

E 20709 F
SONDERDRUCK



Vol. 80 · 1994
2. Quartal · Heft 2

ARSP

*Archiv für Rechts-
und Sozialphilosophie*

*Archives de Philosophie
du Droit
et de Philosophie Sociale*

*Archives for Philosophy of
Law and Social Philosophy*

*Archivo de Filosofía Jurídica
y Social*

Franz Steiner Verlag Stuttgart



Luc Bovens, CU at Boulder

Rawls on Mutual Disinterest and Hume's Subjective Circumstances of Justice

I

J. Rawls is in search of a set of general principles of justice which can guide the particular decisions involved in organizing a just society. In order to attain these principles he sets up a hypothetical construction, viz. the original position (OP). We are asked to imagine that in the OP a group of rational persons who will engage in social cooperation are to decide in advance and under certain constraints upon a set of general principles of justice for their own future society. The principles they will choose are the principles to which a just society must conform. *Mutual disinterest* is one of the constraints which Rawls imposes on the OP. His characterization of this constraint is puzzling at best.

On the one hand Rawls claims that to say that the persons in the OP are mutually disinterested is not to impose certain constraints on the contents of their ends: '[It] does not mean that the parties are egoists, that is, individuals with only certain kinds of interests, say in wealth, prestige, and domination' (Rawls, p. 13).

On the other hand Rawls imposes Hume's 'circumstances of justice' on the OP. Hume argues that questions of justice only arise within particular circumstances:

'tis only from the selfishness and confin'd generosity of men, along with the scanty provision nature has made for his wants, that justice derives its origin. (Hume, p. 495)

Rawls translates 'the selfishness and confin'd generosity of men' into 'mutual disinterest' (or 'the subjective circumstances of justice') and 'the scanty provision nature has made for his wants' into 'scarcity of goods' (or 'the objective circumstances of justice').

There is a contradiction in that Rawls originally claims that to say that the persons in the OP are mutually disinterested is *not* to say that they are egoists or have particular types of interests while the equivocation of mutual disinterest with Hume's subjective circumstances of justice *does* suggest selfishness – and, consequently, a particular type of interest – on the part of the persons in the OP. I will try to resolve this contradiction by arguing that there are two conceptions of mutual disinterest at work in Rawls' theory of justice which each play a distinct role in the OP.

II

Rawls claims that to say that the persons in the OP are mutually disinterested is to say that they 'try to acknowledge principles which advance their system of ends as far as possible' (p. 144). But what are the ends of the persons in the OP? What motivational features can we reasonably impose on persons who are placed behind the veil of ignorance of the OP – i.e. who have no knowledge of their identities?

The constraints on the original position must be 'reasonable' and they are reasonable only if they are 'widely accepted' (Rawls, p. 18 and p. 20). I will argue in this section that there exists a particular motivational feature which qualifies as a widely

accepted constraint and matches a reading of mutual disinterest that does not make the parties in the OP into selfish individuals. Hence, on this reading, mutual disinterest cannot be equivocated with Hume's subjective circumstances of justice.

The OP is a hypothetical construct in which I, as a member of a group of rational persons, have knowledge of a set of possible worlds. Within these possible worlds, I occupy different social positions. I know that exactly one of these possible worlds will be actualized but do not know which one. This epistemic position is shared with the other persons in the group. What does it mean to say that a constraint on the OP is a widely accepted constraint? I propose the following interpretation:

The constraints imposed on the hypothetical construct of the original position are widely accepted if and only if we stipulate similar constraints on the hypothetical constructs that play a role in common moral reasoning concerning justice.

What hypothetical constructs play a role in common moral reasoning about justice? I take the following to be a plausible *guiding question* in common moral reasoning with a level of theoretical sophistication similar to, say, the Golden Rule:

In deciding whether a projected action is just, I ask myself whether I would take issue with the action if I could be just any of the persons affected by the action.

This guiding question contains a hypothetical construct. In this construct I consider the set of possible worlds in which I respectively occupy the place of all of the persons affected by the action in question. I then ask myself whether I would take issue with this action if it were the case that just any of these worlds might be actualized.

The following problem now comes up. If I am to take issue with any action at all in this hypothetical construct, I have to assume that the persons whose identity I will take on in these possible worlds are motivated by certain ends. But what assumptions do we actually make concerning these ends in common moral reasoning?

Before considering this question, I need to make some conceptual stipulations as to the relation between ends and (objective) interests:

- (a) Some person's interest becomes an end for person P, if and only if,
 - (i) P recognizes this interest and
 - (ii) P intends to act upon this interest;
- (b) Unless a particular interest becomes an end for P, it cannot motivate P's action.

It follows from this that the set of P's interests may differ from the set of P's ends in the following ways:

- (1) P may not recognize her interests or may not intend to act upon her interests;
- (2) P may take it as her end to promote or to counteract some other person's interest.

I will now take up my question again. I am asked to consider the set of possible worlds in which I respectively occupy the positions of the various persons affected by the action in question. But what set of ends can I reasonably stipulate for myself in making these identifications? I can see the following possibilities:

- (a) I may be motivated by my own actual set of ends;
- (b) I may be motivated by the actual set of ends of the persons whom I identify with in each possible world;
- (c) I may be motivated by the set of ends which match the actual interests of the persons whom I identify with in each possible world.

The imposition of stipulations (a) and (b) on our guiding question yields counterintuitive results. As to stipulation (a), we merely need to suppose that I would have an

actual set of ends that is consistent with desiring the state of affairs that ensues from some injustice as usurper as well as victim. In identifying with the usurper, I would enjoy, say, the advantages accrued, while in identifying with the victim I would enjoy, say, the attention of being put in the spotlight. As to stipulation (b), it is sufficient to suppose that the victim of some injustice does not recognize her interests, say, because she adopts the usurper's interests as her own ends – as is a common theme in Marxist and feminist literature.

Notice that our guiding question yields counterintuitive results on stipulation (a) and (b) due to a discrepancy between the ends that I adopt in identifying with the persons affected by the action and their actual interests. This suggests that stipulation (c) may provide for a better picture. On this stipulation (1) I do take a person's actual interests to be crucial, and, (2) I do not consider a person's actual attitude towards her own or other persons' interests to be relevant in deciding questions of justice. If we build in stipulation (c) in our guiding question then the victim's end of being put in the spotlight becomes irrelevant – assuming that this end does not match her actual interests. Similarly, the victim's ends matching the usurper's interests become irrelevant as well – assuming that these ends do not match her actual interests. We can thus make our guiding question more in accordance with common intuitions by building in stipulation (c):

In deciding whether a projected action is just, I ask myself whether I would take issue with the action if I could be just any of the persons affected by the action and would in my respective identifications be motivated only by the ends which match the actual interests of these respective persons.

The hypothetical construct we employ in common moral reasoning concerning justice is thus constrained by a very particular motivational assumption: In my identification with a particular person, my ends must match the interests of this person in the actual world or, in other words, I must take an interest in this person's actual interests. Hence, the motivational assumption that a person takes an interest in her own interests qualifies as a widely accepted constraint to impose on Rawls' OP.

I take this widely accepted constraint to be a more detailed description of Rawls' constraint of mutual disinterest: the rational persons in the original position are mutually disinterested if and only if in their identifications with a token person of a social group in the actual world they are motivated only by the ends which match the respective interests of these token persons. They thus differ from the persons in the actual world in their motivational structure in the following respect: In identifying with a token person P they take an interest in P's interests including the interests which did not acquire the status of an end for P in the actual world and they do not take an interest in any but P's interests. On this reading, mutual disinterest is a stipulation to the effect that the rational persons in the OP will take an interest in the actual interests (whatever they may be) of the token persons whom they identify with but it is not an assumption concerning the content of these ends.

This reading of mutual disinterest as a motivational feature matches Rawls' earlier claim that this constraint does not stipulate that the persons in the OP are 'egoists' or 'individuals with only certain kinds of interests'. Yet it is not consistent with Rawls' equivocation of mutual disinterest with Hume's subjective circumstances of justice. In the next section I will consider what role Hume's circumstances of justice play in Rawls' OP and whether there exists an alternative reading of mutual disinterest which respects this equivocation.

III

Let us first consider the objective circumstances of justice. I will construe a piece of quasi-moral reasoning which resembles the original position. Suppose we decide to play a game of cards of which we only have a general notion (cf. Dworkin, p. 69). Before the cards are being dealt we have to formulate a set of fair rules. What are the relevant objective features we have to take into account? Here are a few examples: We need to consider the circumstances under which we are planning to play this game; we need to consider how many decks of cards we will have at our disposal; we need to consider the number of players that will participate. What I wish to point out is that it is irrelevant whether these circumstances obtain *while we are deciding*. Rather, we need to know under which circumstances we will *actually* play our card game.

The OP is an hypothetical construct in which principles of justice are decided for the actual world. The situation is analogous in the following respect: It is irrelevant to stipulate that the condition of 'scarcity of goods' obtains within the hypothetical construct in which the decision for the principles of justice is made. Rather, we need to stipulate that the rational persons in the OP *know* that the condition of scarcity of goods obtains within the actual world, i.e. the world in which the principles of justice they choose for will obtain. It is in this sense that I propose to read Rawls' claim that the circumstances of justice 'characterize' (p. 118) or 'are included in' (p. 130) the OP.

I suggest the same interpretation for the subjective circumstances of justice. This interpretation can be supported similarly by the card-game analogy: In deciding upon fair rules we need to know that the players have such and such features (say, they will be utterly concerned with maximizing their own benefits) *while playing the game*. Whether these motivations obtain during the process of deciding is irrelevant. Consider Hume's more detailed description of the subjective circumstance of justice:

There are (...) particulars in our natural temper (...) which are very inconvenient and are even contrary to the requisite conjunction (of our forces in society). Among the former I may justly esteem our selfishness to be the most considerable. I am sensible, that, generally speaking, the representation of this quality have been carried much too far (...) I am of the opinion that tho' it be rare to meet one, who loves any single person better than himself, yet it is as rare to meet one in whom all the kind affections taken together, do not over-balance all the selfish. (Hume, p. 486–487)

The subjective circumstances of justice refer to a condition in which the character feature of being moderately selfish is commonly shared. Analogous to the objective circumstances of justice, we need to stipulate that the rational persons in the OP know that this condition holds in the world in which the principles of justice they choose for will obtain.

Rawls equivocates mutual disinterest with Hume's subjective circumstances of justice. In the previous section, I have argued for a particular reading of mutual disinterest which passes as a widely accepted constraint on the OP. This reading did not square with Rawls' equivocation. I argued that 'mutual disinterest' refers to a particular motivational feature of the persons in the OP. This is quite remote from the claim that 'mutual disinterest' refers to an item of information that is not shaded from the knowledge of the persons in the OP about the world for which they are choosing principles of justice. I also presented the following definition: a person is mutually disinterested if and only if she takes an interest in her own interests. But this definition does not stipulate that these interests must be of a particular kind viz. moderately selfish interests – as would be consistent with Hume's subjective circumstances of justice.

I want to contend that 'mutual disinterest' stands for two distinct notions in Rawls' theory of justice and that both notions play a very different role in the original position. Consider the following two definitions:

- (a) a person is mutually disinterested₁ if and only if she takes an interest in her own interests (whatever these may be);
- (b) a person is mutually disinterested₂ if and only if she takes an interest in at least some selfish interests.

Mutual disinterest₁ is the weak motivational assumption we impose on the hypothetical constructs that we employ in common moral reasoning. Mutual disinterest₂ is a stronger notion in that it assumes certain features concerning the content of a person's interests. It is only mutual disinterest₂ that matches Hume's subjective circumstances of justice.

What role do both notions play in the OP? Mutual disinterest₁ is a minimal assumption concerning the motivational features which characterize the rational persons in the OP in their identifications with token persons in the actual world. In identifying with the token persons in the actual world they take an interest in the actual interests of these token persons. Mutual disinterest₂ is an item of information about the persons in the actual world for which the persons in the OP are choosing principles of justice and this item is not shaded from their knowledge by the veil of ignorance. Thus, mutual disinterest₁ is a stipulation as to how the persons in the OP *are motivated* in choosing principles of justice. Mutual disinterest₂ is a stipulation as to what the persons *know* about the world for which they are choosing principles of justice.

It is important in its own right to determine what conception of mutual disinterest Rawls has in mind at the various junctions in the text. Furthermore, disambiguating this notion counters a common objection that there is no reason to accept principles of justice that are chosen by rational egoists. The persons in Rawls' OP are not rational egoists. Rather, in identifying with the token persons in society they make the actual interests of the token persons into their ends and they *know* that they are choosing principles of justice for a society in which moderate selfishness reigns. But this is very remote from stipulating a selfish disposition for the persons in the OP. Hence, the objection is contingent on a natural misreading of Rawls' notion of mutual disinterest and can be readily dispelled by disambiguating this notion into a motivational and an epistemic constraint on the OP.

References

- Dworkin, R. 'Hypothetical Contracts and Rights'. In: J. Sterba (ed.) *Justice – Alternative Political Perspectives*. (Wadsworth, Belmont 1980)
- Hume, D. *A Treatise of Human Nature*. (Clarendon Press, Oxford, 1888)
- Rawls, J. *A Theory of Justice*. (Harvard University Press, Cambridge, MA, 1971)