

**RATIONALITY AND THE MORAL SENTIMENTS :
SOME ANIMADVERSIONS ON A THEME IN
*THEORY OF JUSTICE***

Kai Nielsen

**After all, the sense of justice is
a settled disposition to adopt and
to want to act from the moral point
of view insofar at least as the
principles of justice define it.**

**... the liability to moral feelings
seems to be as much a part of the
natural sentiments as the tendency
to be joyful, or the liability to
grief.**

John Rawls

I

Moral philosophers have repeatedly argued that there is an intimate bond between rationality and morality. It is pervasively believed that a thoroughly reasonable person must be a morally committed human being; amoralism or indifference to morality must involve some failure of rationality : a rational human being must also be a morally committed human being.

John Rawls in his monumental *A Theory of Justice* defends this traditional bit of moral rationalism. I shall display the core of Rawls' argument for the claim that reason determines the ends of life in such a manner that it must be the case that the through and through rational person will also be a person of moral principle. I shall then proceed to argue that Rawls has not shown this to be so and that,

Rawls apart, such moral rationalism, though appealing, remains a very problematic matter. In making this assessment of Rawls, I shall first very schematically characterize some of the core features of Rawls' moral theory, including his account of rationality. I shall then turn to those features of his account designed to establish this key thesis of moral rationalism.

As reviewers have repeatedly stressed, *A Theory of Justice* is a powerful, seminal and intricately orchestrated work which will be discussed many years hence. It is a massive, deeply reflective and complex work with a rondo-like and often confusing structure. These features practically insure that the initial appraisals, many of which have been reprinted in Norman Daniels' *Reading Rawls*, will be first approximations in a continuing dialogue.

A Theory of Justice is no less than an attempt to articulate and justify a conception of the principles of justice, together with an account of the values of community and an ideal of the human person, which will provide "an Archimedian point for judging the basic structure of society".¹ (584) Yet, its ambition and scope notwithstanding, the fundamental idea in Rawls' theory is not at all complex: it is the idea that the principles of justice, and indeed the principles of morality generally, are the principles which free and equal rational persons would come to agree on (mutually accept) as the principles which are to regulate their lives together. He takes the central problem of moral and social philosophy to be the problem of "how society should be arranged if it is to conform to principles that rational persons with true general beliefs would acknowledge in the original position", i.e. (roughly) in a position of impartiality (547).

Rawls assumes that by sustained rational inquiry we can come to specify principles which will render determinate the rationale for and the proper assignment of rights and duties and the distribution of benefits and burdens among people whose conflicting interests require resolution in some morally acceptable manner. The principles of justice which he believes would be the outcome of such a rational inquiry and would be the principles which rational persons would choose were they to choose impartially, and where conditions of life are not too harsh, are the following: 1) "each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar liberty for all" and 2) "social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged, consistent with the just savings principle and (b) attached to offices and positions open to all under conditions of fair equality of opportunity". (302) In moderately favorable circumstances, these principles, together with

true factual knowledge, are taken to give us an Archimedean point to assess the morality of institutions and social designs. In such circumstances they are to be taken strictly in their order of priority, i.e. the second principle is to be satisfied only if the first one has been satisfied. The aim is to arrive at a scheme which benefits all and in which the only morally acceptable inequalities are those which are, everything considered, to the advantage of the worst off strata of the society. Rawls would not accept as just a utilitarian principle maximizing either total or average utility, for this might involve sanctioning as just representative members of the society ending up with lower life chances for the benefit of others. Rational persons judging in ignorance of their particular positions in society and judging impartially would not accept such utilitarian principles. They would instead accept, for moderately favorable circumstances, the two principles of justice stated above and would in less favorable circumstances fall back on what Rawls calls a more general conception of justice, to wit the principle that all "social primary goods — liberty and opportunity, income and wealth, and the bases of self-respect — are to be distributed equally unless an unequal distribution of any or all of these goods is to the advantage of the least favored". (303).

These, Rawls would have us believe, are principles all rational human beings would choose where they have full general knowledge of society and human life and where they must be impartial. But to assess this claim and to make headway with the central problem of this essay, we need to have some understanding of what Rawls is talking about when he speaks of rationality.

II

Rawls construes rationality in the self-consciously minimal and antiseptic way in which it is usually understood in bourgeois economic theory. He does not start out by stipulatively defining 'rationality' or by elucidating the concept but proceeds indirectly. We find out whether a person's interests and aims are rational by finding out whether "they are encouraged and provided for by the plan that is rational for him". (409) We, in turn, find out whether a person's plan of life is rational for him by finding out whether "(1) it is one of the plans that is consistent with the principles of rational choice when these are applied to all the relevant features of his situation, and (2) it is that plan among those meeting this condition which would be chosen by him with full deliberative rationality, that

is with full awareness of the relevant facts and after a careful consideration of the consequences". (408) Thus for Rawls, in determining the rationality of action and of moral conduct and generally the rationality of our aims and interests, it is essential to determine what are the *principles of rational choice* and what it is to choose with full *deliberative rationality*.

Rawls, in keeping with his general philosophical program, does not analyze the concept of rationality anymore than he analyzes the concept of justice. He gives the principles of rational choice, as he remarks, "by enumeration so that they replace the concept of rationality". (411) The cluster of principles, which are the core principles of rational choice and are, as Rawls puts it, the most central "aspects of rationality", are the following : (1) *the principle of effective means*, (2) *the principle of inclusiveness*, and (3) *the principle of the greater likelihood*. They are mutually compatible principles, all of which Rawls takes to be tolerably unproblematic. (411-5). They would be a part of any coherent conception of rationality. Stated cryptically they are the following :

1. Given a determinate objective, it is to be achieved with the least expenditure of means or, given the means, the objective is to be fulfilled to the fullest extent. (411-2)
2. Of two or more alternative plans, that plan is to be preferred which would most extensively achieve the desired aims of the other plans and in addition its own desired aims. (412)
3. Where two plans are closely similar, then, *ceteris paribus*, the plan with the greater likelihood of success is to be favored. (412-3)

It is important to realize that in using these principles we do not address the rationality of our wants or desires *per se*. The principles are principles of instrumental rationality enabling us to estimate most adequately how we are most likely to be able to maximize the satisfaction of these wants or realize most fully the maximum of our

aims. The *principle of inconclusiveness* combines with the *principle of effective means* to “define rationality as preferring, other things equal, the greater means for realizing our aims, and the development of wider and more varied interests assuming that these aspirations can be carried through”. (413) However, Rawls does not limit his principles of rational choice to the above three principles. He also takes as an essential aspect of rationality the following :

4. We are to take that course of action which is most likely to realize our most important aims.

In addition to (4) and the other above ‘counting principles’, Rawls adds two further at least putatively unproblematic principles of rationality

5. *Ceteris paribus*, in choosing like plans, an individual is to choose those plans which will best advance his or her interests. (142)
6. *Ceteris paribus*, people are to try to protect their liberties, widen their interests and enlarge their means of promoting their aims, whatever they are. (142)

Lastly, as a final principle of rational choice, we should enunciate what might be called the *principle of rational self-development*.

7. Given that, *ceteris paribus*, people tend to prefer activities that depend upon a large and more complex repertoire of realized capacities to activities that depend on a smaller and simpler repertoire, we are to “realize and train mature capacities”. (428)

These are the core principles of *rational choice* enunciated by Rawls. They, as I have remarked, together with *deliberative rationality*, determine membership in the class of rational life plans. And it is only in the light of a life plan, which has been found to be rational, that an individual’s aims and desires can correctly

be said to be rational. That is to say, to be rational they must be in accordance with such a plan of life. If they do not square with a plan of life which is rational for him, then they will not be rational aims or desires. (We might very well have Parfitian worries about this last cluster of claims.)

However, a rational life plan must not only be in accordance with the principles of rational choice, it must also not contravene *deliberative rationality*. By the latter Rawls means a way of reasoning and acting. A person committed to such a way of reasoning and acting will choose that plan of action, consistent with the principles of rational choice, which he would decide upon "as the outcome of careful reflection in which [he] reviewed, in the light of all the relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realize his more fundamental desires." (417)

We have here what is fundamentally a very Humean-Russellian conception of rationality. Of such a conception it could be justly argued that it does not capture all of what it is we are talking about when we speak of rationality. Rawls, however, is well aware of that but he also realizes that many of the *other* aspects of our common conception of rationality are problematic and morally and perhaps even ideologically non-neutral. If we used instead, in a context similar to that in which Rawls employs his conception of rationality, a morally non-neutral and indeed problematic conception of rationality, we would hardly have a widely accepted base from which we could in some sense derive or ground moral principles of an admittedly problematic sort or preserve the possibility of establishing that taking the moral point of view and rejecting amoralism is something that reason requires.

I could, and on another occasion will, criticize Rawls' conception of rationality, but here I shall accept it as it stands and endeavor to show that, given that conception or any related plausible antiseptic account of rationality, Rawls has not established that amoralism must be irrational or that an amoralist need, because of his rejection of morality, show any loss or diminishment of rationality. Rawls' conception of rationality, even on its own terms, could be refined but that refinement will not make an essential difference to the arguments I shall make about morality and amoralism.

Rawls takes it as a crucial task, in thinking about morality, to determine "whether being a good person is a good thing for that person, if not in general...at least [in the circumstances] of a society well-ordered or in a state near justice..." (397-8) His claim is the rather bland one, "that being a good person is indeed a good". (398)

I say 'bland', for it is a) truistic and b) it does not even suggest an answer to the related and more perplexing question that has exercised some philosophers, namely, whether a human being in such a society who is a through and through rational and informed person must also be a good person (a humane person of moral integrity and decency). In an ordinary way we can, as Rawls acknowledges, assess the rationality of a person's desires. (407-8) Given this capacity, what we want to know is whether in a well-ordered society an amoralist must be an irrationalist or whether in being an amoralist he must have irrational desires. Is his not desiring to be a just man — a man of moral principle — but rather simply desiring to be a man of good morals an irrational desire ?

III

Let me begin this central task of this essay by putting the above problem in Rawlsian terms. Is a desire to act justly regulative of anything that would count as a rational life plan ? (456) Is acting justly part of any person's good ? In answering these questions, part of the task is to show "how justice as fairness generates its own support". In Chapter VIII of his *A Theory of Justice*, Rawls sets himself this task.

If we are to have a well-ordered society with stable institutions, it is crucial, Rawls contends, that there be a shared sense of justice within that society. Recognizing that "a moral view is an extremely complex structure of principles, ideals, and precepts, and involves all the elements of thought, conduct and feeling", Rawls stresses that "many kinds of learning ranging from reinforcement and classical conditioning to highly abstract reasoning and the refined perception of exemplars enters into its development". (461) In showing how justice as fairness generates its own support, we need a reasonably realistic account of how moral development would occur in a well-ordered society in which such principles are instantiated. [We should recall that Rawls characterizes "a well-ordered society as one designed to advance the good of its members and effectively regulated by a public conception of justice". (453)]

Rawls attempts a sketch of moral development. The first stage of moral development, he refers to as the *morality of authority*. It is in this form that morality first becomes a reality for children, though elements of this form of morality carry over into our adult lives. (462) Children in being introduced to morality are not in any position to assess the validity of the precepts or injunctions

addressed to them by those in authority. Children, when they are very young, cannot reasonably doubt the propriety of parental injunctions: without an understanding of the rationale for the injunctions and without the needed background information, they must, to act rationally, simply submit themselves to the *authority* of their parents. Their parents may very well have biased and distorted moral and social views, but the children are in no position to recognize that. Though, we should remember, in a well-ordered society this unfortunate condition of their parents would not (by definition) obtain. (454)

In the context of moral learning, love and trust between parent and child is central and essential, for without it a child is unlikely to accept their moral authority. And without this familial relation, unless he is fortunate enough to have effective and rather constant parent surrogates, he is not likely to gain any effective moral understanding at all. Where there is no such love and trust, he will sense that his parents have a power over him and he will fear them and hence obey them, but he will not accept their authority, where 'authority' has any sense other than their 'power to constrain and instil fear'. He will not, at least *vis-à-vis them*, develop a sense of what it is for there to be a moral or a *de jure* authority. Indeed it is unlikely in such a circumstance that he will develop any such sense in any very effective manner. This is why relations of love between parent and child are so crucial to a well-ordered society. They are, in short, essential empirical conditions for the reciprocity between human beings essential for justice.

So, for moral learning and development to take place, parents must first come to manifest love to the child, meaning very centrally by that that they will "be concerned for his wants and needs" and, most crucially, to affirm his sense of worth as a person. (464) Initially, a child's actions are motivated by "certain instincts and desires" and to the extent that his aims are regulated at all they are motivated by a limited rational self-interest. In a well-ordered society, his parents' relations to him are such that they will do for him as his rational self-love would incline him towards, where this rational self-love is compatible with the principles of justice. (463-64)

Looked at normatively, for the morality of authority to be a genuine part of morality, it must be subordinate to the principles of right and justice. Where this obtains, the parents' authority is, given the child's distinctive circumstances, a legitimate moral authority. The steps whereby we come to introject the moral point of view reproduce the structural scheme of morality. The claim is that

children have a 'morality of authority' and that an appeal to authority has a genuine if limited place in the moral firmament. (461)

The second stage of moral development is what Rawls calls "*the morality of association*". It is, in short, the morality of my station and its duties. The content of this morality is "given by the moral standards appropriate to the individual's role in the various associations to which he belongs". (467) It, of course, takes many forms, since after all there are many associations and roles into which individuals enter. In a well-ordered society, and that *ex hypothesi* is what we are talking about, the principles of justice will regulate the ideals governing these varied associations. (472) In talking of associations, we are talking about everything from the family, school, neighborhood, professional bodies, sports associations and the like up to the community as a whole. (467-68) But in a well-ordered society everyone will have the role of citizen and there will be a full equality of citizenship; everyone is meant to have political views concerning the common good. Thus there will be in such a society "a morality of association in which the members of society view one another as equals, as friends and associates, joined together in a system of co-operation known to be for the advantage of all and governed by a common conception of justice". (472)

Starting with the family itself, one finds oneself in various associations, embedded in certain social structures, in which each member has certain rights and duties. Indeed, as Dahrendorf has argued, we wouldn't even have a social structure or a society if this did not obtain.² In a well-ordered society, a child, trusting his parents, is simply taught by his parents to do the proper thing in the various associations into which he willy-nilly enters. He is taught *what* it is to be a member in good standing in such associations and he is taught *to be* in such good standing. Children learn what it is to be a good daughter, student, mother's helper, companion, sport, choir-boy, neighbor and the like. And indeed, as Rawls points out, our moral understanding increases as we move in the course of life, through a sequence of stations with its attendant duties. And in doing this we will come to have certain ideals appropriate to those roles.

As our understanding of ourselves and our society increases, we will move beyond a morality of association in the direction of having a *morality of principles* as we work out "a conception of a whole system of co-operation that defines the association and the ends which it serves". (468) But in doing this, we will of necessity come to understand the roles of others and to see things from their

perspective. Without this we can have little in the way of moral understanding. This leads us, in a way a child operating exclusively under a *morality of authority* will not do at all, to take note of the importance in morality of motives and intentions. (469)

What generally must obtain for there to be a functioning cluster of associations in a well-ordered society? One central thing is that there must be relations of friendship and mutual trust, where people can rely on one another to do their part. There must be developing relations of *reciprocity*. The situation of an agent entering into an organization in a well-ordered society is parallel to that of a child in his family in such a society. Finding people well-disposed toward him and willing to "live up to their duties and obligations" in the associations into which he enters, he develops, as a matter of psychological fact, "feelings of trust, and confidence" *vis-à-vis* them.

And he acquires attachments to them and a desire to live up to the ideals of the associations. The system will be stable where participants in a system of social co-operation regularly act with evident intention to preserve its just (or fair) rules and where bonds of friendship and mutual trust develop among them, thereby holding them ever more securely to the system. (472)

The principles of justice will be part of the ideals of many of the more complex associations and since this is so, people in those associations move very easily and naturally to a morality of association of the community as a whole, where everyone is viewed as a member of a society of equals and where, as associates, they are joined together in "a system of co-operation known to be for the advantage of all and governed by a common conception of justice". (472) Here the key virtues are "justice as fairness, fidelity and trust, integrity and impartiality". (472) The principal vices are "graspingness and unfairness, dishonesty and deceit, prejudice and bias". (472) Such moral attitudes, Rawls contends, are bound to exist when people become attached to those, who co-operate with them in a fair scheme. (472)

The third and final stage of moral development is labelled by Rawls as the *morality of principles*. In it he tries to account for "the process whereby a person becomes attached to these highest order principles themselves". (473) He recognizes, even in the more complex forms of the morality of association, where a concern develops for the equal citizen, that there is an acceptance of the principles of justice. But our complying with them at such a stage of development is not because we, on reflection and with vivid awareness, simply want to act justly and advance just institutions, but the motive for acceptance "springs largely from... ties of

friendship and fellow feeling for others". Rawls wants to understand how we become attached to the highest order moral principles themselves and come to accept them as intrinsically desirable. This means that our moral attitudes "are no longer connected solely with the well-being and approval of particular individuals and groups, but are shaped by a conception of right chosen irrespective of these contingencies". (475) The *morality of principles* includes, of course, the virtues of the morality of authority and association. But it is more as well, for it organizes the ideals of these moralities "into a coherent system by suitably general principles". (478) We at this stage of moral development become fully mature moral agents. As such we not only want to be cooperative and attain approbation from those around us, we wish, as well, to be *just persons*. (473) We come to recognize how social arrangements in accordance with the principles of justice have promoted our own good and the good of those with whom we are associated and this tends to engender in us "a desire to apply and to act upon the principles of justice". (474) And the having of this desire leads us beyond simply caring about the welfare of those to whom we are bound by particular, family-molded, association-molded, ties of fellow feeling. The building up of such sentiments is crucial in morality. Friendship and the ties of association are not enough, for "while every citizen is a friend to some citizens, no citizen is a friend to all". (474) Our moral psychologies need to be developed to the point where there is the general "acceptance of public principles of justice," the common allegiance to which "provides a unified perspective" from which we can adjudicate our differences. Our psychological development naturally leads to the having of such a *morality of principles* and the having of such principles by human beings is essential to our individual and collective well-being.³

IV

What seems puzzling is how we come to desire, *for its own sake*, to promote just institutions or to act in accordance with the principles of justice. Would they not always be principles which were desired for some other end? They, of course, could be desired for some other end and for themselves as well. Yet it is the latter which remains puzzling. How is it that we come to "want to do our part in maintaining" just arrangements even when doing so may not be beneficial to people we happen to care about, including ourselves? A man with a sense of justice will have such wants, but how does he

come to have them and sustain them? That is to say, *how* does he become and remain a man with a *sense of justice*? *It is also the case that a man with a sense of justice is a man who will be willing "to work for (or at least not to oppose) the setting up of just institutions and for the reform of existing ones when justice requires it"*. (474) In a well-ordered society, we simply desire to do or have these things. We in short have a pro-attitude toward morality and indeed value it for its own sake. (527).

How is it that we come to have these desires? Rawls seems to rely here principally on associationist psychological principles or a kind of Skinnerian reinforcement.⁴ He remarks that once attitudes of love, friendship and mutual trust have been generated, the very recognition that we and those for whom we care the beneficiaries of established and enduring just institutions tends to produce and continue to reinforce in us the corresponding sense of justice. That is to say, once we have these other dispositions and we live in a society with just institutions, we will come to have a *sense of justice*. We develop a desire to apply and to act upon the principles of justice once we realize how social arrangements answering to them promote our good and that of those with whom we are affiliated. "In due course we come to appreciate the ideal of just human co-operation". (474) In such circumstances, when we do not live up to these principles, we also feel guilt. (Pace Skillen and Collier we need not and indeed should not believe that such feelings of guilt must be irrational.)

It looks at least as if this prizing of morality *for its own sake* is just a brute fact about us and, furthermore, it does not look — someone might argue — as if Rawls has even unearthed an explanatory reason let alone a justificatory reason why this should be so.⁵ We (or rather most of us) just want to be moral. I do not wish to make such a strong claim, but I shall argue that Rawls has not shown that there is or can be a reason or set of reasons for prizing morality for its own sake which are sufficient to undermine a consistent amoralist's challenge by showing the irrationality or even the inferior rationality of the consistent amoralist's alternative point of view. This may be one of the points where the giving of reasons comes to an end. This is particularly evident if we stick to Rawls' rather antiseptic conception of rationality.

There is, however, a passage in which Rawls does try to provide us with such a rationale and to show that "the desire to act justly is not a form of blind obedience to arbitrary principles unrelated to rational aims". (476) Put just like that, with the umbrella phrase 'unrelated to rational aims', it may be little better than a truism on

Rawls' part. But put somewhat more strongly, as I think is his intent, the claim would come to the contention that if a man has rational aims and is in a well-ordered society, then he will just desire — and for its own sake — to act justly : a perfectly just society must be part of an ideal that *any rational* man will have, if he has full knowledge and vivid awareness of what is involved.

But his argument why this must be so does not seem to me very clear or convincing. (Indeed, as we shall see, he even makes remarks of his own which hardly square with that claim.) The principles of justice are prized because rational persons can see that it is in their interests to have them when they need to adjudicate competing claims. But this seems only to show their *instrumental* value. It does not show why a rational individual has a sound reason for prizing them for their own sake. It shows why he will want them to be the principles of adjudication between people, but it does not show why he will always want to use them himself rather than always just appear to use them when it is in his interest to do so. That is to say, a rational individual will never want to be *seen* to be unfair and he will see the overall value of principled behaviour in society. That is, he will see why rational individuals will want people to act in accordance with the principles of justice. But it is not so clear that a rational person *must* actually want to be fair, when it is not in his interest to so act, as distinct from wanting simply to seem to be fair. And it is not clear that it is always in a rational individual's interest to be fair. My point is that while it is plausible enough to claim that for rational persons generally it will advance their collective interests if they act in accordance with the principles of justice, it does not follow that it will always be in the rational self-interest of any given individual to be fair or to desire to be fair in every situation in his life where competing claims and questions of justice arise. Rawls typically (but not always) seems to think that being fair will be what reason (rationality) requires of him, but — as far as I can see — he has not given any persuasive grounds for this belief. He has not shown that being fair is something we should do for its own sake, that it is desirable for its own sake or that it is something to be wanted in itself.

Rawls claims — and I agree — that it is a first class blunder to maintain “that the highest moral motive is the desire to do what is right and just simply because it is right and just, no other description being appropriate”. (477) Such a doctrine “of the purely conscientious act”, Rawls holds, “is irrational”. (477) It makes the sense (sentiment) of right and the sense (sentiment) of justice into something quite arbitrary, something without rhyme or reason.

Rawls, of course, believes that this sentiment of justice is not a different desire from that of acting “on principles that rational individuals would consent to in an initial situation which gives everyone equal representation as a moral person”. (478) And, in a very Kantian fashion, he takes that to be equivalent to “wanting to act in accordance with principles that express men’s nature as free and equal rational beings”. (478) We need not à la Ross or Pritchard take the sense of right to be a desire for something which is unique and unanalyzable. Alternatively, and by a de-mythologization or at least de-mystification, we can construe it, as Rawls does, in terms of a *desire to be rational*: to “want to live with others on terms that everyone would recognize as fair from a perspective that all would accept as reasonable”. (478)

I shall make two general comments about this claim. First, the acceptance of the above is perfectly compatible with saying as well that being just is something to be wanted for its own sake, desirable in itself, i.e. intrinsically and not just instrumentally desirable. That the desire to do what is right could only be properly characterized in those Rossian terms could be quite mistaken, even incoherent, while it could very well still be true that being just is something which has intrinsic worth — something which a rational person would want for its own sake. This being so my questions developed two paragraphs back are quite intact. Second, it may — I do not say it actually is — be clear enough that Rawls’ principles of justice are principles “everyone would recognize as fair from a perspective.. all would accept as reasonable” and it still could be the case that an individual (any individual at all) could recognize that while that is the *collectively* reasonable thing to do, it still is not necessarily the rational thing, at least in certain circumstances, for him (as an individual) to do and thus he might *not* come to believe or (more likely) to continue to believe that to be a rational person, it must be the case that he desires to be *just*, particularly where that is construed as something to be wanted for its own sake. He might very well believe instead that when he considers things strictly from his own point of view it would depend on the circumstances whether being just is something which is everything considered desirable.

It no doubt is the case that for most of us — though not for the characters in *Last Exit to Brooklyn* — “among our final ends are the attachments we have for persons, the interests we take in the realization of their interests, and the sense of justice”. (494) But the truth of that at least putative sociological fact is perfectly compatible with its not being the case that there must be a failure in *rationality* in an agent, if he did not desire to be — let alone to strive to be — a

just person where (to put it crudely) being just did not pay.

Perhaps in going through the stages of moral development Rawls outlines — a psychological development necessary for the proper understanding of morality — we would come in “due course... to appreciate the ideal of just human cooperation”. (474) But here we must disambiguate ‘appreciate’. Presumably Rawls takes it to mean ‘to come to see the importance of and *accept as an ideal to govern your life*’. An amoralist — perhaps someone who had once been a full fledged moralist — could not, of course, appreciate it *in that way* and be an amoralist, but he could appreciate it in a manipulative way. That is to say, he could see that human cooperation was necessary both to ward off a Hobbesian state of nature and for people to realize their rational life plans. But he can — without any loss of rationality — be a free-rider or take a purely class point of view and gain the values of human cooperation, while not himself being committed to any principle of fairness. He appreciates the social value of human cooperation, but he has no appreciation of such an ideal as something to which he must commit himself. And in not so committing himself, it is not clear to me that he denies his nature as a free and rational agent anymore (or any less) than if he did so commit himself. (572) Yet Rawls denies this, though he also says things (as we shall see) that would lead one to think that he should accept it. (See his remarks about first-person and free-rider egoism, pp. 486-7.)

What appears at least to be the case is that while sometimes Rawls sees the force of something very much like what I have been arguing, he still characteristically in this arguments fails to acknowledge its force. This is particularly, and crucially so, when he thinks about what a rational person must commit himself to. Rawls, like many a moral philosopher, understandably enough, wants to show that there is a cluster of moral principles which are uniquely rational such that, for certain situations of a determinate kind, any rational agent or at least any fully informed, fully rational agent would have to adopt them as governing his behavior or be diminished in his rationality. If the above argument is correct, it is very doubtful indeed whether Rawls’ key claim is so. But Rawls has some further arguments and before we settle with that conviction about convictions we should look at them.

V

Rawls' second argument for the claim that rational individuals *must* desire to be just, turns on the claim that "the sense of justice is continuous with the love of mankind". (478) But even if this is so — and that it is so is at least plausible — how do we know that rational individuals *must* love humanity? It is rational in a well-ordered society, under normal circumstances, to love one's parents, friends and the like. And it is not irrational to love mankind as well. But why is it, or is it, that a man is in any way faulted in his *rationality* if he does not love mankind? As far as I can see there are no sound grounds for claiming he has been so faulted if he does not love humanity.

Lastly (in this passage) Rawls argues, following Kant, that in reasoning in accordance with the principles of justice men "express their nature as free and equal rational beings. Since doing this belongs to their good, the sense of justice aims at their well-being even more directly". (476) But (as I put it in another context) why *must* it be the case — or is it necessarily the case — that an amoralist, a thoroughly unprincipled man, is any the less a free and rational being for not being a morally good man, as distinct from being merely a man of good morals? It would, in anything like normal circumstances, be disastrous for an individual to be *seen* to be through and through unprincipled, but it isn't clear that he must have engaged in any irrationality if he keeps his unprincipled behaviour a well-guarded secret.

It may be the case that rational individuals in the initial situation would adopt Rawls' principles of justice. But this does not show that a rational individual in a well-ordered society or even in a not so well-ordered society must, where he is only considering his own rational plan of life (a plan for himself), 1) desire to be just or 2) always find it in his rational self-interest to be just. Has Rawls given us adequate grounds for believing that the "common nature of man" is such that if an individual does not act in accordance with the principles of justice, then he has denied his common human nature or has shown that he is not a free rational agent? (I shall return to this question.)

What Rawls has shown, I believe, is that we, or at least most of us, do, as a matter of fact, have such moral sentiments, including a sense of justice, and that this is to be expected on psychological developmental grounds, given our moral education (indoctrination). And, in addition, he has shown that rational agents in the initial position would recognize that these principles — that is at least *some* distinctively moral principles, not necessarily Rawls' own principles

— are the standards it is most reasonable for them to adopt to govern their relations with each other. He may even have given us reason to believe that such a capacity for and propensity towards principled behaviour is, as an outcome of natural selection, an adaptation of humans to their place in nature. (503) But all of this does not establish that it is the most rational thing for *an individual* to do or to want to do. What is in everyone's interest, taken collectively, and what is in most people's interests, taken distributively, need not be in an individuals' interests taken individually. Rawls stresses this himself, though in another context, when he writes :

To be sure, from the standpoint of the original position, principles of justice are collectively rational; everyone may expect to improve his situation if all comply with these principles, at least in comparison with what his prospects would be in advance of any agreement. General egoism represents this no-agreement point. *Nevertheless, from the perspective of any one man, both first-person and free-rider egoism would be still better.* Of course, given the conditions of the original position neither of these options is a serious candidate. Yet in everyday life an individual, if he is so inclined, can sometimes win even greater benefits for himself by taking advantage of the co-operative efforts of others. Sufficiently many persons may be doing their share so that when special circumstances allow him not to contribute (perhaps his omission will not be found out), he gets the best of both worlds: on these occasions anyway things transpire much as if free-rider egoism had been acknowledged. (496-7 italics mine)

It is evident enough that the first-person and the free-rider egoist is an amoralist or immoralist. He doesn't have an eccentric or self-serving morality, he just doesn't have any morality, at all. But what is not crystal clear is why he necessarily must be an irrationalist or indeed why he could not even be a thoroughly rational individual. Rawls recognizes that from the "perspective of any one man" such an egoism "would be still better" than acting from the moral point of view. But, since this is the case, why can't such an egoist be a rational agent? (Remember 'rational' for Rawls is to have no moral force.) Rawls denies that he can be rational. But isn't he just mistaken here? It is irrational in most circumstances to be *seen* to be unfair but need it be irrational *to be* unfair?

Someone might plausibly remark that we should not give a direct

— ‘Yes or No’ — answer to this. If Rawls’ account of the dynamics of moral learning is even near to the mark, rational agents in a well-ordered society — and even in a society with an endurable moral order — are people who have a sense of justice and this means that they will with some regularity act in accordance with it. We will thus, after all, have good *explanatory* reasons why they are moral. But what about *justificatory* reasons? Can we show the free-rider egoist or the classist amoralist who only considers the interests of his class or those he just happens to like that he must have made some *intellectual* mistake in opting for such a plan of life? Can we know that being fair must be an integral part of any *rational* plan of life?

We could show an individual that it was rational for him to want a well-ordered society and that in a well-ordered society his moral education would be such that he (most likely) will feel guilty if he really does act as such a free-rider egoist or classist amoralist. He will, whether it is rational or not, end up with a sense of justice and this sense of justice will make him miserable for being such a free-rider or classist amoralist.

However, such an egoist or classist could respond that once he has such a clear insight into the situation and, if he is rational through and through, he can discount the early effects of moral indoctrination and come to understand what it would be like to view things and indeed to prudently act from an amoral perspective. He wants to know whether it can be known or justifiably believed that such an amoral perspective must be irrational. He could, even more cautiously, put it counterfactually: if he could discount the effects of early indoctrination in the moral point of view so that he could become more autonomous, would there be anything irrational in taking such an amoralist’s perspective?

To this, it might be replied, that there would be such reason, if his very humanity and self-respect are to count for anything, and if they are to be important to him in his image of himself and in the living of his own life. Here we return to some Kantian themes previously mentioned (In this connection, section 74 — particularly pages 486-489 — and section 86 are crucial in *A Theory of Justice*.)

Rawls argues powerfully that “the moral feelings are a normal feature of human life. We could not do away with them without at the same time eliminating certain natural attitudes”. (487-8) He goes on to add “among persons who never acted in accordance with their duty of justice, except as reasons of self-interest and expediency dictated there would be no bonds of friendship and mutual trust. For when these attachments exist, other reasons are acknowledged for acting fairly”. (488) A natural attitude would be the attitude of love

and trust. They overlap with moral feelings, for moral feelings, include centrally such things as remorse, indignation, guilt, shame, approval, joy, trust, love and friendship. But these things could not — logically could not — be part of the psychology of a thorough amoralist. Even resentment and indignation are not—though they are natural attitudes — attitudes an amoralist or free-rider egoist could have. If I have as a rational plan that I will only treat people justly where it pays, I cannot, by definition, have friends or love someone. ‘Bonds of friendship’ does not, semantically speaking, even allow such a relationship. It is a grammatical remark to say ‘If X is my friend, then I cannot deliberately and regularly treat X unjustly when it simply suits my interests to do so’. And it is another grammatical remark to say ‘I cannot simply regularly ignore the interests of those I love when all that is involved in doing so is that I would gain from it’. ‘He loves her but he doesn’t at all care about what happens to her’ is an incoherency. Similarly, while the amoralist could feel anger and annoyance, he could not feel resentment and indignation. (488) As moral feelings, resentment and indignation must be elucidated in terms of principles of right and justice. To have them is to accept such principles as regulative for one’s behaviour. The central thrust of Rawls’ argument occurs in the following passage :

One may say, then, that a person who lacks a sense of justice, and who would never act as justice requires except as self-interest and expedience prompt, not only is without ties of friendship, affection, and mutual trust, but is incapable of experiencing resentment and indignation. He lacks certain natural attitudes and moral feelings of a particularly elementary kind. Put another way, one who lacks a sense of justice lacks certain fundamental attitudes and capacities included under the notion of humanity. Now the moral feelings are admittedly unpleasant, in some extended sense of unpleasant; but there is no way for us to avoid a liability to them without disfiguring ourselves. This liability is the price of love and trust, of friendship and affection, and of a devotion to institutions and traditions from which we have benefitted and which serve the general interests of mankind. (488-9)

This is the Kantian (or for that matter ‘natural law’) motif that a man’s common human nature, his humanity, commits him to taking the moral point of view. The natural attitudes of friendship, love, affection, and mutual trust are things he would have reason to want

if he has *reason* to want anything. If anything is a rational desire, wanting these things are rational desires and the having of such attitudes is the having of rational attitudes. *If* to be rational is to act on those desires that one would have when one is fully informed and vividly aware, then these are things it is always rational to desire. But one cannot desire these things without also being committed to the principles of justice: to trying to act in accordance with these principles not just as a man of good morals would, but as a morally good man would. In short, it is never rational to abandon the moral point of view.

Note that while it might be objected that the use of 'humanity' and 'disfiguring' is normative in the above passage and that they might be thought to be contestable notions, Rawls need not argue against that claim to make his core point, for it is just our unavoidable involvement with the natural attitudes cited above that makes it rationally mandatory on us as individuals to adopt the moral point of view.

However, Rawls cannot justifiably make so strong a Kantian claim on the basis of the evidence he has given us. What he has done is to establish that one cannot be a first-person or a free-rider egoist and have such *natural* attitudes. He has not shown that if one prizes friendship, love, trust and the like, one must have a sense of justice and be committed to principles of right and justice, for one might simply take 'a class point of view' or perhaps even 'a familial point of view' limiting one's concern for people quite deliberately to one's own family and relatives or small circle or class and yet experience friendship, love and trust. That is to say, one could have the attitudes appropriate to 'taking a familial point of view' or 'a tribal point of view' or 'a class point of view' and still have those natural attitudes. Such a man, committed, say, to 'a familial point of view', rejects principles of justice as firmly as the first-person egoist, but he still has these prized natural attitudes.⁶

There are certain passages in Section 86 of *A Theory of Justice* that might be construed as an implicit reply to the above argument. Once we have conceded, as I have, that there are principles of morality which are collectively rational and that it is in the interest of each that everyone comply with them, I cannot, without error, maintain the equal rationality of taking something like a purely class point of view. Recognizing that principles of morality are public and setting myself — my relations to my class and circle of friends and close associates aside — "on a systematic course of deception and hypocrisy, professing without belief" the accepted moral views as it suits my purposes and the interests of myself and my friends, I still

make, Rawls claims, a mistake, for the psychological costs, given my indoctrination, will be too great to make such deception worth it. (570) The necessity of taking precautions, maintaining a pose with its consequent loss of spontaneity and naturalness will make the price of so acting too high to make it a reasonable option for a rational man in a well-ordered society. (Rawls recognizes that what should be done in exploitative and corrupt societies such as ours is another matter.) (570)

What Rawls says *may* be so; it is surely not unreasonable to believe that it is so. But it is precisely the sort of claim that requires considerable, carefully sifted empirical evidence linked with a rather more sophisticated social theory than we have at present. Here we cannot simply rely, as Rawls does, on what it seems reasonable to believe. We need here actual sociological research. Sticking to our armchairs, we can easily develop different scenarios than Rawls', including one in which we have Mafia-like people in positions of security in their own clan or class ignoring moral considerations and still attaining the various psychological reinforcements of which Rawls speaks, while avoiding threats to their security and any extensive need to put on a pose or take elaborate precautions. No doubt such an outlook usually involves rationalizations about the worthlessness of 'the others' outside the clan, but it is not clear that it must. They might not, where they are powerful (say a ruling class), extend their caring beyond their own circle. They might have no love for mankind; indeed they might be quite indifferent to the fate of people beyond their circle and not suffer any failure in rationality, security or ease of life. Indeed the latter two might be enhanced. They would want to be fair to friends and give justice to those they happen to care for, but their familial, clan or class points of view would be their reference points and not the principles of justice — principles partially constitutive of the moral point of view. Yet these amorality could very well have the indispensable natural attitudes of which Rawls speaks (580).

Rawls could reply that I am forgetting that he is speaking of a well-ordered society and in such a society these natural attachments, which are so necessary for human flourishing, would be extended rather more widely than I allow. But it still remains the case that even in such a society a rational agent, rather like a free-loader in our societies, could recognize that his conditioning would very likely take such a moralistic form, and still could conclude — given the admitted strains and costs of moral commitment — only to keep such commitments where he would not be hurt by ignoring them or where he would not hurt his friends and the people closely associated with

him by ignoring them. It is not my concern to show that people actually so reason and so act — we do not have any well-ordered societies either — but simply to claim that there is no reason to believe that such an amoralist, if indeed he exists, would be less rational than Rawls' man of moral principle.

It is rather unrealistic on Rawls' part to counter that such an agent could not select who would be hurt by his unfairness and thus he might very well inadvertently harm those for whom he cares. A moments consideration of Rawls' own example of tax-evasion shows that. In a world of non-tax evaders an individuals' intelligent and reasonably prudent tax-evasion is going to do precious little general harm but it can be very advantageous to himself and to those close to him. Rawls is being Quixotic when he claims that in all recognizable human contexts there are "strong grounds for preserving one's sense of justice". (371)

The social nature of human nature and the important role of what Rawls calls the Aristotelian principle in our life is also not sufficient to make amoralism irrational. We indeed need people to bring to fruition our latent powers, the proper functioning of which is essential for our well being, but there is no reason why, at least for some rational agents powerfully placed, a familial or at least a purely class point of view could not so serve as readily as a moral point of view.

Perhaps, borrowing far more heavily from perfectionist moral theory than Rawls is prepared to do and stressing his seventh principle of rational choice, someone wishing to reconstruct slightly his account to meet these objections, could develop and clearly articulate a more satisfactory self-realizationist account of morality in which he would have established that to realize ourselves — to achieve our full human flourishing — we must be just men and not just men of good morals. That is, of course, a humanly attractive doctrine, yet it is fraught with obscurities and has not been worked out by Rawls or, as far as I know, by anyone else.⁷

Rawls is adamant that "in order to realize our nature we have no alternative but to plan to preserve our sense of justice as governing our other aims." (574) I would no longer make the curt rejection of the coherency of talk about 'realizing our nature' I once did, but it still is a very obscure formula requiring a careful elucidation and defense for it to be something to which we can legitimately appeal.⁸ Marx and some Marxists make something of a beginning here; but Rawls does nothing with this and it is at best a first step.⁹

Similar considerations obtain for the attractive but unestablished doctrine that "the desire to express our nature as free and equal

rational beings can be fulfilled only by acting on the principles of right and justice as having priority". (574) Perhaps if we stress 'equal' something can be done towards its rather truistic establishment; but where 'equal' is dropped, or not given prominence, it remains a morally attractive but quite unsupported claim. We have been given no good grounds for believing that if we would be rational we must be committed to it.

Rawls in an ancient and honorable tradition in moral philosophy wants to get out of reason more than reason can establish. This is not to say that reason is wanton but it is to give to understand that it cannot provide the decision procedure in morals that Rawls envisages. (574-5) The sentiment of justice does indeed reveal what a person is and expresses in a very fundamental way a conception of oneself. But this does not mean that the achievement and sustenance of an individuals' rationality, even in a well-ordered society, is tied to that sentiment. Thoroughly rational people might be unprincipled bastards.

University of Calgary

FOOTNOTES

¹ John Rawls, *A Theory of Justice*, (Cambridge, Massachusetts : Harvard University Press, 1971). All references to Rawls are from *A Theory of Justice* and they are given in the text.

² Ralf Dahrendorf, *Essays in the Theory of Society*, (Stanford California : Stanford University Press, 1968), pp. 151-178.

³ For two striking and indeed important statements of a contrary view see Tony Skillen, "Marxism and Morality", *Radical Philosophy* 8 (Summer 1974) and Andrew Collier "On The Production of Moral Ideology", *Radical Philosophy* 9 (Winter, 1974) and "Truth and Practice", *Radical Philosophy* 5 (September, 1973). While such a morality of principles with its deontological elements, can and indeed often does have the mystifying and harmful ideological effects which Skillen and Collier characterize, it by no means follows that such a morality must be such an ideology or that to attain this last stage of moral development just is to become firmly captive of such a moral ideology. Moreover, without some sense of justice, it is impossible to attain the solidarity, fraternity and, indeed moral

freedom, essential for a fully human society, answering to Marx's conception of such a society. Some of this is brought out in Peter Binn's reply to Skillen and Collier in his "Anti-Moralism", *Radical Philosophy 10*, (September, 1975). See as well Philip Corrigan and Derek Sayer, "Moral Relations, Political Economy and Class Struggle", *Radical Philosophy 12* (Winter, 1975).

⁴Not entirely, however, for see what he says about rationalism on page 495.

⁵The need for drawing this distinction and some of the crucial ways we can fall into confusion here has been importantly re-stressed by E. J. Bond, "Reasons, Wants and Values", *Canadian Journal of Philosophy*, Vol. III, No. 3 (March, 1974), pp. 333-47.

⁶I owe this last point to Professor Grace Dyck. I should also add that there is a curious passage in Chapter IX (page 568) of *A Theory of Justice* where Rawls tries to go around the considerations I have been taxing him with in this essay. He points out there that he is "not trying to show that in a well-ordered society an egoist would act from a sense of justice, nor even that he would act justly because so acting would best advance his ends". He is not even arguing "that an egoist, finding himself in a just society, would be well advised, given his aims, to transform himself into a just man". But then what happens to the tight link between rationality and morality that Rawls is concerned to forge, namely to show that in the world, as we know it, a rational human being *must* — not merely *may* — strive to be a moral human being, e.g. not only do the things a just man would do but do them for the just man's reasons? Rawls in this passage seems to be saying, as I think he should, that there is no such tight link, but then he has lost a grip on what for him is a very fundamental claim, namely that moral principles are the principles that rational human beings would choose to govern their actions. This claim, of course, needs disambiguation but, in one important way — if my central arguments have been correct — it is false and Rawls in this passage in effect agrees that it is false. To counter by remarking that an individual egoist's policies and strategies are not principles, comes down too hard on what may be a correct perception about the ordinary use of 'principles'. However, the essential point is that the individual egoist could have quite teachable strategies that appear at least to be no less rational than moral principles.

⁷Richard Norman makes a start in his "Self-Realization", *Philosophy*, forthcoming. His essay is in part a response to my attempt to exhibit the extensive obscurities and incoherencies in a

self-realizationist ethics in my "Alienation and Self-Realization", *Philosophy*, Vol. 48 (January, 1973).

⁸ For these earlier curt rejections see my "On Taking Human Nature as the Basis of Morality", *Social Research*, Vol. 29 (Summer, 1962) and "Conventionalism in Morals and the Appeal to Human Nature", *Philosophy and Phenomenological Research*, (December, 1962).

⁹ See here Michael P. Lerner, "Marxism and Moral Reasoning", *Social Praxis*, Vol. 2/1-2 (1974), pp. 63-88, and my "Class Conflict, Marxism and the Good Reasons Approach", *Social Praxis*, Vol. 2/1-2 (1974), pp. 89-112.