. One can understand the deliberation in the original position as carried out by a single deliberator, whose principles are voluntarily consented to, because they are recognized as self-imposed. (Hampton, 1980, 338) Rawls admits this: "(...) we can view the choice in the original position from the standpoint of one person selected at random." And from that follows the very important consequence "that the parties have no basis for bargaining in the usual sense." (Rawls, 1971, 139) Thus the selection seems rather characterized by choice than through a contract. Barry thinks that the concept of a contract under the conditions of the original position as Rawls has construed it is entirely trivial. Rawls is thus open to exactly the charge he levels against utilitarianism: "he does not take seriously the plurality or distinctiveness of individuals (..)." (Barry, 1996, 59) Gauthier also rejects the idea that the agreement in the original position can be conceived as a social contract. Since the parties do not know due to the veil of ignorance their ends, abilities, endowments etc., their is no room for bargaining. This is the basis for Gauthier's objection: Rawls's
agreement is not a contract because it is not modeled as an economic bargain. (Gauthier, 1974) (In the following chapter I shall give more details of the critique of Gauthier on Rawls).

In defense of Rawls it is argued that the agreement in Rawls's theory is not to be thought as a compromise between conflicting interests, but more akin to the idea of a rational pre-commitment. Because free and equal moral persons have a fundamental social interest in cooperating with one another in publicly justifiable terms, "[w]e may look upon this desire to realize this interest as a meta-preference for justice, that is, a preference that all one's futures preferences meet the condition of reasonable principles." (Freeman, 1992, 144). Nevertheless, some philosophers remain unconvinced and argue that, though the egalitarian liberalism which Rawls develops and the conception of the good on which it depends are extremely persuasive, the original position serves to model his principles rather then to justify them. They think that Rawls's conclusion would be more persuasively defended by direct moral arguments for liberty and equality. (Nagel, 1975; Scanlon, 1982; Barry, 1996) 354

The main point of criticism concerns, as to be expected, the difference principle. Rawls defended this principle with arguments borrowed from the maximin rule. I already mentioned that he thinks that three features of decision situations that warrant the use of this rule are manifest in the original position. In the first place, he denies that it is possible to have any knowledge of the probable consequences of one's choices. Moreover, it is inappropriate to make use of the

354 Sen also wonders whether a 'transcendental' approach (identifying perfectly just societal arrangements) has any relevance for a comparative approach (ranking alternative societal arrangements). A comparative analysis of justice may be entirely viable and usable without containing any answer to the question: "what is a just society?" (Sen, 2006, 236) We must seek a framework for a theory of justice that concentrates on advancements, an approach that reduces injustice. (See also chapter XXIV).
principle of insufficient reason given the fundamental importance of the decision to be taken. Thus the parties in the original position have to make use of a decision rule that does not take probabilities into account, and such a rule is the maximin rule. This argument is severely criticized. Some critics argue that it is a great leap from wariness about probability calculations to a strategy in which people behave as if they are 100% certain of the worst outcome (Mueller et. al. 1974). Harsanyi uses a similar argument. He believes that “[I]t is extremely irrational to make your behavior dependent on some highly unlikely unfavorable contingencies regardless of how little probability you are willing to assign to them.” (Harsanyi, 1975, 595).355

But philosophers who also believe that probabilities should be totally discounted support Rawls on this question. Nagel thinks “their suppression is justified on the assumption that the proportion of people in various social positions is regarded as morally irrelevant, and this must be the case because it is not thought to be acceptable to sum advantages and disadvantages over persons, so that a loss for some is compensated by a gain for others." (Nagel. 1975, 12) 356

The second reason for referring to the maximin criterion is

355 In a similar vain it is argued, that, while it may be true "that individuals would have no notion of their eventual place in society, they would have some idea of the likely structure of society and some impression that the least-advantaged would represent a minority only". As solipsists, they would be concerned to maximize their own expected utility (Rowley & Peacock, 1975, 142). This kind of criticism ignores that Rawls, in contrast to Harsanyi', has put the parties in the original position behind a thick veil of ignorance. They simply have no idea of the structure of the present society. And furthermore, from a moral point of view the size of the least-advantaged group is irrelevant.

356 He adds that "this aspect of the design of the original position, therefore, appears to be motivated by the wish not to extend to society as a whole the principle of rational choice for one person", which clearly contradicts the statement of Rawls "that we can view the choice in the original position from the standpoint of one person selected at random." But Rawls means that the thick veil of ignorance deprived the persons in the original position of the knowledge of claims or distinctions on which bargaining could be based.
that the parties in the original position care little about the
gain above the minimum stipend they can be sure of by
following the maximin rule. This argument rests on the
assumption that the principles of justice selected will ensure a
socially acceptable minimum, but also on the conviction that
what people need is meaningful work in free association with
others. To achieve this state of affairs great wealth is not
required. (Rawls, 1971, 290) The third reason is that the parties
in the original position will not take a risk, considering what
they might lose if the chances are against them.

With hindsight it can be established that the analogue
between the difference principle and the maximin rule has
only created misunderstandings. Instead of focusing the
attention on the constraints that the original position imposes
on the deliberating parties, it had the effect that the two rules
were equated. Rawls has himself contributed to this interpre-
tation by referring to the difference principle as the maximin
equity rule. (Rawls, 1974) This formulation strengthened the
idea that, instead of providing a rule legitimating to restrict
inequality, the difference principle demanded that under all
circumstances priority should be given to the needs and
wishes of the worst-off members of society. A demonстраtion
of this interpretation is Harsanyi's criticism of the difference
principle. Harsanyi vigorously repudiates the difference
principle, because in his view it discriminates against the
interests of people, who happen to be rich, or gifted, etc. The
difference principle "(...) would require us to give absolute
priority to the interests of the worst-off individual, no matter
what (..)." (Harsanyi, 1975, 596) The utilitarian moral code, on
the other hand, gives equal a priori weight to every person's
legitimate interests. Hare also suggested that the difference
principle allows us "to impose any loss, however great, upon
a better-off group in order to bring a gain, however small, to
the least advantaged group, however affluent the latter's
starting position." (Hare, 1975, 107) 357

357 ‘This criticism can be reversed. The difference principle also allows
tremendous gains to the well to do as long as the position of the less well
This kind of criticism makes it clear that the discussion of Rawls's principles of justice is very confused. There seems to be a lack of awareness of the difference between distributive justice and allocative justice. Rawls's principles of justice apply only to the main public principles and policies that regulate social and economic inequalities. The rules of allocative justice apply directly to individuals and associations and are to be followed by them in their transactions. When he discusses allocative justice, Rawls does not seem to differ in opinion with Harsanyi and Hare. For then he says "(..) the allocative conception of justice seems to apply when a given collection of goods is to be divided among definite individuals with known desires and needs. The goods to be allotted are not produced by these individuals, nor do these individuals stand in any existing cooperative relations. Since there are no prior claims on the things to be distributed, it is natural to share them out according to desires and needs, or even to maximize the net

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th></th>
<th>II</th>
<th></th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>y</td>
<td>1</td>
<td>20</td>
<td>1</td>
<td>20</td>
<td>2</td>
</tr>
</tbody>
</table>

i and j are individuals, the numbers present primary goods and x and y are alternative policies. Case I is the case referred to above. In this case policy x should be chosen, because it seems that the position of j has been improved at the expense of i, the least well-off. This is a situation that Rawls wanted to exclude, as a safe-quad against exploitation. In case II, to favour policy y seems to grant j a superior claim, to favour x grants i a better claim. In this case the lexical difference principle should prevail, which entails that when the lowest shares are equal under alternative policies, we should move to the next lowest position and maximize benefits for it. This means we should adopt policy y. In case III policy y will also be the preferred one. Rawls attached no intrinsic value to equality. There is no objection to a policy that makes no one worse off and some better off. Rawls does not want to reward envy. Envy is not a moral feeling. (Rawls, 1971, 533) There is a significant difference between the maximin rule and the leximin rule; most discussions refer to the first.
balance of satisfaction." (Rawls, 1971, 88) The objection that the difference principle enjoins continuous corrections of particular distributions and capricious interferences with private transactions is consequently based on a misunderstanding. The difference principle is to justify inequalities from a starting point of equality, not the other way around.358

In the next section I will delve deeper into the conflict between Rawls and his utilitarian opponents and compare their conceptions of the original position.

7. Rawls and utilitarianism

Rawls aspired to show that utilitarianism is inferior to his own theory. Using the overview presented by Smith Goldman (1980), his arguments can be divided in two categories. Some are designed to show that utilitarianism is inadequate without invoking the contractarian apparatus, that is, the original position, the veil of ignorance, and so forth. Smith Goldman refers to these arguments as "extra contractarian" arguments.

The contractarian arguments take the form of attempting to establish that the parties in the original position would choose Rawls’s principles of justice rather than the average principle. In fact this was the discussion about the difference and the average principle in the previous section. In this section I want to argue that the choice between the difference or the average principle is already implicated in the conception of the original position.

The major extra-contractarian arguments are: "i) the argument that utilitarianism generates prescriptions which

---

358 On Rawls’s conception of the morally arbitrary all differences in achievement are based on morally arbitrary factors. The "natural lottery, which distributes genetic endowments; the "social lottery", which distributes home and school environments; and the "lottery of luck", that distributes illness, accidents, and so on. There is no case at the most basic level of justification for anything else than equality in the distribution of primary good. (Rawls, 1971, 74; 104)
violate our considered moral judgments concerning what is just and unjust; ii) the argument that the reasoning in favour of utilitarianism illegitimately "merges persons" and iii) the argument that utilitarianism requires making interpersonal comparisons of utility that have no scientific basis." (Smith Goldman, 1980, 351) I will first discuss these objections toward utilitarianism, thereafter, instead of merely criticizing utilitarianism; I will compare the constructions of the original position.

In arguing that utilitarianism generates prescriptions that violate our considered judgments on justice, Rawls may have two arguments in mind. First, utilitarians are only concerned with the size of the cake (i.e., maximizing average utility), not how it is distributed. In Rawls's opinion a differential reward structure can only be defended when the compensation due to the "inequality surplus" is big enough and the size of the share of the least well-off group is larger than under the circumstances of the original practice with equal compensation for everyone.

The second argument that utilitarianism might violate moral intuitions is based on the assertion that a utilitarian doctrine may require unacceptable sacrifices of liberties. I doubt that this is really a strong argument. When the question is raised whether basic liberties should always be given priority, then the opinions of Rawls and Harsanyi don't seem to differ that much. Harsanyi presents the case of underdeveloped countries where economic growth may not be set in motion without concentrating a good deal of power in the hands of government and perhaps even some curtailment of civil liberties. "[J]udgement about any particular policy must always depend on the balance of the advantages and disadvantages it is likely to yield (.).

(Harsanyi, 1975, 602) Rawls presents a similar argument. Under the general conception of justice particularly dire conditions may make it rational to sacrifice basic liberties for the sake of other goods such as increased security or economic development. (Rawls, 1971, 247)
Rawls' second argument against utilitarianism is that it fails to take seriously the distinction among persons. Rawls’s accusation is directed against the model of the impartial observer with complete empathetic identification in which society is taken as if it were one person. But this is just one model of social choice, and, for instance, surely not Arrow’s. Moreover, we have seen that Rawls is accused of making a similar error, namely that the agreement in the original position is merely the choice of a representative individual. Alternatively Rawls could have referred to another content of the impersonal strain in utilitarianism, namely that benefits and burdens can be freely weighted against each other, even if they come to different people. Rawls rejects this impersonal view of benefits and burdens, since someone’s burden cannot be compensated by benefits to someone else.\[^{359}\]

Rawls' third argument against utilitarianism arises from the fact that utilitarianism requires us to make interpersonal comparisons of utility. In his view it is possible to arrive at an objective comparison of the distribution of primary goods, but it is not possible to measure and compare the satisfaction of preferences. However, how are we to compare the benefits derived from rights, or liberties, or powers, or opportunities? Rawls argues that this is not a serious problem for the following reasons. "The fundamental liberties are always equal, and there is a fair equality of opportunities; one does not need to balance these liberties and rights against other values. The primary goods that vary in their distribution are the powers and prerogatives of authority, and income and wealth. (...) The only index problem that concerns us is that for the least advantaged group. (...) We try to do this by taking up the standpoint of the representative individual of this group and asking which combination of primary goods it would be rational for him to prefer." (Rawls, 1971, 93/4)

\[^{359}\] Parfit argues that there can be moral reasons that allow that the burdens of one person are outweighed by the benefits of another person. (Parfit, 1984, section 115)
Arrow thinks that when this problem could be resolved by reducing everything to a single primary good, the same could be done in the sum-of-utilities approach. (Arrow, 1973) Rawls keeps on denying that the index-problem is comparable. "The theory of primary goods is a generalization of the notion of needs, which are distinct from aspirations and desires." (Rawls, 1974, 643) By comparison it is much more difficult to know what maximizes average utility than to determine what kind of policies sustain the needs of the least well-off group.

The extra-contractual arguments against utilitarianism are not equally powerful and sometimes not powerful at all. To demonstrate the inferiority of utilitarianism stronger arguments are needed. Will the comparison between the alternative conceptions of the original position provide us with reasons to conclude in favour of one of the two moral theories? This is the question to which I will turn now.

Hare declared that what the hypothetical person would choose in hypothetical choice theories has to be determined by the conditions to which he is subject (Hare, 1975, 87). These conditions must, deductively, determine the choice. But when Rawls’ construction produces the difference principle he shows great embarrassment and suggests that “Rawls may have reasoned that, since an 'economical veil' would make him into a utilitarian, he had better buy a more expensive one.” (Hare, 1975, 91) But Rawls has simply done what he, according to Hare himself, had to do, namely, to deduce the principles from the constraints that characterize his construction of the original position. These constraints are not meant to get specific principles, or to avoid some, but to secure the idea that moral considerations are compelling. The question is how one can choose between these conceptions of the original position. Utilitarians like Harsanyi and Hare argue that, in order to secure the idea that fair principles are designed, the parties in the original position only need to be

---

360 Of course, this argument recoils on Hare, who can be charged with adopting a construction of the original position that delivers him the average principle.
ignorant of their roles as individuals in the real world. With this thin ('economic') veil the normative consequences of the "ideal observer" and the "rational contractor" would be the same (and these normative consequences would be utilitarian). This is so, because, in both constructions only that information in the original position would be excluded that is necessary to achieve impartiality and, together with the application of Laplace's principle, this would lead to the average principle. Rawls disagrees; he acknowledges that under the conditions of the 'ideal and impartial observer' the average principle is likely to be selected, however on the additional condition of sympathetic identification. But in Rawls's description of the original position the representative individuals are characterized by mutual disinterest. Moreover, Rawls uses a thin theory of the good (as contrasted to the 'thick theory of the good' of the utilitarians) and the reason for doing this is that something is good only if it fits into ways of life consistent with the principles of justice already chosen. The theory of the good is, therefore, restricted to the bare essentials. This thin theory about the good is to secure the premises about primary goods that are required to arrive at the principles of justice. "The Kantian doctrine, therefore, proceeds in the opposite direction [compared with utilitarianism]: it starts by allowing to the parties no information and then adds just enough so that they can make a rational agreement. (...) The parties are not to be influenced by any particular information that is not part of their representation as free and equal moral persons. (...) (Rawls, 1980, 549) Utilitarianism, on the contrary, adopts a thick theory of the good because of the focus on maximizing social well-being. What is right is what maximizes the good.

Thus we have two very different models of the original position. One with a thick veil of ignorance and a thin theory

---

361 "In the one case perfect knowledge and sympathetic identification results in a correct estimate of the net sum of satisfaction; in the other, mutual disinterestedness subject to a veil of ignorance leads to the two principles of justice." (Rawls, 1971, 187)
of the good. This construction has to guarantee that the rational deliberations are in such a way embedded in moral considerations that the principles agreed to are reasonable. The other with a thin veil that is combined with a thick theory of the good, and this construction of the original position has to guarantee that the chosen principle is rational; i.e., that social well-being is maximized by giving equal weight to the equal preferences of all individuals. What must be our criteria to choose between these two views?

The utilitarians believe that Rawls’s construction is biased because of its extreme emphasis on risk-aversion. Without this bias he would have assumed equiprobability and when he had assumed equiprobability he would, without doubt, have chosen that principle which would give equal weight to the preferences of all concerned. Did not Rawls admit that he deduced the difference principle from the comparison with the maximin rule and isn’t this the rule a person will apply “(...) for the design of a society in which his enemy assigns him his place.” (Rawls, 1971, 152)?

But Rawls also stipulated that it is essential not to allow the principles chosen to depend on a special attitude toward risk. What must be shown is that choosing as if one had such an aversion is rational given the unique features of that situation. (Rawls, 1971, 172)362 Thus, he does not assume that the parties are particularly risk-averse, but that they, under the constraints to which they are exposed, and because of the extensiveness of the losses they would suffer if they were the least-advantaged, they had good reason to adopt the conservative attitude expressed by this rule. They seemed to have followed the advice of Popper, to be not so much concerned with the maximization of happiness as with the minimization of misery. 363

362 Given the features of the original position, reasonable people would choose principles as if they were highly risk-averse. (Rawls, 1974, 649)
363 In contrast to the model of the impartial observer the parties in Rawls’s original position are in a position of radical uncertainty, for not only do they not know which social position they will occupy, they don’t even know the range of social positions. Neither do they know how bad
Rawls’s conception of the original position results in the difference principle as Harsanyi’s conception leads to the average principle. It is almost impossible to choose between both principles, unless we follow Dworkin and invoke a deeper theory. In the deeper theory that Dworkin presents, utilitarianism is a goal-based theory and Rawls's theory is a rights-based theory. (Dworkin, 1975) Dworkin distinguishes two basic rights: the right to liberty and the right to equality. A right to equality may be described in the following way: individuals have a right to equal concern and respect. Being a highly abstract right, it can take different forms. Someone might argue that it is satisfied by arrangements that provide equal opportunity; or by a system that guarantees equality of income and wealth; or that equality means that the interests of each individual are weighted equally. Rawls’s original position may now been seen as a device for testing these competing arguments. The original position is well designed to enforce the abstract right to equal concern and respect that must be understood as the fundamental concept of Rawls's

the worst positions are and therefore there is no sense in assuming equiprobability. The difference between Rawls’s model and the utilitarian model in this regard is the difference between uncertainty and risk. Ellsberg’s experiments have shown that individuals are not prepared to apply the principle of insufficient reason when they cannot assign probabilities to events. Aversion to uncertainty cannot be captured as risk aversion. Even when probabilities are known it appears that most people are risk-averse when confronted with a situation in which their is a small chance to lose and that they only are prepared to take great risks if there is a small chance to win. This pattern became known as "the common consequence effect". (see chapter III) See also the maximin solution in game theory (note 347).

364 Dworkin suggests that the agreement in the original position "must be seen as a product of a deeper political theory that argues for the principles through rather than from the contract. We must therefore try to discover the features of a deeper theory that would recommend the device of a contract as the engine of justice, rather than the other theoretical devices Rawls mentions, like the device of an impartial spectator (.)." (Dworkin, 1975, 37) According to Dworkin a political theory may be goal-based, right-based or duty-based.
deeper theory. "The right to equal respect is not on this account a product of the contract, but a condition of admission to the original position. (...) This is a right that does not emerge from the contract, but it is assumed, as a fundamental right must be, in its design. (...) Justice as fairness rests on the assumption of a natural right of all men and women to equality of concern and respect, (..)." (Dworkin, ib., 51) The utilitarian position is goal-based. The parties in the original position have to make a choice between alternative social systems; and rational people will choose the social system providing the highest level of average utility.

The preference between the alternative constructions of the original position must be based on a meta-preference for an approach in which what is right is what contributes to the good or an approach in what is right is what is fair.365 For utilitarianism all that matters is aggregate well-being, Rawls’s main concern is the establishment of institutions that contribute to a fair distribution of well-being. In the utilitarian original position claims to goods are secured, while the Rawlsian original position captures the insight that those claims that will later arise must be put forth in relation to institutionalized rules that are fairly established or adopted by agents.

8. Conclusion

Concerning the procedure which the parties in the original position adopt in order to reach a decision, Rawls proposed imagining that they are given a short list of alternative conceptions of justice and that the parties proceed by taking two pair-wise comparisons. "The first choice is between a

---

365 From Rawls' point of view the difference principle reflects the priority of the right in the original position (De Marneffe, 1990) We could imagine that the parties in the original position were to reflect on the question: 'how would you have liked that the other parties had acted, with regard to the definition of the good and its distribution, when it appeared that you ended in the worst-off position in society.'
conception defined by the principle that average utility be maximized and a conception defined by two principles that express a democratic idea of justice.” (Rawls, 1974, 142) The second pair wise comparison is “between the two principles of justice and a variant in which the second principle reads as: social and economic inequalities are to be adjusted so as to maximize average utility consistent with a fair equality of opportunity.” (Rawls, ib., 143)

But in the Revised Edition that appeared in 1999 Rawls writes that were he to write *A Theory of Justice* now, there would be two things he would handle differently and one of these concerned the two comparisons. Now he would let the parties first decide between the two principles of justice, taken as a unit, and the principle of (average) as the sole principle of justice. “In the second comparison, the parties would decide between the two principles of justice and those same principles but for one important change: the principle of (average) utility is substituted for the difference principle.” (Rawls, 1999, xiv). He continues that using these comparisons has the merit of separating the argument for the equal basic liberties and their priority from the arguments for the difference principle itself. “The arguments for the equal basic liberties are at first glance much stronger, as those for the difference principle involve a more delicate balance of considerations.” (ib.,xiv) It is better to recognize that the case for the difference principle is less evident and is unlikely to have ever the force of the argument for the two prior principles.

Barry observed that no other aspect than the effort to show that the difference principle can be derived from the original position, as Rawls specifies it, has attracted more commentary and met with such uniform rejection. The equation of the difference principle with the maximin rule had the effect of giving the impression that the parties in the original position were exclusively concerned with the worst that

---

366 This is, in my view, only slightly different from the original formulation.
could happen to them. Although people had reasons to be risk-averse in this special situation, this negative image has been taken up by his opponents to ridicule the difference principle and the arguments in favour of it.367 This aspect, combined with his criticism of Rawls's construction of the original position, which is quite similar to the criticism of utilitarians, has led to Barry's suggestion to go straight to Rawls' underlying moral intuitions and scrap the formulation in terms of choice in an original position.

Barry proposes taking an alternative argument to defend the difference principle. He suggests taking an equal distribution as just in the first instance: it provides a benchmark. If it is possible to make everyone better off, then it is legitimate to make a move from the benchmark. And indeed, "the difference principle is precisely a principle for picking out one point at the Pareto frontier." (Barry, 1989, 237) Barry's reading of the difference principle, results in just another formulation of the well-known Pareto principle: if everyone gains from inequality it would be irrational not to prefer the situation with inequality to the situation without it. Without invoking the original positions, he comes to a conclusion that is quite similar to Rawls's own: "Injustice, hence, is simply inequalities that are not to the benefit of all." (Rawls, 1971, 62)

This would render the difference principle rather innocuous. Though it morally condemns that the rich get richer to the detriment of the poor and it delivers arguments against exploitation, at the same time it does not exclude huge income inequalities.

367 This is illustrated in the following description of Rawlsian man in the original position, he is: "(..) a strikingly lugubrious creature: unwilling to enter into situations that promises success because it also promises failure, unwilling to risk winning because he feels doomed to losing, ready for the worst because he cannot imagine the best, content with security, and the knowledge he will be no worse off than anybody else because he dares not risk freedom and the possibility that he will be better off (..)." (Barber, 1975, 299)