



An "Equal Power" Theory of Right

PETER GEORGE WOOLCOCK

DEPARTMENT OF PHILOSOPHY

UNIVERSITY OF ADELAIDE

APRIL, 1984

Awarded 2-2-85

FOR IRIS

TABLE OF CONTENTS

PAGES

6 - 10	Chapter 1	: The Veil of Ignorance
11 - 41	Chapter 2	: Unforced Agreement
42 - 57	Chapter 3	: Rawls' 'Circumstances of Justice'
58 - 97	Chapter 4	: Rawls' 'Formal Constraints of the Concept of Right'
98 - 114	Chapter 5	: Impartialism and Non-Persons
115 - 145	Chapter 6	: Assumptions about Personal Identity
146 - 160	Chapter 7	: Epistemological Assumptions
161 - 174	Chapter 8	: Autonomy, Rationality and Perfectionism
175 - 193	Chapter 9	: The First Principle of Justice
194 - 221	Chapter 10	: The Priority of Liberty
222 - 235	Chapter 11	: The Difference Principle
236 - 245	Chapter 12	: The Primary Goods
246 - 266	Chapter 13	: Impartialism, Entitlement and Desert
267 - 292	Chapter 14	: 'Oughts', 'Wants' and Egoism
293 - 311	Chapter 15	: Reflective Equilibrium
312 - 335	Chapter 16	: Why Be Impartialist?
336 - 344		: Selected Bibliography

ABSTRACT

In this thesis I propose a theory of right in which the rightness of an act is determined by its conformity to the principles that people would reach by unforced agreement. I call this theory 'impartialism'. I begin with an investigation of the conditions under which an agreement can be said to be unforced. I argue that this occurs when the parties to the agreement are in a position of equal power, behind a veil of ignorance and want not to be coerced more than they want to coerce. When these conditions are met, determinate answers can be given as to which acts should and should not be done in a particular set of circumstances.

As impartialism and Rawls' contractarianism both derive normative principles from the veil of ignorance, I assume that Rawls' theory will provide a useful guide to the normative principles of impartialism. Rawls' use of the veil has been much criticised so I look at these criticisms to see if they also apply to the use of the veil in impartialism. I show that most of these criticisms lose their force if the scope of the veil is not constrained in the ways Rawls requires. I show that these constraints are not inherent to the nature of the veil but are external restrictions placed on the veil to produce the particular normative conclusions Rawls is after. In particular, I show that the scope of the veil does not need to be limited to cases that fall within the circumstances of justice, or to principles that meet the formal constraints of the concept of right.

I then look at various claims to the effect that the normative conclusions derived from the veil are vitiated due to a bias built into the veil by certain metaphysical or ethical assumptions. In particular, I show that metaphysical or ethical assumptions are not involved in deriving from the veil normative conclusions such as the rightness of religious tolerance, the moral priority of persons over non-persons, and the inviolability of autonomy. I also show that the veil does not presuppose any particular

epistemology, or any particular metaphysical theory of personal identity.

I then look at the substantive normative principles that Rawls deduces from his theory. I show that the priority of liberty can be derived from impartialism but the second principle of justice cannot, although principles can be that do limit inequalities of wealth. I show that the veil does not have to be coupled with the Difference Principle, the primary goods or the Aristotelian Principle to generate intuitively plausible principles. I also look at how contractarianism can handle desert and entitlement.

In the remainder of the thesis I look at three issues. Firstly, I defend how I have used such words as 'ought' and 'want'. This defence commits me to the view that egoism cannot be dismissed as irrational. Finally, I look at the extent to which there is a rational basis for preferring impartialism to other theories of how we should act, including egoism. I argue that neither reflective equilibrium nor Hare's methodology provide such a basis. I conclude by demonstrating that there will be a tendency to publicly justify one's actions in impartialist terms in real life. I also show that argument can resolve conflicts of desires only if the parties to the argument are impartialists.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University and, to the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the text of the thesis. I consent to the thesis being made available for photocopy and loan if applicable if accepted for the award of the degree.

PETER GEORGE WOOLCOCK

ACKNOWLEDGEMENTS

I would like to thank the following people :

My supervisor John Chandler for his help and encouragement; the staff and post-graduate students of the Philosophy Department, University of Adelaide, for their comments on preliminary drafts of most of the chapters of the thesis; Mavis Partridge for typing the thesis; and my wife Iris and my children Dylan and Buffy for their support.



Chapter 1

The Veil of Ignorance

Philosophers have proposed a variety of methods by which to determine which acts we ought to do, and which we ought not. If the acts we ought to do are 'right' acts and the ones we ought not to do are 'wrong' ones then these methods are methods for determining which acts are right or wrong. Perhaps the most famous philosophical method for determining which acts are right or wrong is utilitarianism. Its method is to determine the right act in terms of which maximises aggregate utility (Classical Utilitarianism) or average utility (Average Utilitarianism). Utilitarianism as a method determines right or wrong in terms of a particular content such as happiness or want-satisfaction. Not all methods for determining right and wrong do so in terms of a particular content. Kant, for example, proposed a purely formal method, namely, 'Act as if the maxim of your action were to become through your will a universal law of nature'. (1948, 84) More recently, Hare has proposed that the rightness of an act is determined by whether the person who is considering it can universally prescribe it (1963). Hare's method, like Kant's, is a purely formal one, although he now thinks it is sufficiently strong to entail utilitarianism. (1981)

In this thesis I will be concerned by the kind of method for determining which acts are right or wrong that John Rawls puts forward in his 'A Theory of Justice'. (1972) His method determines which acts are right or wrong by reference to the principles that people would agree to behind a veil of ignorance. He describes the veil of ignorance as a situation in which the parties are deprived of certain information, namely, 'they do not know their place in society, their class position, or social status,

nor do they know their fortune in the distribution of natural talents and abilities. It is assumed that they do not know their conception of the good, that is, their particular final ends; nor, finally, their own distinctive psychological dispositions and propensities, and the like'. (1980, 522-523) Rawls says that the people behind such a veil are in an 'original position of equality'. (1972, 12)

Rawls wishes to use the veil of ignorance as a method for arriving at principles of justice, rather than principles of right. His primary interest is in the principles that the parties behind the veil would agree to concerning the basic structure of society, that is, 'the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social co-operation'. (1972, 7)

Nonetheless, as Farrell (1980, 188) notes, Rawls does have a theory of right, that is 'his general theory of normative ethics, in other words. This theory, which Rawls calls "rightness as fairness" and which has received surprisingly little critical attention, would have us think of the principles of normative ethics as the outcome of hypothetical deliberations that take place behind a carefully constructed veil of ignorance'.

Rawls elaborates his notion of 'rightness as fairness' as follows, 'the concept of something's being right is the same as, or better, may be replaced by, the concept of its being in accordance with the principles that in the original position would be acknowledged to apply to things of its kind'. (1972, 111)

It is my intention in this thesis to investigate the consequences of adopting the veil of ignorance as a method for determining which acts are

right or wrong. This will involve discovering what guidance, if any, the veil gives us about which particular normative principles we should follow. I will also be concerned with seeing if there are any convincing reasons for preferring the veil of ignorance to other methods of determining which acts are right or wrong. This will involve seeing whether there are any convincing reasons for preferring to adopt the veil's advice as to which acts are right or wrong when its conclusions differ from those of other theories. It will be my contention that the veil is to be preferred to any other method.

Rawls' version of the veil of ignorance, however, has been widely criticised as a method for determining which basic structure for society is 'just'. These criticisms may also apply to it as a method for determining which acts are right or wrong. In this thesis I will show that these criticisms are not so much criticisms of the veil of ignorance itself but, rather, are criticisms of other aspects of Rawls' theory which lead him to place restrictions on the range of application of the veil. These restrictions, however, are not inherent to the nature of the veil itself. I will argue, in fact, that the veil is a formal method for determining which acts are right or wrong and that many of the objections to it are a consequence of Rawls' attempt to import content into his particular application of the veil.

Each of the various methods listed earlier for deciding which acts are right or wrong are methods for eliminating disagreement between people about which act should be done. If everyone accepted classical utilitarianism as the arbitrator of such disputes, for example, then the only basis for disagreement over what should be done is disagreement over the nature of the facts. Rawls sees the veil as a method by which

disagreement can be eliminated. It is necessary, in his view, for the veil to exclude the kind of information it does 'if no-one is to be advantaged or disadvantaged by natural contingencies or social chance in the adoption of principles. Otherwise the parties would have disparate bargaining advantages that would affect the agreement reached. The original position would represent the parties not solely as free and equal moral persons, but instead as persons affected by social fortune and natural accident. Thus, these and other limitations on information are necessary to establish fairness between the parties as free and equal moral persons and, therefore, to guarantee that it is as such persons that they agree to society's basic principles of justice'. (1972, 523)

If we are to come up with a single determinate answer to what should be done in a particular situation, then we do need a method by which to eliminate all hypotheses except one. If people disagree as to which act is the right one, then this method will eliminate all disagreement. Why, however, should it be Rawls' method rather than one of the utilitarian ones, or Kant's, or Hare's, or any of the many other candidates?

Rawls himself suggests, in the passage just quoted, that his method guarantees that the agreements the parties reach are ones they reach as 'free and equal moral persons'. He values, then, a method that has two features of note. Firstly, that it decides what is right or wrong as a result of agreement rather than fiat. Secondly, that this agreement is one reached by free and equal moral persons. Not everyone, however, values these features in a method for giving a determinate answer to what should be done in a particular situation. If we share Rawls' values then he might well be able to convince us that his method is either the only one that meets these values, or meets these values better than any other method. If we don't share his values, however, then he needs to make a

case as to why a method that embodies these values is to be preferred to one that doesn't.

In this thesis I will argue that Rawls is correct to value a method that decides what is right or wrong in terms of the agreement of free and equal rational agents. Like Rawls, I will try to clarify what conditions an agreement needs to meet in order to be so described. I will argue that it needs to be an unforced agreement, and that this is only possible if agreement is reached in a position of equal power. Other conditions will also need to be met, including that the parties argue from behind the veil of ignorance. The substantive moral conclusions to be drawn from the position of equal power will be similar, but different in important respects, to those Rawls obtains from his theory. In the next chapter, I will examine in detail the conditions that need to be met if an argument is to be described as occurring in a position of equal power and the conditions it needs to meet if it is to result in a determinate conclusion.

Chapter 2

Unforced Agreement

1. Introduction

In this chapter I intend to identify the conditions that must apply to an agreement for it to be an unforced agreement. I will be concerned with agreement between rational agents. I will be using 'rational' here in Rawls' sense, namely, that 'a rational person is thought to have a coherent set of preferences between the options open to him. He ranks these options according to how well they further his purposes; he follows the plan which will satisfy more of his desires rather than less, and which has the greater chance of being successfully executed'. (1972, 143)

Rawls goes onto make the special assumption that the rational person does not suffer from envy. I will not build this requirement into my use of the word 'rational'.

Two rational agents A and B will be said to be in disagreement when there is a conflict of desires between them. The desires of A and B will be said to conflict when the situation is such that, if A's desire is satisfied then it is logically impossible for B's desire to be satisfied. This means that B's desire can only be satisfied if the situation is changed, for example, if the amount of some desired material were increased. The most straightforward case of a conflict of desires is when A wants to do some action X whereas B wants A not to do X. I will call a case of this kind an 'actual' conflict of desires. A potential conflict of desire exists when the relationship between A and B is such that, were B to know what A wanted to do in this particular case, then B would want A not to do it. If A's desire is in conflict with B's, then the two

desires will be said to be 'incompatible'.

One way in which A and B can try to resolve their conflict of desires is to find a principle P that both of them are prepared to follow that gives a determinate answer as to which desire is to have priority in any particular kind of situation. If A's desire is to do X and B's desire is that A not do X, then the principle P will determine whether or not A should do X in this particular case. A and B, therefore, could seek to resolve their conflict of desires by trying to find a principle they both could agree to that would give them a determinate answer as to what should be done in this case, as well as a guide as to what should be done in future cases of a similar nature. It may be the case, however, that in proposing P or any other principle, one of the parties is not sincere. The sincerity of the parties in agreeing to a principle has considerable bearing on the conditions that must apply if the agreement is to be unforced.

2. Sincere and Insincere Agreement

If A and B agree that people should act in accord with the principle P, then their agreement is 'sincere' when each of them really believes that people should act in accord with P. Neither of them is being hypocritical, in that he is not publicly endorsing one principle while privately endorsing another. If there is agreement between A and B when one of them does not believe in the principle publicly agreed upon, that is, does not really believe that people should act in accord with it, then I shall call that person's agreement an 'insincere' one. If I am arguing with someone over whether or not people should act in accord with some principle P, I assume that they are sincere in their attempt to reach agreement, that is, that they are looking for a principle that they can sincerely publicly

agree with me on. If I know that their agreement is insincere, then I know that there is some principle they haven't disclosed which is the real guide to their actions and which will override the principle we have supposedly agreed upon whenever their desires are incompatible with mine. If I know that their agreement is insincere in this way, then it is rational of me to have reservations about the reliability of their acting according to our publicly agreed principles. It may be rational of me to require that our agreement be reinforced with sanctions in a way that may not be necessary if their agreement were sincere. Using sanctions to enforce an agreement, of course, does not make their agreement sincere. Rather it just prevents or inhibits, their acting upon the principles they really do hold sincerely.

3. Forced Agreement

One kind of insincere agreement is 'forced' agreement. Forced agreement occurs when there is unequal power between A and B. There is unequal power between A and B when B agrees with A because of some difference between A and B which, if removed, would result in B's not agreeing with A. Such a difference will be called an 'advantage'. Greater physical strength is a typical advantage, as is greater intelligence, or greater sophisticated ability, or greater bloody mindedness, or patience. If A is physically stronger than B, then A may be able to use this advantage to get B to agree here when B would not have agreed with him were they equal in physical strength. Having greater physical strength enables A to exercise sanctions against B. If B doesn't act as if he agrees with A, or doesn't say he agrees with A, then A can do things to B that B doesn't want done whereas B cannot do these same things, or even equivalently unpleasant things, to A. All advantages can be translated into physical strength terms. If A is able to get agreement due to some advantage, then were A and B equal with respect to this advantaging characteristic and with

respect to all others except physical strength, then A would only be able to get the same agreement from B by resorting to the use, or threatening the use, of sanctions. By the same token, any advantage could be translated into any of the others but it will be convenient to take physical strength as the standard as it is the most readily recognized and familiar.

If A has superior power to B, then B can either insincerely agree to A's principle P or sincerely disagree with A and suffer the consequences. Forced agreement, then, can be one kind of agreement over principles. As I indicated earlier, however, if B's agreement is insincere, then A maintains B's agreement solely by sanctions. As a rational agent it may be worth B's while to work towards a redress of the balance of power so that A no longer has the superior power. A has not offered B a principle that B sincerely endorses as to why B shouldn't do this if he can manage it. The only reason for B's agreeing with A (or his acting as if he agrees) is A's power to exercise sanctions. Once these disappear, B's reason for agreeing with A evaporates. Their agreement, then, is inherently unstable.

4. False Sincere Agreement

B may agree with A because A has kept relevant information secret. If there were no differences between A and B with respect to knowledge then B would not have sincerely agreed with A. B only sincerely agrees with A because A has been able to avoid the use of sanctions against B by deceiving B in some way. If A and B were in a position of equal power, that is, there was no difference between A and B that gave either an advantage over the other, then B would not have agreed with A. It is only because A has an advantage over B that, were A and B equal with respect to all advantaging characteristics including this one but excluding physical strength, then A would have to use sanctions against B to get the same agreement. B's agreement,

then, can still be forced even though it's sincere. It is forced whenever it would not be maintained were A and B in a position of equal power. It is a 'false' sincere agreement. B seems to consent to P but his consent is a 'pseudo-consent' not a 'genuine' one. False sincere agreement, then, is another kind of agreement that A and B can reach over principles but it, too, is really a form of forced agreement. The mere fact that agreement is sincere does not show that it is not forced. Whether it is forced or not is shown by whether it would be maintained in a position of equal power.

5. Conflicts in Positions of Unequal Power

Consider a case where A and B have incompatible desires but A has superior power to B. Suppose A sincerely proposes some principle P to resolve the conflict of desires while B sincerely proposes some principle Q. Suppose, further, that both A and B know that, in a position of equal power A would not agree to Q and B would not agree to P. Should A use his superior power over B to get B to agree with him, that is, should A force B to act in accord with P and to publicly say that everyone should act in accord with P, even if B's agreement is insincere?

If A does, then he does so even though he cannot give B any reason that B would find acceptable in a position of equal power, that is, in a position where A and B are equally capable of determining which of P and Q is the better principle. A has not convinced B that P is the better principle. He has not persuaded B, who is as capable of assessing the merits of P and Q as A is, that P deserves priority over Q. If A, then, having the superior power, uses that superior power to enforce P on B, what reason can he have? Whatever his reason, it is one that is not unacceptable to B in a position of equal power. It is not that P has some feature that any

rational agent would acknowledge gives it superiority to Q. This is shown by the fact that B has already refused to acknowledge its superiority when he had equal power with A. If A chooses P, then B can only interpret this choice as being based on who holds the principle, not on any quality internal to the principle itself. If A had sincerely proposed to B in a position of equal power that P be chosen just because it was A's principle, then B would have rejected this unless Q was the principle that everyone would agree with A. In this case, however, there would be no conflict of desire between A and B. As long as there is a disagreement between A and B over ultimate principles, B could not, as a rational agent, accept it as a reason for choosing P over Q that P is the principle A prefers. This would be for B to accept that a principle had priority regardless of its nature merely because A rather than B wanted it. As a rational agent with desires different from A's, B could not agree to this.

In forcing B to agree to P just because he has superior power even though B refused to agree to P in a position of equal power A shows himself to be the kind of person who believes that whatever principles he holds are to have priority over any principles that B holds, merely because it is he who holds them rather than B. If this is so, then what really matters to A is not that everyone follow P but that everyone follow whatever principle he, A, believes in. It is really a case of A placing himself above any principle. He derives his principle P from the meta-principle that any principle he holds is to be preferred to any different principle held by anyone else. It is also a case of A ranking himself as more important than, or superior to, anyone else when it comes to deciding what actions anyone should do in a particular situation. He is claiming that his views have priority over everyone else's regardless of whether or not other people see that there is good reason why this should be so. He is dismissing other people as equal, or even significant, contributors to the

decision. No wonder, then, that, in a position of equal power, he fails to convince them. He has no arguments.

If A, then, is prepared to make B agree to P in a position of unequal power even though B refused to do so in a position of equal power, then he shows himself to be the kind of person who advocates a principle, not because of any merit the principle would have in the eyes of any rational agent, but merely because it is the principle that he happens to prefer for reasons that are specific to him as an individual.

6. Equal Worth of Persons

The position of equal power is a position in which A has to treat B as a person worthy of the same respect as himself. He has to give B's desires the same consideration as his own. He must treat B as an end in himself, not just as a means to his own ends. He cannot ride rough shod over B's interests in pursuit of his own. When he proposes a principle he must take into account whether it serves B's purposes as much as his own. He must give B's concerns equal weighting with his own if he wants to get B's co-operation. A may not believe that B is as worthwhile a person as himself but, if he cannot get B's agreement by force, deception or any other advantage, he must act as if he believes it.

If A is prepared to make B agree in a position of unequal power to what B refused to accept in a position of equal power, then A does not regard B as a person whose worth is equal to his own. In fact, he regards B's priorities as totally subservient to his own. He may not be vindictive towards B, that is, he may not be frustrating B's desires just out malicious pleasure, but, whenever B's desires conflict with his own, he has no compunction in sacrificing B's highest priorities for A's most trivial ones. At least, this is so if B is incapable of sanctioning A sufficiently. When it

comes to arguing, then, as to whose preferences are to prevail, A has already decided the issue, and his decision is only alterable by sanctions or their equivalent. I will call such a person an 'egoist'. His relation to more standard definitions of 'egoism' will become clear later.

7. The Veil of Ignorance

If A and B, then, are to reach agreement in a position of equal power, they will have to be prevented from arguing as egoists even if that is what they are. They will have to be prevented from giving priority to a principle just because it's the principle they happen to prefer. It is this insistence that a principle be given priority because it is A's that prevents A from agreeing with B in a position of equal power. Likewise, B's insistence that his principle override A's regardless of any intrinsic merit of the principle but merely because he is the person who holds this principle also prevents B agreeing with A in a position of equal power. It is A's or B's attitude that he is to be preferred to the other, even though there is no relevant difference between them, that renders a consensus impossible. This is not to say that there are no differences that could exist between A and B in a position of unequal power that justify A's being preferred to B but that the decision that this difference does justify such a preference is to be made in a situation where the decision is not already biased in favour of the difference in question.

In order, then, for A and B to agree about which of P or Q is the better principle, they will have to ignore whose principle P or Q is. They will have to argue as if each of them doesn't know which is the principle he personally favours. A and B will have to argue about the relative merits of P and Q as if the question has not already been settled by who holds the principle. They will have to argue as if they are behind Rawls 'veil

of ignorance', or something very similar to it. They will have to argue as if they are impartial between A and B, rather than partial to their own case. It is as if they stood outside of their particular identities and regarded A and B as two characters in a scenario. It is as if they had become two spectators, J and K, who were looking at a drama in which a character, A, had a particular desire whereas B had an incompatible desire and J and K were given the task of coming up with a rule to govern this kind of conflict without knowing if they were ever going to occupy A's role or B's, or that there was any greater chance of their occupying one rather than the other.

The argument, then, as to which of the principles P or Q is the one that A and B should follow, is to be conducted by A and B in their guise as J and K. Each of them has to argue as if he doesn't have any idea which of A or B he is more likely to be. If J did try to calculate the odds of his being A or B so that he could manipulate K into agreeing to a principle that favoured whichever of A or B that he, J, really was, then he would be reasoning as an egoist. His reasoning would begin with the meta-principle that priority should be given to whichever person in a hypothetical case he turns out to be. He shows himself disposed to treat his case as more deserving in advance of any agreement in a position of equal power that cases like his are more deserving than cases unlike his.

If J and K have to argue as if they are behind a veil of ignorance this will eliminate all areas of disagreement between them that rest solely on their disposition to give priority to their own case. It will force them to treat their own case as of no greater worth than anyone else's because they have to argue as if they do not know which one of the parties in a hypothetical situation they are.

Our discussion so far has isolated two methods by which agreement can be reached when A and B have incompatible desires. One method is that of force. This method is only possible in a position of unequal power and will occur if the party with the greater power is the kind of person who thinks that priority should be given to a principle just because it is his, that is, if he is an egoist. If A and B are not egoists, then they will continue to disagree in a position of unequal power because neither will force the other to agree. The other method is the method that makes agreement in a position of equal power possible by limiting the kinds of arguments A and B can use to those that do not rely on their giving priority to a principle just because it is theirs. It makes agreement possible by eliminating a major cause of disagreement, namely, the freedom to treat one's own case as superior to that of other people for reasons unacceptable to those other people. It forces A and B to find reasons for treating their own case as superior that are acceptable to other people in a position of equal power. This is done by restricting the arguments A and B can use in a position of equal power to only those they would use if they were ~~also~~ behind a Rawlsian-type veil of ignorance. They have to argue as if their meta-principle is to act in accord with whatever principles are agreed to by people in a position of equal power behind a veil of ignorance.

8. Ordinary Life

Even though A and B may be able to reach agreement on incompatible principles to resolve conflicts of desire if they are in a position of equal power of what relevance is this to those of us who are almost never in a position of equal power? People in ordinary life are rarely, if ever, in a position of equal power. Nevertheless, even in ordinary life, A and B may come to argue over which of the two ultimate principles P and Q people should follow. Even in ordinary life they can argue as if they are in a

position of equal power and as if they are behind a veil of ignorance. The point of the discussion so far has been to show that, if they agree in ordinary life but haven't argued as if they were in a position of equal power and as if they were behind a veil of ignorance then their agreement is a forced one. It is an agreement that has occurred because one or the other of J and K has either consciously or unconsciously manipulated the other into agreement, either by deception or by the direct use of sanctions. In ordinary life, if B has agreed to P even though he would not have done so in a position of equal power, then A only maintains that agreement by deception or by sanctions. B has no reason to agree with A other than the sanctions J can muster against him.

When we argue with each other about which principle people are to follow we assume that the argument is not one to be settled by force. The analysis so far is intended to show that we are mistaken about this if the parties to the argument have made no logical or factual errors yet refuse to argue as if they were in a position of equal power and as if they were behind a veil of ignorance. If the person we are arguing with does not sincerely argue like this, then he already has the attitude that whatever he believes is to have priority over whatever anyone else believes. The only way we can stop him putting this into effect is if we have the power to control him. That power may consist of bluffing him into agreeing with us by arguments that falsely lead him to think that what we believe is what he believes but this, as I have said before, is just a substitute for force. If he is not to be deceived in this way, then any argument other than showing him the consequences in terms of sanctions of his disagreeing with us is pointless. This kind of openly coercive argument will only be effective if he believes we have the power to put these sanctions into effect and that we also have the will. If he believes these, and he wants to disagree with us less than he wants to avoid the sanctions, then, as a

rational agent, he will act as suggested by the consequent of the hypothetical 'If you don't want to suffer the sanctions we can employ, then agree with us'.

That a person argues as if behind the veil in a position of equal power means that he has not already closed the argument in his mind in favour of whatever principle he holds, that he is prepared to rank principles in terms of criteria other than which particular individual holds the principle. In fact, he is prepared to rank principles in terms of criteria any person could accept, provided that person was in a position of equal power with others and behind a veil of ignorance. Only if the argument is conducted in this way can we eliminate the insolubility forced on the argument by the insistence of some people that they be given priority over everyone else. In ordinary life, a person who will not argue this way, and who is making no factual or logical errors, is thereby showing himself to be someone who insists that he be given priority over everyone else. When we enter into arguments with people about whether P or Q should be followed we do not expect the argument to be vitiated right from the start because someone insists that he be given priority and therefore that the argument be settled in favour of his principle. If he had openly declared this at the beginning we would have abandoned any type of argument except ones that threatened sanctions or amounted to it. Instead we assume, when we enter the argument, that those we are arguing with regard their views as having no automatic priority over ours, that the point of the exercise is to get us to consent freely and willingly to the same conclusion as them, that they will value only that conclusion that is arrived at by the free and willing consent of both parties and that they respect the fact that we are capable of free and willing consent, that is, that it matters to them that the conclusion they themselves act on is one that would be acknowledged freely and willingly by any rational agent as the right one. If we

know from the beginning that our opponents did not have this view, that they had no respect for the desires or beliefs of agents other than themselves, that they did not feel that it was a measure of the rightness of what they wanted to do that other rational agents would freely and willingly agree that it was right, then we would regard the whole activity of arguing with them in anything other than coercive terms as pointless.

9. Velorians and Conflictants

We see, then, that the fact that there is rarely, if ever, a position of equal power between the parties in ordinary life does not render irrelevant the principles that would be agreed in such a position. Instead, it is these very principles that are the ones that we would follow if we treated those who argue with us as agents whose worth is equal to our own.

As these principles are ones we would agree to if we were arguing as if we were J and K, rather than as we really are, that is, A and B, I will take all arguments about the principles that should be followed to resolve conflicts of desire between us as arguments between us in our roles as J and K. It will be useful to have a set of terms that distinguishes between the persona we adopt as J and K and the persona we have as actual participants in the conflict of desires. I will call J and K 'velorians' from 'velos' meaning 'veil' because they are the persona we adopt when we argue as if behind the veil of ignorance. I shall call A and B the 'conflictants' because they are the persona we have as parties to an actual or possible conflict. Hare calls the parties behind the veil by the abbreviation POPs ('people on the original position') and the parties to the conflict of desires by the name POLs ('people in ordinary life'). (1975, 89) His POPs are the same as my velorians and his POLs the same as my conflictants.

10. Coercion

We have, then, eliminated two kinds of principle that J and K could use to resolve which of the desires of A and B is to be acted on and which is not. J and K cannot use principles that people would only adopt if they were in a position of unequal power, and J and K cannot use principles that people would only adopt if they were not behind a veil of ignorance.

Eliminating these principles means that J's and K's chances of agreeing on a principle to cover any particular situation have increased, as the eliminated principles are no longer contenders. Unfortunately, however, the position of equal power and the veil of ignorance are not powerful enough to produce a single determinate principle for those cases where B's desire is that A do X but A does not want to do X. These cases raise the issue of whether J and K would agree that, in such a situation, B should coerce A to do X.

I will use the term 'coercion' in what Bayles (1972) calls its 'dispositional' sense.⁽¹⁾ Bayles says that A (dispositionally) coerces B to do X if, and only if, (a) Person A intends that B do X (b) A further intends to harm B if B doesn't do X (c) A threatens B with harm if B doesn't do X (d) B does X (e) B could have acted otherwise had he so chosen (f) B would have chosen otherwise had he not been threatened. I will follow this account of the use of the term 'coercion' with the minor modification that I will count A as trying to coerce B even when A is

(1) There are alternative analyses of the term 'coercion'. Pennock (1972, 3) suggests that Bayles' Point (d) doesn't always fit common usage. Held (1972) and McIntosh (1972) contend that Bayles' Point (c) is too narrow. They believe offers or other promises of benefit can be coercive. Nozick (1969), Girt (1972) and Wertheimer (1972) side with Bayles.

bluffing, that is, I will still regard A as trying to coerce B to do X even when A doesn't really intend to harm B if B doesn't do X but nonetheless intends B to think that A will harm him. This is a modification, then, to Bayles' Point (b).

My concern with the meaning of 'coercion', however, is not to give a correct linguistic analysis or to show how it is related to words like 'freedom' and 'power', although these are important philosophical tasks. My only concern is to adopt a meaning that is clear and close to common usage, which is why I think it desirable to modify Bayles' Point (b). Except for this variation, though, I think Bayles' analysis captures how most people use the word. Whether it does or not, however, is not the issue. Even if Held (1972) is correct that offers are coercive I am interested in agreements that result from the use of sanctions so I would have to talk of 'threatening' coercion so that ordinary language users wouldn't think I might mean 'offering' coercion. As I will only be meaning the sanctioning variety I will use Bayles' analysis.

Consider the case, then, in which B wants A to do X but A doesn't want to do X. Would J and K agree that B should coerce A to do X? J and K have to adopt a principle to cover this kind of case as if they have no more idea of whether they will be in A's role than they have of whether they'll be in B's. Suppose B suggests as a principle to govern this case that A do X against A's will whenever B wants him to. This principle not only prescribes what A is to do, namely, he is to do X rather than some other action Y, it also prescribes the conditions under which he is to do X, namely, whenever B wants him to. If A wants to know why he is to do X rather than Y in a particular situation, these conditions constitute an explanation of why he is to do X rather than Y. He is to do so because B wants him to. This principle, then, not only prescribes what to do, it

also supplies the reason for doing it. In fact, the reason is the strongest part of the principle in that A is to do X whenever B wants him to but, if B doesn't want him to do X, the principle leaves it open as to whether or not A is to do X.

Accepting B's principle, then, means accepting it as a reason for A's doing X that B wants A to do X. J and K, however, do not know if they are A or B. If they were B, they would promptly accept B's principle but if they were A they would equally promptly reject it. No rational agent A will agree to a principle that allows B to coerce him to do X just because B wants him to do X, especially as the nature of X is not specified. It may be the act that A most wants not to do.

Nonetheless, as a rational agent, A may be prepared to accept B's principle if B is prepared to accept the symmetrical principle, that is, if B accepts the principle that A coerce B to do X whenever A wants B to do X. Actually, the symmetrical principle need not be cast in terms of A's coercing B to do the act X. It could be cast instead in terms of any act equivalent to X. An act Y will be said to be 'equivalent' to X when A wants to do Y as much as B wants to do X or A wants not to do Y as much as B wants not to do X. A, after all, may not want to coerce B to do X so he will not necessarily accept a symmetrical version of B's principle if he is required to coerce B to do precisely the same kind of act as B wants him to do. Furthermore, B may himself quite willingly do X. If this is so, A will not agree to B's principle if B agrees to the symmetrical version because this would mean that A agrees to A's being made to do something A doesn't want to do but that B wants him to do, without thereby gaining the parallel power to make B do something that B doesn't want to do but that A wants him to do. A, then, will only agree to the symmetrical version of B's principle if A has the power to make B do something that he doesn't want to do that is as

unpleasant to B as A's doing X is to A. It does not need to be the particular act X itself. Nonetheless, for expository convenience, I shall express principles as if both A and B had the same attitude to X, i.e. either they both wanted to do it or both didn't want to do it and that their desires in each case are equally strong. When a particular conclusion I wish to draw rests on this assumption, however, I will take into account the degree to which the conclusion would have varied if the assumption didn't hold.

If both A and B are prepared to accept what I will call the 'symmetrified' version of B's principle, then, when J and K consider the principle from the viewpoints of each of A and B, they will be prepared to adopt it, provided that the principle makes sense when symmetrified. Not all principles, however, will make sense when symmetrified in that they will fail to be guides to action because they give self-defeating advice. Symmetrification, then, can be used to eliminate principles in much the same way Kant tried to use the Categorical Imperative or many modern philosophers try to use 'universalisation'.

As it turns out, B's principle, at least in the form used so far, is self-defeating when symmetrified. If B has the power to make A do X against A's will, then A cannot have the power to make B do Y against B's will, at least where Y covers whatever acts B has to perform to coerce A to do X. B's principle, then, when symmetrified, has logically impossible results for particular cases. It can, however, be modified to avoid this consequence by making both A and B free to try to coerce each other, or by giving them each the right to try to coerce each other, that is, by adopting a principle to the effect that no outside party interfere with their attempts to coerce each other.

J and K, then, can agree to B's being free to coerce A provided A is free to coerce B. J and K, however, will only agree to this if A is prepared to agree to B's being free to coerce A and B is prepared to agree to A's being free to coerce B. Both A and B, therefore, need to be people whose desire to coerce the other is stronger than their desire not to be coerced. They must want the chance to coerce the other so much that they are prepared to risk being coerced instead of doing the coercing.

There is no guarantee, however, that A and B both want to coerce the other more than they want not to be coerced. ^{Think about better} There are two symmetrical principles that A and B could adopt here, either Principle P that A not be free to coerce B provided B is not free to coerce A or Principle Q that A be free to coerce B provided B is free to coerce A.

There are, then, three possibilities, namely, Possibility (1) that both A and B accept Principle P, Possibility (2) that both accept Principle Q and Possibility (3) that A accepts one of P or Q while B accepts the other. The position of equal power and the veil of ignorance, however, rule out Possibility (3). As J and K do not know which of A or B they are they will have to choose a principle that covers them both. They will therefore have to decide between both A and B being bound by P or both being bound by Q.

Possibility (2) only obtains if A and B both want the other to do X more than they want not to be made to do it themselves. Possibility (1) only obtains if A and B both want not to be made to do X more than they want the other to do it. This means that J and K now have to arbitrate between these desires. They have to determine which desire is to be satisfied - the desire that the other person do X against his will even if this means that one may be made to do X oneself against one's will or the desire

that one not be made to do X against one's will even if this means that the other person can't be made to do X against his will.

Both of these desires have passed the test of equal power and the veil of ignorance. They do not assume that priority is to be given to one's own case. Anyone who accepts Principle P accepts equal limitations on his own behaviour as on that of others. If A accepts Principle Q then he is prepared to accept that X may be done to him against his will as the price that X may be done to B against B's will. A is not claiming anything for himself that he is not prepared to claim for the other person. He is not giving his own case special treatment. He is not, that is, an egoist.

Those who desire that X be done to the other against his will more than they desire that it not be done to them against their will I will call coercionists. Those who desire that X not be done to them against their will more than they desire that X be done to the other against his will I will call noncoercionists.

Both the noncoercionist and the coercionist are prepared to accept this symmetry of reason-giving. Neither is claiming that what is to be a reason why his desires should be satisfied is not to be a reason why other people's desires should be satisfied. He is not claiming, as the egoist does, that the relevant difference as to why something is a reason in his case and not in the case of others is that it is he whose case it is.

How, then, are J and K to settle the dispute between the noncoercionist and the coercionist? Settling the dispute amounts to deciding, not merely whether Principle P is to be preferred to Principle Q but, further, whether X is to be done to A or to B, if Principle Q is adopted. A settlement of the dispute requires a determinate outcome when there is a conflict of desires between A and B. If both A and B want X done to the other less

turns
press. v. r

than they want X not done to themselves, that is, they would choose Principle P, then J and K would agree on the determinate outcome that X is to be done to ^{by} neither A or B, even if they turn out to be A who wants X done to B or B who wants X done to A. ^{At the X} If both A and B want X done to the other more than they want X not done to themselves, that is, they would choose Principle Q, then, even if J and K agree to follow Principle Q, they have still not resolved the question of whose desire is to be satisfied. Is it to be A, who wants X done to B but not to A, or B who wants X done to A but not to B?

With Principle P, no-one is to be free to do X to either A or B just because he wants to. 'Freedom', here, is a term that indicates how the law stands. He will not be free under the law. Alternatively, the code of behaviour of the society will rule it out. He is still free in the sense that he will not be bound hand and foot, incapable of doing X to either A or B, nor will his thoughts be monitored so that his mere intention to do such an act is immediately detected and thwarted. Rather, he is to be constrained by sanctions or threats of them, by an efficient detection system, etc.

If A is free to do X to B and B is free to do X to A, then, given that A wants to do X to B and symmetrically for B, then which of these desires are J and K to agree is the one to be satisfied? We have already seen that A and B will agree to Principle Q if they both want to be free to do X to the other more than they want the other not to be free to do X to them. What J and K have to choose between is Principle P or Q. Now, if A and B have accepted Principle Q, they have accepted that the outcome of who is to do X to whom is to be determined solely by the natural capacities of A and B, and by luck. This means that, in most cases, whichever of A and B has the superior power will be the one who does X to the other. In

accepting Principle Q, then, A and B are accepting that whoever does X to the other against his will is to be determined either by superior power or luck. This fact supplies J and K with a basis for choosing between Principle P and Principle Q if J and K agree that which of A or B does X to the other is not to be determined by superior power or luck. If it is a premiss that J and K share in the arguments they have about which of A's or B's desire is the one to be satisfied that this is not to be determined by superior power or by luck, then they will reject (Q) and agree on (P). Further, anyone who accepts (Q) thereby accepts ultimately that which of two incompatible desires involving coercion is to be satisfied is to be resolved by superior power or luck. If we discover that the person we are arguing with holds this view, then, as rational agents, it would seem appropriate either to abandon the argument and to get ourselves the necessary superior power, or, alternatively, to change the argument into a mere substitute for superior power.

We have, then, two types of people. The first type contains both the egoist and the coercionist. The egoist has the view that conflicts of desires are to be resolved in favour of satisfying the desires of whichever party to the conflict he is. This means that, if person A is an egoist, then, when he argues with person B over what principles to adopt to resolve a conflict of desires between them, he has already decided to use whatever force or deception is necessary to satisfy his own desires to the full regardless of the cost to B. He will only modify the satisfaction of those of his desires that thwart the satisfaction of B's to the extent that B has the power to make him.

The coercionist has the view that conflicts of desires are to be resolved in accord with the principles of whoever has the superior power or the luck. If A is a coercionist arguing with B over which principle should

resolve a conflict of desires, then A already has the belief that it should be the principle of whichever of A or B is able to have his way. If A is able to get his way by deception, or superior power, or luck, then A is prepared to force B to resolve the conflict in accord with A's principle regardless of the effect of this on B's desires or principles. Both the egoist and the coercionist are not sincerely committed to argument as a way of resolving their dispute. Their engagement in argument is a pretense. No matter what agreement is reached by argument they will dishonour it as soon as a shift in the balance of power permits it. Even if agreement could be reached in a position of equal power as a result of argument, they have no sincere intention of abiding by it if they have the superior power or the better luck. They remain partial to their own cases, regardless of the distribution of power. I shall, therefore, call them 'partialists'.

A person who sincerely intends to abide by the agreements people would reach behind a veil of ignorance, in a position of equal power, even when he has the superior power in real life, I will call an 'impartialist'. As we have seen, a person will only sincerely have this intention if he is not a coercionist. This means that people behind the veil cannot argue as if they wanted to coerce people more than they wanted not to be coerced, even when what they want to coerce others to do is to resolve conflicts of desires in accord with some unselfish principle rather than in accord with whichever outcome satisfied their desires no matter who they were. Nagel (1975) is concerned that this aspect of the veil makes it biased. I shall turn to an examination of whether this is so.

11. Ideals

The egoist's only concern in the selection of principles for resolving conflicts of desires is that they result in the satisfaction of his desires, whatever they happen to be. Both the impartialist and the coercionist, on the other hand, are not concerned that the desires they personally happen to have as conflictants in the conflict of desires are the ones that are satisfied by whatever principle is used to resolve the conflict. Rather, they are concerned to see that the conflict is resolved in accord with their preferred principle, regardless of the effect this has on them personally. I will describe their desire to see conflicts of desires resolved in accord with a principle, rather than in favour of whatever desires they happen to have as parties to the conflict, as their 'ideal'.

A patriot is a person with the ideal that his nation should triumph regardless of the effect on him of its doing so. A saint is a person with the ideal that the tenets of his religion be followed regardless of its effect on him. Hare's fanatic follows ideals in this sense. His Nazi who continues to believe that Jews should be killed even when it turns out that he is a Jew is a man with an ideal. Even utilitarianism is an ideal in this sense. A sincere utilitarian does whatever is required by the principle of utility regardless of its effect on him in particular. Admittedly, its effect on him is part of his utilitarian calculations but if these calculations require a great sacrifice, even death, of him, then he will not refuse to do it because it is him that the principle is having this particular effect upon.

The egoist, then, has no ideal. The coercionist, however, has an ideal but, as we saw in the previous section, it is an ideal he is prepared to force on others even though they would not agree to it in a position of equal power. Impartialism, then, will reject those ideals that would only

be adopted by J and K in a position of equal power behind the veil of ignorance if the conflictants A and B were both coercionists.

In order to do so, however, it must deny J and K knowledge of which ideal, or conception of the good, each has, that is, it must treat conflicts between ideals as just another kind of conflict of desires. J and K know that A has one ideal and B has another, incompatible one but neither J or K knows which of A or B they are likely to be. Impartialism, then, is identical to Rawls' theory with respect to denying the parties behind the veil any knowledge of their own conception of the good. Nagel (1975) believes that this actually biases the veil in favour of some conceptions of the good. It does so, he thinks, because the veil must combine with some conception of the good if it is to produce principles to which the parties behind the veil give unanimous agreement. He claims that Rawls introduces this conception of the good into his theory through the 'primary goods'. The conception of the good that he thinks is favoured by Rawls' theory is a 'liberal individual conception according to which the best that can be wished for someone is the unimpeded pursuit of his own path, provided it does not interfere with the rights of others'. (1975, 10). It discriminates against conceptions of the good that 'depend heavily on the relation between one's own position and that of others'. (1975, 9). This liberal individualist aspect of Rawls' theory has also been criticised by other writers, for example, Fisk (1975), Francis (1980). I argue in Chapter 11 that impartialism does not need the primary goods. Nagel regards the primary goods as the source of Rawls' particular liberal individualist bias. His argument, therefore, provides no reason to think this exact bias is also in impartialism. Nagel's point, however, is that any theory which denies the parties behind the veil of ignorance any knowledge of their conception of the good will only get unanimous agreement out of these parties if it has imported a conception of the good in elsewhere. He says (1975, 8)

'It is true that men's different conceptions of the good divide them and produce conflict, so allowing this knowledge to the parties in the original position would prevent unanimity. Rawls concludes that the information must be suppressed and a common idea substituted which will permit agreement without selecting any particular conception of the good ... Another possible conclusion, however, is that the model of the original position will not work because in order to secure spontaneous unanimity and avoid the necessity of bargaining one must suppress information that is morally relevant, and moreover suppress it in a way that does not treat the parties equally'.

Nagel, I think, is right about this. Impartialism, then, must use the veil in combination with some conception of the good. The concept of the good that impartialism combines with the veil is that it is good that conflicts of desire be resolved by unforced agreement. This concept is given substance in the position of equal power and the non-coercion requirement. There could, then, be a conflict of desires between Group A that favoured this concept of the good and Group B that didn't. We could apply the veil to this conflict. If Nagel is right, the veil by itself will be unable to resolve the conflict, except that it will filter out any egoistic resolutions of the conflict. With the other two impartialist axioms however, the veil also filters out coercionist resolutions of conflicts of desires, that is, it filters out coercionist ideals. Nagel thinks this denies the parties access to morally relevant information. Certainly, it filters out information that would enable the parties to agree that A and B are to resolve conflicts in whatever manner suits whichever of them has the greater power or luck. It only filters out morally relevant information then if information that enables the parties to agree to settle conflicts according to the will of the strongest is morally relevant. Even if we define 'morality' so that this is morally relevant, it doesn't show why we should want the veil to take account of information that is morally relevant in this sense. Any argument that we should take account of this information is an argument that we should not be impartialists. Nagel's criticism, then, is a challenge to show why we

should adopt the conception of good that produces unanimous agreement when combined with the veil. If applied to impartialism, it is the challenge to show why we should give unanimous consent to impartialism as an ideal rather than to some partialist ideal. I take up this challenge in Chapter 16. Nagel appears to accept that egoistic resolutions of conflicts of desires are morally irrelevant ones, so he is not also asking for a justification as to why we should prefer impartialism to egoism or why, for that matter, any ideals are to be preferred to egoism, even if they are coercionist. I will also discuss why we should prefer impartialism to egoism in Chapter 16.

12. Impartialism and Ordinary Life

An impartialist, then, believes that everyone should act in accord with the principles that people would reach in a position of equal power and behind a veil of ignorance, provided that these principles did not entail that the ultimate determinant as to which of A or B had his way was whichever of them possessed the superior power. As I noted earlier, people in ordinary life are not behind veils of ignorance nor are they usually in positions of equal power. Nonetheless, they can argue about who should do what as if they were impartialists. Anyone who knowingly does not do so is a person who can only be persuaded by sanctions. They are committed to superior power as the only resolution of disagreements between persons. The only kind of argument that is not merely a substitute for sanctions is argument within impartialism. It is the only form of argument that takes as its premiss the full equality of the participants in the argument. If person J uses impartialist arguments then he is assuming neither that his case has automatic priority over that of other people nor that the outcome of the argument is properly determined by accidental inequalities between the participants. A partialist, however, is committed to one of those two

assumptions. This is not to say that an impartialist will not agree that sometimes the outcome of an argument is properly determined by certain inequalities between people but that which inequalities these are should itself be determined in a situation where these inequalities are neutralised. If, for example, it is agreed by J and K that greater physical strength properly determines the outcome of an argument, this cannot be because J has used his greater physical strength to get K to act if he agrees with J about this.

When two people in ordinary life have an argument, then, either it is the case that they treat each other with impartialist respect, that is, as impartialist equals, or the argument is really just a substitute for force. Impartialism provides a test whereby to identify the nature of the persons arguing. When we argue with each other we assume that the other person is seriously committed to a resolution that has a basis other than superior power. If we discover that this is not so, then this will change the point of the argument. I will discuss exactly how the point of the argument is changed and what consequences this has in more detail later. What I wish to do now is to look at the kinds of principles that impartialists would agree to. As we have seen, impartialists will adopt principles that would be rejected behind a Rawls-like veil of ignorance. I will therefore look in some detail at the principles Rawls thinks would be adopted behind his veil and the kinds of arguments he advances for them. Before I do that, however, I would like firstly to clarify the two-party nature of the argument so far and, secondly, to distinguish between fallible and infallible impartialists.

13. Two-Party Conflicts of Desires

I have cast the discussion so far in terms of two parties. These parties are the individuals in the conflict A and B who have been required to adopt the viewpoint of the velorians J and K who treat A and B as characters in a scenario. I have done so because all conflicts of desires involving more than two parties can be reduced to conflicts between two parties. Suppose there is a conflict of desires between A and B where both A and B are not individuals but groups. Say that A wants some state of affairs R to obtain whereas B does not want R to obtain. In not wanting R to obtain, however, it does not follow that B is unanimous that some other state of affairs obtain instead. Suppose that some of B want S to obtain while others want T. We, then, have Conflict 1 in which we can label A as A₁ and B as B₁. Let's resolve this conflict in favour of B, that is, some state of affairs other than R is to obtain. We now have Conflict 2. Let's suppose that the people who were in A now become members of one of the two groups in B. Let's call those who want S by the name A₂ and those who want T by the name B₂. Hopefully, J and K will resolve this conflict, giving us a determinate answer as to which state of affairs should obtain. This process can be adopted no matter how many parties there are to the conflict. The more parties there are, the greater the number of conflicts generated out of Conflict 1 before we reach Conflict N where the conflict is resolved. Likewise, the greater number of parties A and B before we get to A_N and B_N.

I will not, then, be using the term 'conflict of desire' in the sense it is often used in the literature, for example, Williams (1973, 167). He uses the term to refer to conflicts between the desires of the same person. I will only use the term, however, of conflicts between the desires of more than one person. The desire of A has to be incompatible with the desire of

B before I will describe it as a conflict of desires. The term 'conflict of desires', then, will be used to cover 'interparty' conflicts. It may be that one of the differences between rationality and morality is that rationality has to do with conflicts of desires within one party, whereas morality has to do with conflicts of desires between different parties.

14. The Impartialist Axioms

As a result of the discussion in the previous sections I have identified three conditions that must hold of the parties engaged in arguing about resolving a conflict of desires if they are not the kind of people who either want superior power to resolve the conflict in their own favour or who want the conflict resolved in favour of whoever has the superior power or luck. Those conditions are that they must argue as if they are in a position of equal power, as if they are behind a veil of ignorance and as if they are noncoercionists. Each of these conditions can be translated into an axiom for the theory of right I call 'impartialism'. The 'equal power' axiom says that people are to act in accord only with principles agreed to in a position of equal power. The 'veilorian' axiom says that people are to act in accord only with principles agreed to behind a veil of ignorance. The 'noncoercionist' axiom says that people are to act in accord only with principles that would be chosen by people who want not to be coerced more than they want to coerce.

15. Fallible and Infallible Impartialists

We began this discussion with two people in ordinary life A and B. Step by step we limited the ways in which they go about resolving conflicts of desire between them until we had arrived at the conditions that would have to prevail if their resolution of the conflict was to be achieved by an unforced agreement. These two people A and B only reach unforced agreement

if they settle their conflict as if they were in a position of equal power, behind a veil of ignorance and as if they were noncoercionists. They have to consider their own actual cases as A and B as hypothetical cases scrutinised by themselves in the role of J and K. Even though A and B argue as if they are noncoercionists in a position of equal power, behind a veil of ignorance they cannot avoid bringing to the argument whatever limitations they possess. The kinds of conclusions they will reach, no matter how conscientious they are in their efforts to reason as velorians, may turn out to be quite different to the conclusions that velorians without these limitations would have reached. Just because A and B try to reason as if they were J and K they still retain the knowledge of A and B, the susceptibilities to logical error of A and B, and so on. It is even possible that they unconsciously argue as if they know which of A and B they really were but this is something they might be able to overcome by an effort of the will once it is pointed out to them. They cannot overcome factual ignorance and the inability to see the point of an argument by the same means. At their best, then, A and B will still be very fallible rational agents. This fallibility will affect which principles it is they reach unforced agreement on. What we get in the real world, then, as a result of people applying the impartialist method, will be a set of provisional principles. They are provisional because they are not necessarily the principles that sincere impartialists would agree upon if they have their facts right, they have included all the relevant facts, and they have made no logical errors. While these provisional principles are the ones that an impartialist will have to work with in any particular place at any particular time, he would prefer to follow the set of unprovisional principles. These are the principles that truly infallible and omniscient velorians would agree to. They are the ideal to which impartialists aspire when they disagree with each other about which principles the velorians really would have endorsed.

When they argue with each other about which act is the right one to do in this situation, they are out to demonstrate that this is the act that would be required by whatever principle J and K would enjoin to govern this kind of conflict of desires were J and K infallible and omniscient. If A and B are both sincere impartialists they will amend their opinions in the direction of the unprovisional impartialist principle. If A can show B that B's resolution of the conflict rests on a factual or logical error or oversight then B will modify his position.

16. Conclusion

We now have the essential apparatus we need to determine whether or not a particular resolution of a conflict of desires is one that would obtain unforced agreement or not. This apparatus can be used to establish action-guiding principles to cover conflicts of desire in ordinary life. Rawls has already derived a number of such principles from the veil of ignorance. Rawls, however, believes that the scope of the veil needs to be limited in various ways. These limitations affect the kinds of principles the veil will generate as the resolution of conflicts of desires. I will now look at the limitation Rawls places on the veil to see whether these same limits need to be placed on its use in impartialist theory.

Chapter 3

Rawls' "Circumstances of Justice"1. Introduction

In the previous chapter I showed what conditions the argument between two parties with conflicting desires would have to meet if the principles they were to adopt to resolve the conflicts were impartialist ones, that is ones where the parties to the argument assume neither that the outcome of the argument should favour whichever desires they happen to have nor that the outcome of the argument should favour whichever of the parties has the superior power or the better luck. The conditions that make an argument an impartialist one were that the parties J and K argue (1) as if they were in a position of equal power (2) as if they were behind a Rawls-like veil of ignorance and (3) as if they wanted to coerce others less than they wanted NOT to be coerced. The conditions can be imposed on any argument over conflicting desires. Rawls, however, restricts the application of his veil of ignorance to those conflicts of desires whose resolution raises questions of justice. As he has selected the features of his veil of ignorance so that it results in principles that he considers 'just' ones, it may be that his veil of ignorance needs to be modified in various ways to make it suitable for impartialist arguments where the conflict of desires does NOT raise questions of justice. If this is so, then the features of the initial situation in which J and K are to be placed if their principles are to be impartialist ones may NOT be the same as those of the initial situation of parties whose principles are to be 'just' ones.

Rawls believes that the principles that the parties behind the veil will produce will only result in a just basic structure of society in the conflicts of desire they are designed to resolve occur within the 'circumstances of

justice'. Rawls does not apply the veil of ignorance to conflicts of desire that lie outside the circumstances of justice. Nonetheless, the veil may still be able to produce principles that resolve these conflicts. It is not my purpose here to examine whether Rawls is right in claiming that only those conflicts of desires that fall within his circumstances of justice are capable of just resolution. Rawls, like a number of other contemporary political philosophers such as Lucas (1966) and Richards (1971), has developed his account of the circumstances of justice from that first presented by Hume. The problem with both Hume's and Rawls' account of the circumstances of justice as limitations of the application of the concept of justice are discussed in detail by Hubin (1979). My concern, instead, is whether or not there is any good reason to limit the scope of Rawls' veil to only those conflicts of desires that fall within his 'circumstances of justice'.

Rawls divides the circumstances of justice into two categories, namely, (1) objective circumstances and (2) subjective circumstances. I will look at the objective circumstances first.

2. The Objective Circumstances of Justice

I wish to consider three of the objective circumstances of justice that Rawls proposes. I will call these (1) the Man Friday circumstance, (2) the humanhood circumstance, and (3) the moderate scarcity circumstance. I will look at each in turn to see whether there is any good reason why the veil of ignorance should not be used to resolve conflicts of desire that lie outside those circumstances.

3. The Man Friday Circumstance

Rawls says that,

'the circumstances of justice may be described as the normal conditions under which human co-operation is both possible and necessary'. (1972, 126)

This implies that the circumstances of justice do not cover cases of what humans do in situations where co-operation is neither possible nor necessary. One of these is the Robinson Crusoe case of someone marooned alone and incommunicado from the rest of the world on a desert island. Rawls only uses the veil of ignorance to resolve conflicts of desire that occur within the circumstances of justice. If this is the limit of its application, then it would appear to have no role to play in providing principles about how Robinson Crusoe should behave on his desert island. It would appear to have a role to play only after Man Friday turns up. Until then, it is not possible for Crusoe to engage in human co-operation, if only for the purely tautological reason that it takes two to co-operate. Until Man Friday is present there is not the same necessity to co-operate either as there is once he turns up. Crusoe may at times be unable to do things alone that he would be able to do if Friday were there but, then, neither does he have to defend himself from the things Friday might wish to do to him.

Nonetheless, people do disagree about how a person should behave when alone, even when he is alone on a desert island. For example, many people would think that Robinson Crusoe shouldn't masturbate on his desert island, whereas others would think that he should if that's what he feels like doing. Is this conflict one in whose resolution the veil of ignorance has no role to play due to the fact that human co-operation in such a case is neither possible nor necessary?

We will be in a better position to answer this question if we are clear about the roles of the two parties J and K. As we have seen, J and K are not the ones who have the desires which are in conflict. Rather, they propose principles for the conflict of desires that exists between two parties A and B. The veil of ignorance is merely a device by which to limit the principles that J and K propose to ones that do not depend on their being one particular person rather than another.

Suppose, then, that A wants to masturbate on his desert island. It is the role of J and K to determine whether or not he should. This involves their proposing a principle to cover A's situation. A's situation, however, is not one in which there is a conflict of desires. A conflict of desires, as I said in the previous chapter, has been defined to require at least two parties. J and K, therefore, have no task to perform, that is, they have no conflict of desires to arbitrate. Robinson Crusoe's case does fall outside the scope of the veil of ignorance but only because it provides no conflict of desires which J and K have to settle by finding a principle they would regard as acceptable regardless of whether they turned out to be A or B. If J and K are to have a conflict of desires to resolve with respect to Robinson Crusoe-type cases, then they cannot consider cases where only the desires of the person on the desert island are taken into account. Instead, they will have to consider the case where A and B have a conflict of desire over how Crusoe should act. Suppose A wants Crusoe to do whatever Crusoe feels like doing provided this doesn't physically harm Crusoe whereas B wants Crusoe to abstain from masturbation. J and K are now in a position to postulate principles about which of A's and B's desire is to be satisfied here. In doing so, they will be coming up with a principle which determines how Crusoe should act on his island, even though he is alone. The argument between J and K as to which of A's or B's desire is to be satisfied being an impartialist one, J and K will have to argue as if they were behind the veil of ignorance, even though their

reasoning will produce a principle that determines how a person should act who is outside Rawls' circumstances of justice. We see, then, that the veil of ignorance can be used to produce principles concerning cases like the Robinson Crusoe case which lie outside the circumstances of justice, provided J and K argue about which of A's or B's desire about Crusoe's behaviour is the one to be satisfied. It is only when J and K argue directly about whether or not Crusoe's desire is to be satisfied that they cannot use the veil to come up with any principles. This, however, is not due to any limitation inherent in the veil. It is not as if the principles generated by the veil cannot be about Robinson Crusoe-type cases. Rather, it is due to the fact that the situation to which the veil is being applied does not contain a conflict of desires, and therefore there is nothing for the veil to resolve. It would appear, then, that the veil can apply to any Robinson Crusoe-type case as long as it is construed in terms of the desires of A and B about Robinson Crusoe's behaviour. The Man Friday circumstance is not a restriction on the application of the veil in an impartialist argument.

4. The 'Humanhood' Circumstance

Rawls does not regard the circumstances of justice as covering cases of what humans do to non-humans. As he only applies the veil to cases that fall within the circumstances of justice, this means that he does not think that the veil provides principles about how humans should treat non-humans.

In the previous section, we saw that an apparent limitation of the veil was due, not to any inability of the veil to handle principles about any kind of topic, but rather to its only applying to situations where there was a conflict of desires. The 'humanhood' circumstance throws a similar

light on the applicability of the veil. Just as the veil can only apply to a case if the case involves a conflict of desires, so the veil can only be applied by J and K if J and K are humans, or, at least, persons. It is not that the veil cannot be used to resolve conflicts of desires about how humans should treat non-humans. As we shall see, it can be used for this purpose. As with the Man Friday circumstance, the humanhood circumstance doesn't tell us anything about the scope of the principles generated by the veil. It does not put limits on what these principles can be about. Rather, it reminds us that the parties behind the veil, that is, J and K, have to be capable of argument. They have to be rational, self-conscious agents, able to understand the consequences of their decisions, concerned about their own future states as well as their present ones, and so on. J and K have to be the kinds of beings that make argument about the principles for resolving conflicts of desires meaningful. I will follow Singer (1979, 75) and call such beings 'persons'. As the only beings capable of Rawlsian-type contractarian argument appear to be humans, this may suggest that a device such as the veil only applies to cases of how humans treat humans. This is not so. One set of humans may want humans to exploit animals regardless of the sufferings to animals this causes. Another group of humans may want to stop this kind of exploitation. There is here a clear case of a conflict of desires. J and K therefore can employ the impartialist apparatus, including the veil, to resolve this conflict of desires. In deciding which of A's desire or B's is to be satisfied, J and K will be coming up with principles about how humans should treat animals, even though the animals themselves were NOT, and were incapable of being, parties to the generation of these principles. Once again, a distinction needs to be drawn between what the principles that the veil generates can be about and the conditions that need to be met before there is an argument at all. There is only an argument if the parties to it are persons, just as there is only an argument if there is a conflict of desires.

Just because Rawls restricts his use of the veil to the generation of principles about how humans should treat humans, it doesn't follow that the veil cannot be used to generate principles about how humans should treat non-humans. It merely means that Rawls' theory will provide us with no guidance in this area unless supplemented in some way. As Haksar says,

'You cannot make such deals with beings, such as animals, who are below the minimum level of rationality. Now as a result of these deals you may enter into certain obligations with regard to those with whom you have made them. Rational contractors can have special obligations towards non-rational contractors. Let us call these contractarian obligations. Now rational human beings can have two kinds of duties: contractarian obligations and non-contractarian obligations (or duties).' (1979, 25)

What this discussion of the 'humanhood' circumstance shows us is that principles need not be only about the relationships between rational agents even though the parties who develop these principles must themselves be rational agents. The veil can be used to arrive at principles that constrain how rational agents treat non-rational agents such as animals or, possibly, even such things as trees, rocks and eco-systems. While Rawls may be right that principles about the relationships between humans and non-humans cannot be principles of justice, nonetheless they are principles that resolve conflicts of desire between humans as to how non-humans should be treated. Rawls has concerned himself only with principles that fall within the circumstances of justice. His doing so, however, does not show that conflicts of desires about matters lying outside these circumstances do not also merit resolution or that the veil cannot be used in producing resolutions.

5. The 'Moderate Scarcity' Circumstance

Rawls believes that questions of justice only arise in conditions of moderate scarcity. Conditions of moderate scarcity are to be contrasted with what I will call conditions of 'extreme scarcity' on the one hand and conditions of 'superfluity' on the other. Conditions of extreme scarcity are those where conditions are 'so harsh that fruitful ventures must inevitably break down'. Conditions of superfluity are those where 'natural and other resources' are 'so abundant that schemes of co-operation become superfluous'. (1972, 127)

Rawls may be correct in his view that questions of justice do not arise in conditions of extreme scarcity or of superfluity. It may still be the case, however, that conflicts of desire can occur about what happens in these circumstances, as they did about what happens in Robinson Crusoe-type situations and situations involving the treatment of non-humans by humans. It may be the case, for example, that one set of people in conditions of superfluity want to engage in the mindless destruction of an endless supply of plant-life whereas others do NOT want them to. J and K will have to use the veil of ignorance to come up with principles to resolve this kind of conflict. In conditions of extreme scarcity, there will undoubtedly be conflicts of desire. Rawls exempts both kinds of conflict of desire from the scope of the veil because he sees them as cases either where human co-operation is not possible, as in the case of extreme scarcity or else as cases where human co-operation is not necessary, as in conditions of superfluity. Exactly what the scope of the veil is, then, in his opinion, will depend on how he understands the notions of co-operation being 'possible' and its being 'necessary'. No matter how extreme the scarcity, it is difficult to see that co-operation is not possible, even in cases where people's lives are at stake. Consider the case of people

adrift in a lifeboat faced with the decision of throwing one or more overboard in order to save the rest. It may be that Rawls is correct in believing that such a case falls outside the circumstances of justice, although, even in such a case, it would seem that there are fair and unfair ways of deciding who has to die.

Two historical cases described by Glover (1977, 204) illustrate this point. He says,

'In the case of U.S. v Holmes (1842) the defendant was a member of a ship's crew in an overcrowded lifeboat, who obeyed the orders of the mate and helped to throw overboard sixteen male passengers. The judge, directing the jury to convict him on the charge of manslaughter, said that the first people to have been sacrificed should have been chosen by lot.'

In the other case, R v Dudley (1884), it was found that shipwrecked sailors who killed a cabin boy for food were guilty of murder, despite the likelihood that, without this act of cannibalism, all would have died, with the already weakened cabin boy probably dying first.'

As I said earlier, however, it is NOT my purpose here to dispute Rawls' account of 'justice'. I am interested, instead, in whether there is any good reason why the veil cannot be used to resolve conflicts of desires in such cases. If A wants to throw his enemy off the boat, and B wants them all to draw straws then there is a conflict of desires that J and K can try to resolve by an appropriate principle arrived at from behind the veil. The extreme scarcity involved does not seem to present an obstacle to the use of the veil in this situation. Admittedly, this is a situation in which human co-operation is possible, in that the people on the boat could agree that one of them has to be thrown overboard and that this person is to be the one who draws the short straw. As a consequence, Rawls might argue that this isn't a situation of extreme scarcity. Such a move, however, runs the risk of making the notion of 'extreme scarcity' an empty

one. Likewise, even in a situation of superfluity, co-operation may be necessary for cases like the mindless destruction of abundant plant-life. To say that the mere fact that co-operation is necessary for this kind of case means that the situation isn't one of superfluity runs the risk of rendering the notion of 'superfluity' empty.

So far, then, we have seen that there are no good reasons to restrict the use of the veil to resolving only those conflicts of desire that fall within these objective circumstances of justice. Our discussion of the objective circumstances of justice has, however, shown us (1) that there must be a conflict of desires involved in a case before the veil can be used to arrive at principles about how people should act, (2) that the veil can be used to arrive at principles about how humans should treat non-humans provided that the parties who agree on the principles are beings of a kind capable of arriving at, and living in accord with, the principles they determine, (3) that the veil can be used to arrive at principles about how people should behave in conditions of superfluity and extreme scarcity provided the notions of 'superfluity' and 'extreme scarcity' are not defined in such a way as to make co-operation in such conditions logically impossible.

6. The Subjective Conditions of Justice

In Rawls' view, not only do certain objective circumstances have to exist before questions of justice arise but also certain subjective conditions have to be met. I will be interested in three of these, which I will call the 'dissensus' circumstance, the 'mutual disinterest' circumstance and the 'fallibility' circumstance. Like the objective circumstances, these are circumstances that Rawls believes have to apply to people in ordinary life before their situation is one that is properly the subject of questions of justice. Both the objective and the subjective circumstances

of justice, then, are meant to obtain for the characters A and B, not for the parties J and K. J and K only have a job to do when A and B find themselves in the circumstances of justice, either objectively or subjectively. As we have seen, J and K can still arbitrate conflicts of desires between A and B even when A and B are not in the objective circumstances of justice. These conflicts of desire may have nothing to do with justice as such, if Rawls' account of justice is correct, but they are still conflicts which the veil of ignorance can be used to resolve. I will now look at Rawls' subjective circumstances of justice in order to see whether J and K can try to resolve conflicts outside these circumstances by using the veil of ignorance.

7. The 'Dissensus' Circumstance

I borrow the word 'dissensus' from Alasdair MacIntyre (1983, 590) where he says that dissensus is the obstacle to meaningful conversation between the various groups in society because "there are too many rival conventions, too many conflicting anecdotes; and the repetition of assertions and denials does not constitute conversation". It is precisely this kind of dissensus that the impartialist conditions of argument are intended to overcome.

In Rawls' view, the disputes between people that the veil of ignorance resolves arise because each person has his own plan, or conception of the good, that leads people 'to have different ends and purposes, and to make conflicting claims on the natural and social resources available.'

(1972, 127)

This raises the question of whether the veil of ignorance has a role to play only if there is an actual conflict between A and B or whether it also applies to hypothetical conflicts of desire?

I can see no reason why the veil cannot be used to anticipate the principles that will be needed to resolve conflicts of desires that might happen. Also, it is possible that the agreement reached by the ordinary people A and B is a case of forced consent as discussed in the previous chapter. It is, therefore, not even the case that consensus between A and B obviates the role of the veil. It may well be that there is a lack of dissensus between A and B solely because one of them has acted contrary to the principles that J and K enjoin A and B to follow.

8. The 'Mutual Disinterest' Circumstance

Rawls introduces the notion of 'mutual disinterestedness' as follows,

'Although the interests advanced by these plans are NOT assumed to be interests in the self, they are the interests of a self that regards its conception of the good as worthy of recognition and that advances claims in its behalf as deserving satisfaction. (1972, 127)

He elaborates the notion as follows,

'The assumption of mutually disinterested rationality then, comes to this: the persons in the original position try to acknowledge principles which advance their system of ends as far as possible. They do this by attempting to win for themselves the highest index of primary social goods, since this enables them to promote their conception of the good most effectively, whatever it turns out to be. The parties do not seek to confer benefits or to impose injuries on one another; they are not moved by affection or rancor. Nor do they try to gain relative to each other; they are NOT envious or vain. Put on terms of a game, we might say: they strive for as high an absolute score as possible. They do not wish a high or a low score for their opponents, nor do they seek to maximize or minimize the difference between their successes and those of others. The idea of a game does NOT really apply, since the parties are not concerned to win but to get as many points as possible judged by their own system of ends.' (1972, 144-145)

What is the effect of the mutual disinterestedness circumstance with respect to the veil of ignorance? It limits the veil's application to only those conflicts of desires between A and B that are not due to the vanity of either A or B, nor due to their envy, or vindictiveness or beneficence. Conflicts of desires with these causes, however, continue to exist. While Rawls might be right in thinking that such conflicts do not involve questions of justice, there still seems to be a role for J and K behind the veil to find principles that resolve conflicts of desire between A and B caused by envy, vanity, etc. Rawls' concern is that, unless the veil's application is restricted to conflicts of desire that fall within the mutual disinterestedness circumstance then the veil cannot produce a principle that yields a determinate result as to whose desire is to be satisfied. It is not necessary, however, for impartialism to impose the mutual disinterest condition on the parties behind the veil, even if Rawls has to. The same result can be derived from the impartialist axioms. This can be seen by an examination of the case of envy.

9. The 'Envy' Principle

Suppose there is a conflict of desires between A and B where A wants B not to do X because A envies B. A, then, is prescribing the principle that B not do X whenever A envies B's doing X. Symmetrified into a form that is not self-defeating this becomes, 'A is to be free to stop B's doing X because A envies B and B is to be free to stop A's doing X because B envies A.' J and K, however, will reject this principle because it fails to pass the non-coercionist axiom. If A wants not to be coerced from doing X more than he wants to coerce B into not doing X then he will not agree to this principle. Likewise for B. It appears, then, that J and K will not accept being envious of what a person is doing as a reason why coercion can be used to stop his doing it. A similar case could be mounted as to why it is not a reason to stop someone having something that one is envious of him.

It is a principle, then, that is deducible from the impartialist axioms that envy is not to be a reason for coercion. I will call this principle the 'envy' principle.

10. The 'Fallibility' Circumstance

Another circumstance of justice is human fallibility. Rawls puts it this way,

'I also suppose that men suffer from various shortcomings of knowledge, thought and judgement. Their knowledge is necessarily incomplete, their powers of reasoning, memory and attention are always limited, and their judgement is likely to be distorted by anxiety, bias, and a preoccupation with their own affairs. Some of these defects spring from moral faults, from selfishness and negligence, but to a large degree, they are simply part of man's natural situation. As a consequence individuals not only have different plans of life but there exists a diversity of philosophical and religious belief, and of political and social doctrines.' (1972, 127)

Even if it were the case that humans had complete knowledge, unlimited and unerring powers of reasoning, memory and attention, and undistorted judgement, it remains to be shown that there would not still be conflicts of desires between them. It may well be that there would be far fewer than there are now but inadequate knowledge etcetera may not be the cause of all conflicts of desires. J and K behind the veil of ignorance, then, may still have a role to perform in settling conflicts of desires between A and B, even when A and B are infallible.

11. Conclusion

Impartialism resolves conflicts of desire by having J and K argue as if they were behind a veil of ignorance with respect to which of A and B they are. Rawls does not see the veil of ignorance as contributing to the resolution of (1) conflicts of desires about how people should behave when they cannot influence each other (the Man Friday circumstance), (2) conflicts of desires about how humans should treat non-humans (the Humanhood circumstance), (3) conflicts of desire occurring in circumstances of superfluity or extreme scarcity (the Moderate Scarcity circumstance), (4) situations where there is already consensus (the Dissensus circumstance), (5) conflicts of desire due to envy, vanity, malevolence or benevolence (the Mutual Disinterest circumstance) and (6) conflicts of desire between infallible beings (the Fallibility Circumstance).

I have argued that, while the veil of ignorance when applied to each of these circumstances may not result in J and K producing just principles for governing the behaviour of A and B, it may still result in principles that either resolve a conflict between A and B or else, in the dissensus circumstance, determine whether there should be a consensus or not. By restricting the veil to the circumstances of justice (1), (2), (3), (5) and (6) Rawls neither eliminates the conflicts of desire involved nor resolves them. By restricting the veil to circumstances of justice (5) Rawls runs the risk of A exploiting B by getting B's agreement through some form of false consciousness, that is, by what I have called 'false sincere agreement'. As I have shown, however, there are cases of actual or hypothetical conflicts of desires between A and B that lie outside Rawls' circumstances of justice but where the veil of ignorance can be used to determine whether or not any resolution of the conflict is a forced, or an impartialist, one.

Rawls also wishes J and K to limit their reasoning so that it only results in principles that meet what he calls the 'formal constraints of the concept of right'. I will now turn to an examination of whether or not there is any good reason to restrict impartialism's use of the veil in this way.

Chapter 4

Rawls' 'Formal Constraints of the Concept of Right'1. Introduction

We have already seen that Rawls wishes to restrict the scope of the veil to those cases that fall within the circumstances of justice. Even if it is allowed that it is only for cases within the circumstances of justice that the veil comes up with just principles to resolve conflicts of desires, we saw that there is no reason to believe that the veil will not come up with principles for resolving conflicts outside the circumstance of justice. In Rawls' view, these principles do not result in just resolution of conflicts, not because they result in unjust resolutions instead but because, if Rawls is right, the notion of justice is not relevant to the cases they deal with. If what we are looking for is a comprehensive theory of right, then either it will have to cover conflicts of desire outside Rawls' circumstances of justice or else arguments will need to be adduced to show why these conflicts are not properly the subject of a theory of right. I will treat any conflict between people about any matter whatsoever as one that can be considered from an impartialist point of view, that is, as one where the resolution of the conflict is either the result of force or else is the result of genuine consent.

Not only does Rawls restrict the application of the veil to the circumstances of justice, he also wants to restrict the reasoning of the parties behind the veil in such a way that the principles they propose to resolve conflicts of desires must meet what Rawls calls the 'formal constraints on the concept of right'.

2. The 'Formal Constraints of the Concept of Right'

Rawls puts certain restrictions on the way in which the parties behind the veil can reason. He says,

'The situation of the persons in the original position reflects certain constraints. The alternatives open to them and their knowledge of their circumstances are limited in various ways. These restrictions I refer to as the constraints of the concept of right since they hold for the choice of all ethical principles and not only for those of justice. If the parties were to acknowledge principles for the other virtues as well, these constraints would also apply.' (1972, 130)

He explicitly rejects any attempt to justify his choice of these constraints by an appeal to the meaning of 'morality'. He says, 'The merit of any definition depends upon the soundness of the theory that results; by itself, a definition cannot settle any fundamental question' (1972, 130). This refusal to rest his theory on a clarification of concepts arouses Hare's ire. Hare says,

'There is in fact a vast hole in his 600-page book which should be occupied by a thorough account of the meanings of these words, which is the only thing that can establish the moral rules that govern moral argument. If we do not have such an account, we shall never be able to distinguish between what we have to avoid saying if we are not to contradict ourselves or commit other logical errors, and what we have to avoid saying if we are to agree with Rawls and his coterie.' (1975, 85)

This criticism would lack force if Rawls had stipulated what he meant by various terms, that is, a failure to define terms does make it harder to see where Rawls' conclusions might be a result of a shift in meaning. This, however, isn't really what Hare has in mind. He is concerned that Rawls doesn't take as a given the meanings that the 'moral' words have in ordinary language. Hare himself has a theory about what the various entailment relations are between moral terms in their ordinary language use. What Rawls is doing is saying that, even if Hare has these entailment relations right, these relations are either neutral between moral theories,

whereupon a detailed analysis of them is unnecessary except for clarity's sake or else that they are not neutral between theories, whereupon clarifying them only shows what one is committed to if one adopts the moral theory embodied in ordinary language. It doesn't show us why we should adopt the moral theory embodied in ordinary language. Where, perhaps, Rawls is open to criticism here is his implication that the meanings of the moral words he uses in his theory are contextually defined, that is, we only know their meaning by seeing the role they play in his theory as a whole. This kind of holistic account of the meaning of words does make it difficult to separate factual errors from logical errors, so to this extent Hare is right.

I will now turn to Rawls' 'formal constraints on the concept of right' to see if there are any good reasons why the principles that the parties behind the veil choose to resolve conflicts of desires have to conform to those constraints.

3. Generality

Rawls requires the parties behind the veil to adopt principles that are 'general', that is, 'it must be possible to formulate them without the use of what would be intuitively recognised as proper names or rigged definite descriptions'. (1972, 131)

What is puzzling here, is why Rawls should think this constraint needs to be placed on the reasoning of the parties behind the veil. If J and K are unable to tell which of the people in ordinary life A and B they are more likely to be, then it would seem that J and K would, as rational agents, refuse to agree to principles such as 'Do whatever A says'. Further, this rejection of such principles would be a consequence of their using the veil

rather than a restriction placed on its use prior to any reasoning by J and K. As a conclusion drawn by using the veil it would have whatever authority the veil itself has. As a condition placed on the use of the veil it cannot borrow the veil's credentials to give it a warrant, so it stands in need of some independent justification. There may well be two groups of people in ordinary life who disagree precisely over the issue of whether or not the principles that resolve conflicts of desires are only to be general (Group A) or can include proper names and rigged definite descriptions (Group B). What we have here is a conflict of desires over which kinds of principles are to settle other conflicts of desire. Rawls has decreed that the answer shall favour Group A. This however, may not be the position that J and K would adopt behind a veil of ignorance. If it is not, then it is a partialist position, that is, one that would only get forced rather than unforced agreement. Rawls, then, will have to argue for the position of Group A as if he were behind a veil of ignorance. This means that he cannot presuppose in this argument the unacceptability of principles involving proper names or rigged definite descriptions.

It might be argued in Rawls' defence that generality is a necessary condition of a principle's being a 'moral' principle. This may be so, in that one can either stipulate it to be so, or show that this is the ordinary language use of the word 'moral'. This, however, merely renames the groups A and B. Whereas A used to be the group that wanted only general principles to resolve conflicts of desire, it is now the group that wants only 'moral' principles to resolve conflicts of desire. Likewise Group B, which allowed principles containing proper names or rigged definite descriptions, is now the group that allows principles other than 'moral' ones to resolve conflicts of desire. Either way, no argument has been given as to why 'moral' principles are to be preferred to other kinds of principles. We

are left exactly where we were when we wanted reasons why general principles should be preferred to non-general ones.

Rawls states that generality rules out egoism. Stipulating generality as a formal constraint on the concept of right, then, has given him an easy victory over the egoist. It does not, however, throw any light on what's wrong with egoism. As I have shown, the three constituents of impartialist reasoning, namely, the position of equal power, the veil of ignorance and noncoercionism, also rule out the egoist. Egoism can be rejected as a conclusion of impartialist reasoning. This means that the egoist is someone who can only get his way in a conflict of desires by force, never by genuine consent. If we know this about an ethical position, then it gives us important guidance about how to conduct arguments with people who hold such a position. This guidance, however, is not available for ethical positions we have refused to subject to the scrutiny of impartialist reasoning. In fact, the acceptability from the impartialist viewpoint of the principles we reach are rendered dubious precisely because they have not had to compete with the principles we excluded. It may be that most of us, or even all of us, have an intuition that egoism is an unacceptable guide to action, that it is not the way to determine which side of a conflict of desires is to have its desires satisfied. One of the effects of the veil, however, is to reduce reliance on intuition. If our intuition coincides with what people would genuinely consent to, then egoism will be rejected behind the veil. It should not be excluded in advance of such consideration in the fear that it might not be rejected and the intuition that it should be rejected outweighs the appeal to the veil. To do so is not to take the veil seriously but to use it only when one knows in advance that it will conform to one's prejudices.

Rawls presumably places the generality constraint on the reasoning of the parties behind the veil because he thinks that the veil itself does not guarantee that J and K will only choose general principles. At first glance, this may seem an unnecessary fear in that J and K are unlikely to agree to the principle 'Do whatever A says' if neither of them knows the likelihood of his being A or B. First appearances, however, are deceptive. While the veil does mean that J and K will not agree to principles that use proper names or rigged definite descriptions of the conflictants, it does not mean that J and K will not agree to principles that use proper names or rigged definite descriptions of velorians. If J behind the veil has superior power to K it is possible for J to get K to agree to a principle that favours whoever J will be once the veil has lifted, even though neither J nor K knows who this is. J, for example, can use force to get K to agree to a principle such as, 'Do whatever you are told by the conflictant that J becomes' or, 'Everyone is to do X except the conflictant that J becomes'.

J and K will not agree to principles of this kind under impartialist constraints on their reasoning because J and K are in a position of equal power behind the veil. As a consequence, J does not have the power to get K to agree to non-general principles that favour whichever conflictant that J will turn out to be. Without this power, J will not be able to get K, or any other rational agent whose desires might conflict with K's, to agree to such non-general principles. Rawls, however, does not include any equal power constraint as one of the conditions placed on the reasoning of the parties behind the veil. This leaves J and K open to adopt non-general principles on the basis of their unequal power. As a result, he has to remove the possibility of their adopting non-general principles by fiat. The exclusion of non-general principles is not deducible from the basic premisses of his theory but is imposed on the theory. His justification for doing so, like his justification for all aspects of his theory

that are not deducible from the veil itself, is that this achieves the best reflective equilibrium between theory and intuition. This, however, is an unsatisfactory form of justification, as I will show later, if what we are out to do is to minimise the possibility of disagreement between rational agents about the principles on which we should act. The impartialist theory at least has the virtue that it stands or falls totally on what is deducible from its three constituent devices, the position of equal power, the veil of ignorance and noncoercionism. This, of course, is also a potential weakness in that, while it strengthens the theory's explanatory power, it increases the chances that it will clash with intuition, that is, if we pursue the scientific parallel, it increases its falsifiability. The appropriateness of comparing moral to scientific theories will also be considered when I discuss reflective equilibrium.

4. Universality

Rawls requires that the principles governing the behaviour of people in the real world be 'universal in application'. This places a restriction on the kind of reasoning that can be done by the parties behind the veil. They must reason only in those ways that lead to universal principles.

To be universal, principles need to meet a number of criteria.

- (1) 'They must hold for everyone in virtue of their being moral persons'. This means that everyone should be able to 'understand these principles and use them in their deliberations'. This means that there be 'an upper bound of sorts on how complex they can be, and on the kinds and number of distinctions they draw'. (1972, 132)
- (2) 'Principles are to be chosen in view of the consequences of everyone's complying with them'. This places at least two kinds of restriction on principles.

- (i) 'A principle is ruled out if it would be self-contradictory, or self-defeating, for everyone to act upon it'.
- (ii) 'Should a principle be reasonable to follow only when others conform to a different one, it is also inadmissible'. (1972, 132)

Rawls, then, places four constraints on the kinds of principles that the parties behind the veil can use to resolve conflicts of desire. I will call these (1) the 'moral persons' constraint, (2) the 'complexity' constraint, (3) the 'self-defeatingness' constraint, and (4) the 'exemption' constraint. Each can be stated as follows:

- (1) The 'moral persons' constraint, that is, a principle must hold for any being which is a moral person merely because that being is a moral person.
- (2) The 'complexity' constraint, that is, the principles must be such that all moral persons can understand these principles and use them in their deliberations. This means that principles must not be too complex or draw too many kinds and numbers of distinctions.
- (3) The 'self-defeatingness' constraint, that is, it should not be self-contradictory or self-defeating if everyone acted on it.
- (4) The 'exemption' constraint, that is, it should not be the case that the principle is reasonable for a particular person to follow only if everyone else follows some other principle.

The 'moral persons' constraint can be seen as an extension of the generality constraint. The generality constraint prevents any particular velorian (J or K) from tailoring principles about conflictants (A or B) to further the desires of whichever conflictant he turns out to be. The 'moral persons' constraint prevents any group of velorians from tailoring

principles about conflictants to further their own partialist desires. The possibility of the parties behind the veil furthering their own partialist desires behind the veil exists as long as the parties behind the veil are not treated as being in a position of equal power. Just because they have to reason as if they don't know what their sex is in ordinary life, or their colour, or wealth, or religion, or whatever, it doesn't follow that their choice of principles about conflictants will not be determined by what they know about themselves as velorians. If they know what sex they are as velorians, or what race etc., then, even with the generality constraint, they can tailor principles to favour certain groups of conflictants. Which group they will favour depends on whether, as velorians, they are egoists or coercionists. If they are egoists, they could propose principles such as 'Everybody is to do X except conflictants whose velorians were male, or white' or 'Do whatever you are told by a conflictant whose velorian was white'. There is no proper name or rigged definite description in such a principle. Any particular egoist would be less happy with such a principle than with one that required X of everyone except him, but at least he is one of the group of whom X is not required (assuming X to be something he would wish to avoid if he could). If the velorian is a coercionist, then he will not care what happens to the conflictant he turns out to be as long as the ideal that he has behind the veil is furthered in ordinary life.

Unlike the generality constraint, however, the 'moral person' constraint does not allow him to put forward principles that 'All conflictants are to act so as to further my ideal as a velorian'. Rather, he can put forward principles such as 'Conflictants are to further only those ideals that are ideals held by whites, or males' that is, he increases the chances that his ideals as a velorian will be followed by conflictants but he can't guarantee it as he can with the generality constraint if he has superior power over other velorians. He will, therefore, be prepared to propose principles

such as 'Everyone is to do whatever is required to achieve the ideals of white velorians'. Neither the egoist or the coercionist, however, will be able to get all the other velorians to agree to this principle if all proposals are construed behind the veil as two-party disputes in a position of equal power. Rational agents who do have different desires to these egoists or coercionists as velorians, or who might as conflictants, will only agree to those non-universal principles if they were coerced into doing so in a position of unequal power.

Impartialism, then, does not need the 'universality' constraint to guarantee that the parties behind the veil will not agree to egoistic or coercionist non-universal principles. The 'position of equal power' and the 'non-coercionist' requirements achieve the same effect.

Katzner puts a somewhat different interpretation on the 'moral persons' constraint.

'If there is a right to life it extends to all moral entities (all beings whose life or death makes a difference), not merely to all men, all women, all Americans, or any other such class. Similarly, if lying is wrong, it is wrong for everyone capable of telling the truth.' (1980, 49)

Katzner's account could commit Rawls to the view that all principles applying to conflictants must contain no exception clauses, that is, the velorians are not to formulate principles of the form 'Everyone (conflictants) has a right to life, except convicted murderers' or 'Everyone (conflictants) shall tell the truth, except philosopher kings'.

Presumably neither Rawls, nor Katzner, would mean the universality constraint to be such a strong requirement, as both the principles given have been understood by moral philosophers as legitimate candidates for moral

principles. All moral philosophers, I would think, would expect a moral system to contain principles with exceptions built into them, eg, 'Everyone must work for a living except the disabled' and so on.

It is more plausible to interpret Rawls as I have suggested. This means that his universality constraint is a constraint on how the velorians are to understand the principles they adopt. They are to see the principles as applying to all of them, that is, they cannot rig up a principle so that it makes reference to exempting them in their velorian form, even if it allows exemption among conflictants.

As can be seen from this discussion, Rawls' theory runs the danger of generating a kind of infinite regress of veils. The velorians cannot be allowed to use their knowledge of who they are as velorians to construct principles with exception clauses built into them that favour them individually as conflictants. This could be (temporarily) overcome by positing a pre-original position, that is, we have 'people in the pre-original position', who don't know anything about themselves in the original position, but this merely poses the same problem at a later remove. As we could always suspect at the pre-original position level that the people there had rigged the principles for the velorians on the basis of knowledge about themselves as people of the pre-original level, we could never feel secure about the impartiality of people in any position.

It is clear enough, then, why Rawls ought to rule out the velorians adopting non-universal principles, if a non-universal principle is one that favours certain conflictants on the basis of the features they possess as velorians. He is simply trying to avoid the same problem for velorians knowing which group they belong to as velorians as he was with velorians knowing which individual conflictant they were, i.e., which proper names

or rigged definite descriptions fitted them. Some solution to this kind of problem needs to be found before Rawls' theory can even get off the ground. He has chosen to limit by fiat the range of principles the velorians can use, relying on our accepting this limit as a reasonable one for a theory of justice. As I indicated earlier, it may be reasonable if the kind of theory we are after is a theory of 'justice', but its striking us all as reasonable does depend on our sharing a common concept of justice. If we don't, Rawls will need to show why it is reasonable for a theory to only count as a theory of justice if it is universal in his sense, i.e., it contains no non-universal principles that favour certain conflictants on the basis of the features they possess as velorians. Even if it does gel with the linguistic intuitions we all have about the meaning of the word 'justice', he still has to show why 'just' and, therefore, necessarily universal theories are to be preferred as guides to action rather than theories like egoism, which his constraints have been selected to rule out. A theory of right would have to show why egoism should be ruled out. If I can show why partialism should be ruled out then this will have shown why egoism should be as egoism is a form of partialism. It will not be possible, however, to use the veil to show why partialism should be ruled out because the veil is just a method for deciding what actions are permitted, required or forbidden if partialism is ruled out. There will need to be independent arguments in favour of impartialism and against partialism. My use of the veil, however, does not have this problem of an infinite regress of veils, because it requires the parties behind the veil to reason as if they were in a position of equal power and as if they were noncoercionists. However, just as the veil cannot be used to show why partialism should be ruled out, neither can the position of equal power or the requirement of noncoercionism because these are merely devices for ruling partialism out. There will also need to be independent arguments showing why we should reason as if in a position of equal power and as if

we were noncoercionists, as well as arguments showing why we should reason as if we were behind the veil. Essentially, those arguments are those given earlier to the effect that only such reasoning avoids coercion and embodies a person's genuine consent. I will take up later the question of why we should adopt principles based on genuine consent rather than force.

The 'formal constraints of the concept of right' are constraints that Rawls places on the kinds of principles that the velorians can use to resolve conflicts of desires between conflictants. The aim of the velorians is to come to agreement on the principles that govern certain kinds of cases in ordinary life. There is no guarantee that those principles will not be very complex, or that they will not draw very many kinds and numbers of distinctions. If the principles are very complex, etc., then it may be very difficult for conflictants to act in accord with them. Principles that are difficult to understand and which, therefore, are likely to not be followed properly, are not very satisfactory as guides to action, so Rawls suggests that the velorians be constrained to come up with relatively uncomplicated principles. This, however, is not the only solution, although it is the one that fits in with Rawls' rule-contractarianism. Another solution is that the velorians develop simple Level 1 rules of thumb to govern the standard cases, that is, they adopt principles about the kinds of dispositions to action that conflictants should possess. When a case is contentious due to its complexity, then the velorians need to engage in more complex reasoning, that is, reasoning as if we were velorians we need to engage in more complex reasoning. Rawls would prefer that we merely applied the rules of some institution or social practice such as promising which our reasoning as velorians has shown to be a better practice than any other. This, however, means that he needs to establish that his rule-contractarianism is superior to act-contractarianism (and, as I argued earlier, this he hasn't done). Farrell (1980) shows that Rawls has NOT established this.

I will now turn to the 'self-defeatingness' constraint. The self-defeatingness constraint is only applicable if principles can be divided into two kinds, namely, those that have self-defeating consequences if everyone follows them and those that don't. What, however, is it for everyone to follow a principle? Suppose we have a principle such as 'Break a promise whenever doing so is profitable' which I will call the principle 'E'. This has the form, 'Do an action of class X under circumstances C'. To whom is it addressed? Presumably to anyone, or everyone, in circumstances C. The form of the principle then, can be expressed as 'For all js (individuals), if j is in C, then j is to do X', that is, for all js, if j is in a situation where breaking a promise is profitable, then he is to break the promise.

This particular principle is, presumably, one of those that Rawls, along with Kant, regards as self-defeating if everyone follows it. It is self-defeating because, in Kant's words, 'it would make promising, and the very purpose of promising itself impossible, since no one would believe he was being promised anything, but would laugh at utterances of this kind as empty shams'. (1948, 90)

Assuming Kant's argument to be correct, then the consequences of everyone following this principle are self-defeating, so Rawls would not let the parties behind the veil consider such a principle. If only some people, however, up to a threshold figure, followed this principle, its consequences might not be self-defeating. Why should Rawls want to prevent the parties behind the veil taking this into account?

Katzner suggests that doing so thereby enables Rawls to ensure that the parties behind the veil do not adopt any act utilitarian principles. He says,

'It should be clear that although all moral principles may well conform to the first or weak ("moral persons") interpretation of universality, some do not conform to the latter (stronger) interpretation (the "self-defeatingness" and "exemption" constraints'... The stronger interpretation ... by elevating the hypothetical consequences of everyone's doing something over the actual consequences of some people doing it precludes basing morality on what will actually happen. Although one might argue that a theory based upon hypothetical consequences is more reasonable than one based upon actual consequences, this cannot be a formal constraint of the concept of right. For, if it is, some of the chief competitors to Rawls' theory of justice are ruled out of the moral domain. In particular, act utilitarianism becomes a non-moral theory.' (1980, 50)

If Katzner is right about the effects of the self-defeatingness constraint, then it is a constraint that prevents the parties behind the veil resolving a particular conflict of desires, namely, the conflict of desires between Group A who wants to resolve other conflicts of desire by taking into account the actual consequences of doing X rather than Y and Group B who wants to take into account only the consequences of doing X rather than Y under hypothetical conditions. Rawls is wanting to commit J and K behind the veil to siding with Group B. If he does so, then the effect of this is to make any conclusions J and K reach behind the veil unacceptable to members of Group A. This kind of strategy, then, fails to get the agreement of all rational persons. Rawls has to come up with reasons convincing to Group A why their position is excluded. He cannot appeal to the impartiality of the veil because he has ensured by fiat that the veil is not impartial between Group A and Group B. There is, however, no impartialist reason for excluding Group A's position as one of the positions the parties behind the veil might consider. It may be that, upon consideration, Group A's position is rejected by J's and K's use of the impartialist devices and the actual consequences of everyone's doing X rather than Y are not taken into account in resolving conflicts of desire over X and Y. If this happens, however, it is by the application of the veil, not because of a constraint on the veil. As a consequence, the rejection of A's

position would show us that it was a partialist position, whereas Rawls' rejection of it shows us merely that he wants a non-utilitarian theory.

Rawls, however, may have had something else in mind. He could be out to prevent the parties behind the veil taking into account the consequences of only some people, rather than everyone, acting as the principle requires in specified circumstances. Why would a rational mutually disinterested agent be concerned with the consequences of only some people acting in accord with a principle rather than with everyone's doing so? It would only be of interest to him if he thought this would lead to those actions being performed which achieved his desires rather than someone else's in a conflict of desires.

Suppose the velorians are considering whether or not to endorse E. Now, in endorsing E, they are endorsing it for everyone in circumstances C. Suppose, however, they know that most conflictants won't do X in C, e.g., most conflictants will keep their promises even if breaking the promise would be profitable. If a particular velorian, say J, knew that he was a conflictant who would follow E, then it would be in his interests to advance E as a principle to be endorsed by the parties behind the veil. If he did not know which conflictant he was he may still be prepared to advance E but, knowing he was such a conflictant, he would certainly advance it. This is not a case, then, or has not yet been shown to be a case, where J would advance E only if he knew which conflictant he was. Nonetheless, the fact that he would definitely advance E if he knew which conflictant he was might make us suspicious that he would only advance it if he knew. After all, he is prepared to advance it as long as he knows which conflictant he was without any further consideration of whether or not he would still be prepared to advance if he didn't know which conflictant he was. If it hasn't been shown yet that people would only endorse

E if they didn't know whose goals E furthered, then it hasn't been shown that E is a principle that impartialists wouldn't endorse. However, if a person J is prepared to endorse E even if he doesn't know whether E is an impartialist principle, then he shows himself to be a partialist. This is what J does when he is prepared to endorse E just because he knows that only some conflictants follow E, that the goals that will be furthered by E in an interpersonal goal conflict between the conflictants that follow it and those that don't will be the goals of those that do, and who knows that he is one of the conflictants who follow E. Such a person cannot be an impartialist, because he is prepared to endorse E even though he doesn't know whether or not the actions that E permits, requires or forbids are those that further only those goals that people would agree to allow each other to pursue even if they didn't know whose goals they were.

While such a person shows himself to be a partialist, it doesn't follow that impartial people wouldn't endorse E. This is because the partialist who endorses E even if only some people follow it, only does so because he knows that he is one of the conflictants whose goals are furthered by E if most other conflictants don't follow E. Once he is placed behind the veil, he may no longer endorse E but, then, it could turn out that he may. It remains an open question, then, at this stage, as to whether all principles that are self-defeating if all conflictants follow them but advantage one group rather than another if some conflictants follow them are partial. It remains, then, an open question as to whether or not act utilitarian principles are partial ones.

The 'exemption' constraint is the reverse of the 'self-defeatingness' constraint in that it rules out considering principles in terms of the consequences of only some people following them if you are not one of those who follow it, whereas the self-defeatingness constraint is concerned

with the consequences if you do follow the principle. The same points as were made for the 'self-defeatingness' constraint apply for the exemption constraint. It rules out taking actual consequences of adopting the principle and allows only the hypothetical consequences of everyone following it, even if some people actually don't. This means it rules out act utilitarianism for the reasons Katzner gives. It also means that a person is a partialist if he would endorse a principle just because he knows that most people will follow it, that the goals it furthers in a conflict of desires between conflictants who follow it and those who don't will be the desires of those who don't and who knows that he is one of those who don't, regardless of whether he would still endorse it if he didn't know whose desires it would further. The principle itself, however, may still be an impartialist one but whether any, or all such principles are partialist or not is, at this point, an open question.

All of the universality constraints, then, appear to perform a similar role. They try to prevent a velorian adopting a principle either because it furthers the desires of those conflictants who were members of the same group of velorians or else because it furthers the ideal of some group of velorians of which the proposer is a member, regardless of who he turns out to be as a conflictant. It is a group version of the generality constraint. All of these constraints, however, are not to be imposed on the thinking of the velorians but, rather, are to be conclusions of their reasoning behind the veil when constrained by the position of equal power and the requirements of noncoercion. Under these conditions, they will reason as if the 'moral persons' constraint applied, that is, they would not resolve conflicts of desires with principles that favour one group of velorians over another. The 'complexity', 'self-defeatingness' and 'exemption' constraints however, have not been shown to be constraints that the velorians would agree should apply to the principles that resolve conflicts of desires.

5. The 'Publicity' Constraint

Rawls requires that the parties behind the veil choose between principles on the assumption that the principle they chose will be known by people in the real world to be the principle on which they act. He says,

'They suppose that everyone will know about these principles all that he would know if their acceptance were the result of an agreement. Thus the general awareness of their universal acceptance should have desirable effects and support the stability of social co-operation. The difference between this condition and that of universality is that the latter leads one to assess principles on their being intelligently and regularly followed by everyone. But it is possible that all should understand and follow a principle yet this fact not be widely known or explicitly recognised.'
(1972, 133)

The 'publicity' condition, however, has the same problems as the generality and universality conditions. It excludes by fiat a range of positions actually held in the real world, rather than demonstrating that these positions would be rejected by the parties behind the veil. There is nothing in the requirement that we reason as if we didn't know whether we were A or B that imposes the publicity condition on our reasoning, although once we begin reasoning as if behind a veil of ignorance we may conclude that the only principles that should govern the actions of people in the real world are principles about which everyone in the real world is fully informed. Katzner, for example, claims that the presence of the publicity condition is what distinguishes contractarian from non-contractarian theories. Rawls makes remarks that support this view. He says, 'It is characteristic of contract theories to stress the public nature of political principles'. (1972, 16)

Katzner concludes from this that the publicity condition as a formal constraint of the concept of right arbitrarily excludes all non-contractarian theories of right. He says,

'The reason why it is illegitimate for Rawls to claim that publicity is a constraint for all moral principles should be obvious. Not all moral theories are contractarian. Indeed, the theory which Rawls acknowledges to be the most likely alternative to his i.e. utilitarianism, is not normally conceived of in contractarian terms. Thus Rawls must either insist that most of the proponents of utilitarianism are not presenting moral theories or else acknowledge that publicity is not a constraint which applies to all moral principles.' (1980, 52)

While Katzner's criticisms are correct, a few comments need to be made to put them into context for our discussions concerning impartialism. Rawls may wish to insist that the proponents of utilitarianism are not presenting moral theories. If he does, this is merely because he has given the word 'moral' a stipulative definition that excludes theories that don't conform to the publicity condition. He can use the word 'moral' this way if he wishes, but, as I stated earlier, this kind of philosophical legerdemain merely leaves us with the problem of why we should prefer moral theories (ones that meet the publicity condition etc.) about what we should do to non-moral ones (e.g. utilitarianism)? In the real world, there may be group A that supports only those principles that meet the publicity condition, and group B that does not. All that the veil requires of J is that he reasons as if he doesn't know whether he is a member of A or B. If the publicity condition is included as a constraint on how J reasons, then he reasons as if he knows that he is a member of A, which contravenes this requirement.

Nonetheless, it may be the case that, even without the publicity constraint, any person J behind the veil will conclude that people in the real world should only act in accord with those principles about which everyone in the real world is, or could be, fully informed. If J does reach this conclusion, it should only be as a consequence of the constraints imposed on him by the veil itself.

Haksar makes the point well. After noting how Rawls believes the publicity constraint to rule out Plato's Noble Lie in the Republic, Haksar says,

'But why must we accept the publicity condition as a universal constraint on the model? Rawls points out that the publicity condition is natural. But if it is so natural it should be possible to derive it from the rest of the contractarian model rather than have it as a separate and universal constraint on the model.'
(1979, 169-170)

A rather intriguing consequence of defining contract theories so that they conform to the publicity constraint in advance of the outcome of the discussion behind the veil is that my impartialist theory fails to be a contract theory. It is a contract theory, however, if contract theories include those theories that conclude that the publicity constraint should apply to the principles to be followed in the real world.

The force of this criticism of the publicity constraint can be seen from an example given by Haksar. The 'egalitarianism' he refers to is meant to cover Rawls' contractarianism. He says,

'Anti-utilitarians sometimes criticize utilitarianism on the grounds that it may involve a kind of deception. Utilitarianism at the deepest level may commit one to producing a race of non-utilitarians. For if people are brought up as utilitarians this may (as a matter of fact) make them do actions that lead to bad results from a utilitarian point of view. But now there is a similar paradox in the case of egalitarianism. It may be that even if egalitarianism is true, preaching it and implementing it under certain social conditions may create so much havoc that it harms the egalitarian cause; it may raise people's expectations far beyond what the state can satisfy, and lead to misery, envy and violence. In a hierarchical society where the lower castes have not just less liberty but also are acknowledged to have less worth than the top ones it is quite possible that, even from the point of view of self-respect, the worst off are better off under such a society than they would be in a society where egalitarianism was publicly acknowledged and where there was much dissatisfaction. Now under such conditions a maximin egalitarian (though not a strict egalitarian) may well find the caste society the lesser of the two evils.'

Of course many moralists, such as Kant and Rawls, would think that any moral doctrine that is true at the deepest level, but cannot safely be turned into a public moral code, is an immoral one. Now such thinkers disvalue deception very strongly. Rawls follows Kant in stressing what he calls the publicity condition: the true political principles, according to Rawls, should be the basis of public charter. But now Rawls is also an egalitarian at the deepest level; the conditions of his original position are designed to ensure this equality. It is not obvious to me why the publicity condition must be accepted by egalitarians. Rawls uses the publicity constraint to construct his model, for 'the parties assume that they are choosing principles for a public conception of justice. They suppose that everyone will know about these principles all that he would know if their acceptance were the result of an agreement'; this condition automatically (that is to say, without considering the relevant merits and demerits) rules out 'such devices as Plato's Noble Lie in the Republic'. But why must we accept the publicity condition as a universal constraint on the model? Rawls points out that the publicity condition is natural. But if it is so natural it should be possible to derive it from the rest of the contractarian model rather than have it as a separate and universal constraint on the model. The truth is that it is no more difficult and unnatural to construct an egalitarian model that has the other constraints (such as the veil of ignorance) but not the publicity constraint.' (1979, 168-170)

Haksar's criticisms of the publicity constraint occur as part of his attack on the unargued for value that Rawls gives to autonomy as a feature of real world people that must be preserved. In the real world, however, there may be group A that supports only those principles that preserve the autonomy of people in the real world and group B that supports principles that don't. All that the veil requires of a person J is that he reasons as if he doesn't know whether he is a member of A or B. If he reasons as if there is an autonomy constraint on the principles that the people in the real world must follow, then he reasons as if he knows that he is a member of A, which contravenes the impartiality of the veil. This means that the kind of conclusion that Haksar sketches as a possible outcome of impartialist reasoning is not excluded. In my view it needs to be excluded, but it cannot be excluded by fiat. It will be a major task to be taken up later to show that a respect for autonomy can be deduced

from the impartialist conditions without any ad hoc constraint on the reasoning of the parties behind the veil.

As we have noted, making publicity a requirement of a theory's being a moral theory has the effect of ruling out all non-contractarian theories as candidates for moral theories. This, presumably, is not Rawls' intention, that is, while he wants to show that contractarianism is superior as a moral theory to non-contractarian theories, he does not want to do this by the linguistic fiat of excluding non-contractarian theories from the domain of the moral. What he does want to do, however, is rule out the situation where the velorians agree that a principle such as the principle of utility is to be used to resolve conflicts of desires while all or most conflictants do not know that it is, following instead some lower level principles such as 'Keep your promises' rather than the real principle 'Keep your promises if this maximises utility in this case', (or whatever). Without the publicity constraint, the velorians would, it seems, be able to adopt utilitarianism of the kind just described where all, or most conflictants think they are acting on the basis of one kind of principle when the principle they are unwittingly following is quite different.

A moral theory which does not meet the publicity constraint does raise interesting problems from a contractarian point of view. It would be less question-begging, however, if these questions could be answered by applying the contractarian machinery rather than by an ad hoc restriction of its application. What Rawls effectively does is to say that the conflict of desires between those conflictants who want to resolve other conflicts of desires by use of public principles (Group A) as opposed to those who allow some non-public principles (Group B) is not itself to be a conflict of desires resolved by the veil. It is to be resolved in favour of Group A

prior to the use of the veil. While this gives Rawls the answer he wants it does so at the cost of his answer being unconvincing to members of Group B.

What, then, are the questions raised by a theory that doesn't meet the publicity condition that a truly impartial application of the veil would have to answer? Firstly, such a theory raises problems about the relationship between the velorians and the conflictants. The velorians are only an imaginary device. If only they know what the real principle is, then no-one actually knows it. The only real people are the conflictants, the people in ordinary life. This means that the velorians have to reason as if at least some conflictants know the real principle, that is, the non-public principle. Moreover, these must be the conflictants in power. They must be able to organise society so that all other conflictants are trained to obey the rules of institutions like promising but are not trained to think about the justification for those rules. We end up, then, with Platonic philosopher-kings, which raises all the well-known questions about what qualities such people need to possess, and how their corrupt use of power can be prevented, and so on.

Does the veil rule out conflictant philosopher-kings or are they an option by which conflictants can resolve conflicts of desires? Any actual moral argument is between conflictants, that is, any argument about the principles by which to resolve conflicts of desires is between conflictants.

Impartialism requires, as does Rawls, that conflictants argue as if they were velorians. Suppose a group A of conflictants has decided that velorians would accept philosopher-kings and that philosopher-kings would require that all conflictants except members of A should follow Principle 2 whereas members of A should follow Principle 1 from which Principle 2 is derived. Suppose that there is a conflictant named B who wants to know

why he should follow Principle 2. There is now a conflict of desires between B, who doesn't want to follow P2, and A who wants him to. In fact, A probably thinks that B should be coerced to. Nonetheless, as impartialists, A and B have to argue as if they were in a position of equal power behind a veil of ignorance with a noncoercionist attitude to coercion. This means that they have to argue as if they were J and K, neither of whom know which of A or B he was. This doesn't mean that they don't know which of A and B they are, but they have to agree to adopt Principle 1 or to refuse to adopt Principle 2 as if they didn't know.

Suppose J is one of the conflictants in A and K is the conflictant named B. J can't agree to Principle 1 as if he knows he was A. In a position of equal power, whatever one party knows so does the other. J, therefore, either has to present no arguments that rely on his knowing that Principle 1 is the justification of Principle 2 or else has to inform K of Principle 1. If he chooses the former, then he is unlikely to get K to agree to Principle 2, which means that B doesn't need to follow Principle 2 in ordinary life. If he chooses the latter, then K knows that Principle 1 is the justification for Principle 2 as well as for other lower level principles, which means that B, who was not a member of A, knows it, which means that Principle 1 will be contravened everytime any conflictant not in A asks for the justification of principles derived from Principle 1. Impartialist restrictions on argument, then, force the holder of non-public principles to make them public to anyone who challenges the conclusions derived from them. The only way this can be prevented in ordinary life is to guarantee that members of B never challenge principles derived from Principle 1. This could only be done by removing their autonomy as rational agents. As we shall see, however, the impartialist argumentation leads to the conclusion that this would not be agreed to behind the veil.

Once B challenges Principle 2, and A offers Principle 1 as justification, it is also possible that B challenges Principle 1, that is, that B doubts that the parties behind the veil would agree to the principle that everyone except members of A should follow Principle 2. The members of A, then, would have to argue as if they didn't know they were members of A as to why the parties behind the veil would agree to Principle 1. They cannot use their own belief as conflictants that the parties behind the veil would suppress conflictants who challenged Principle 2 to use coercion to suppress B, because B may be challenging, or may go on to challenge whether the parties behind the veil would agree to anything of the kind. They could only act to suppress B if they were arguing as if they knew they were members of A in the conflict between A and B. This shows them to be acting as partialists, that is, as people committed to the resolution of the conflict in their own favour by the use of sanctions.

6. Ordering

Rawls requires that a moral theory be 'able to order all the claims that can arise (or that are likely to in practice)'. (1972, 134)

Katzner criticises this requirement as follows:

'What is to be done when justice requires one thing and utility another? According to Rawls, these principles must be ordered. The theory must tell us which one takes precedence when the two conflict. And this is a constraint of the concept of right. In other words, any theory which does not do this is not a moral theory.

Surely this is too strong. It requires us to say that all those theories which Rawls labels 'intuitionist' - i.e., all theories which rely on the use of intuition to resolve conflicts between principles - are not moral theories. To insist that the reliance on intuition is a defect in any moral theory in which it appears is one thing. But to say that they are not moral theories is something else. It involves much too narrow a conception of what is moral and what is not.' (1980, 52)

While Katzner's criticism is well-taken, it is not the one of central concern to those interested in the impartiality of Rawls' candidates for formal constraints of the concept of right. As I indicated earlier, for any definition of what it is to be moral, there can be disagreements in the real world as to whether or not one ought to do what morality requires rather than what it forbids, in whichever sense of 'morality' has been offered. In the real world, therefore, there may be group A that supports only those principles that are required by 'morality' in the sense some speaker has offered, and group B that supports principles regardless of whether 'morality' as defined in this conflict requires them. Any party behind the veil J is required to reason as if he doesn't know whether he is a member of A or B. If he reasons as if there is a constraint on him to reach conclusions consistent with the principles of group A, then he is arguing as if he knows he is a member of A rather than B. This contravenes the impartiality requirement. In particular, if group A believes that principles must be ordered, and group B does not, then J is not to argue as if he already knew he was a member of A. The conditions imposed by the veil must be sufficient to lead him to require ordering or else ordering is not required.

Rawls requires that the ordering between principles be transitive, that is, 'if, say, a first arrangement of the basic structure is ranked more just than a second, and the second more than a third, then the first should be more just than the third'. (1972, 134) This eliminates the use of 'intuition', which cannot guarantee transitivity, therefore, as we have seen above, transitivity cannot be a prior condition on the reasoning of the parties behind the veil, unless it can be shown to be required by whatever is entailed by the parties being rational agents able to discuss what acts should be done.

Rawls adds a further requirement to ordering, that is, not any old way of arriving at an ordering is good enough. It is a formal constraint of the concept of right that an ordering cannot be determined by

'the appeal to force and cunning ... Thus I assume that to each according to his threat advantage is not a conception of justice. It fails to establish an ordering in the required sense, an ordering based on certain relevant aspects of persons and their situation which are independent from their social position, or their capacity to intimidate and coerce.'
(1972, 134)

Now, it is precisely this latter that the veil is meant to eliminate, that is, it is hoped that J's not knowing whether he is A with superior power, or B with the inferior power is sufficient to ensure that the decisions J or K or any one else behind the veil are not coercive in an unfair, unjust or partial manner. Nonetheless, coercion may still be required, as we shall see, but it is coercion that the parties behind the veil agree should be applied in the real world but which they agree to behind the veil in a situation of total non-coercion. The sole basis for agreement behind the veil is the rational calculation of individual desire-maximisation where there is no point in J's threatening K to favour A rather than B because J doesn't know that A's desire will be his desire because he may be B with some incompatible desire.

While it is hoped that the veil condition will result in the parties agreeing to use methods in the real world other than physical conflict and the resort to arms, this cannot be a prior requirement on the argument behind the veil. After all, there may be people in the real world who are members of some group A that believe that might is the only right. If J argues behind the veil on the assumption that members of A are wrong, then he argues as if he knows he is not a member of A, which contravenes the impartiality requirement.

As we have seen, the three impartialist axioms together rule out 'might to right', but the veil by itself doesn't rule it out. When I come to show why impartialism should be adopted as a theory of right one of the positions I have to show should be rejected is the view that might is right.

It is possible that the velorians conclude that there are two or more principles that are equally important determinants of actions. This means that two groups of conflictants, Group A and Group B, could want to do incompatible actions X and Y, where X is required by one principle and Y by the other. If this situation arose, the velorians would have no way of deciding with action should be done. Rawls says, 'It is clearly desirable that a conception of justice be complete, that is, able to order all the claims that can arise (or that are likely to in practice)'. (1972, 134) Rawls, therefore, requires that the velorians not leave it the case that the principles are equally important. They must rank principles. Furthermore, they cannot rank principles by some arbitrary means, e.g., flip a coin as to which one to give priority to. At least, they cannot do so if this does not produce a transitive relation. If there are three principles P, Q and R which the velorians decide are equally important then, if flipping a coin ranks P higher than Q, and Q higher than R, then P must also rank higher than R.

Clearly, the conflictants can't be left in a situation where the rules they follow permit them to do either X or Y where these are incompatible actions. This puts the conflictants in a situation of deciding which shall be done by force. It is this that the veil is supposed to avoid. However, just because the rules that the conflictants follow must be ordered in a complete and transitive manner it does not follow in Rawls' system that the velorians arrive at principles that are ordered in a complete and transitive manner,

as long as the velorians decide which of X or Y is to be done. They could do it on intuition, or by flipping a coin, or whatever, and it could be that the relationships between their decisions are not transitive. Of course, this couldn't occur in too many cases otherwise the conflictants would never know what to do (this harks back to the complexity component of the universality constraint).

If, however, the conflictants are the velorians, as is the case in ordinary life, the fact that P and Q are equally important does lead to the situation where even if the conflictants think as velorians, there is no solution as to which of X or Y is to be done. This is not solved by the conflictants relying on their intuitions because, even if they try to think as velorians, there is nothing available to velorians, whether in Rawls' model or mine, that gives the velorians any guidance as to which intuitions are right and which wrong. This means that conflictants of Group A will have velorian intuitions favouring X, and there is no reason why they shouldn't, whereas conflictants of Group B will have velorian intuitions favouring Y, and there is no reason why they shouldn't. They are still no closer to a solution as to which of X or Y should be done. If Group A has superior power, it could resolve the situation by forcing Group B to follow A's preference but then A's members are not behaving as if they didn't know whose desires were furthered by X's being done rather than Y. If they really didn't know, then they might have opted to force Y rather than X, that is, if they really didn't know, they would have chosen as a random procedure such as coin-flipping would have chosen. Their bona fides in this direction is best demonstrated by their actually choosing such a random procedure. This eliminates intuition as an acceptable procedure.

7. Finality

Rawls puts forward as a formal constraint of the concept of right that 'the parties are to assess the system of principles as the final court of appeal in practical reasoning. There are no higher standards to which arguments in support of claims can be addressed, reasoning successfully from these principles is conclusive'. (1972, 135)

This is meant to be a constraint on the parties in the original position, that is, behind the veil of ignorance. Now, in one respect the constraint of finality is redundant, at least from the viewpoint of impartiality. There may be people in the real world who do not regard the decisions of the parties behind the veil as final. They constitute Group A, as opposed to Group B, who do treat these decisions as final. A person's impartiality in real-world decision-making, however, is a function of the degree to which he acts in accord with the principles he would choose behind a veil of ignorance, consequently what the parties behind the veil would do is final from the viewpoint of impartiality. There is no higher court of appeal, provided the veil conditions have correctly operationalised the concept of impartiality. If what we are trying to decide is whether or not an act or institution is the one that impartial people would prefer, then the decisions of the parties behind the veil is final.

What is of interest is whether the finality of these decisions is something the parties behind the veil accept prior to arriving at them, or whether it is one of the conclusions they reach behind the veil? It is, presumably, an assumption they have prior to discussion. This does not render its impartiality suspect, however, because, even if the parties behind the veil were to come to the (self-referentially) paradoxical conclusion that people in the real world shouldn't act in accord with the principles agreed to by

the parties behind the veil, it would be their coming to this conclusion that rendered it right for people not to follow the conclusions reached by the parties behind the veil. As it is, of course, people cannot (logically) follow this advice because, if they followed it, they wouldn't be following it, and if they didn't follow it, then they would be following it. This means that the parties behind the veil, if they are rational agents, cannot offer this advice as a serious guide to actions. It needs to be an assumption, therefore, of the parties behind the veil that their advice is final, that is, they argue as if they knew that their advice is the ultimate court of appeal. It is not a question they are free to discuss, in that their function would be rendered futile were they to conclude their decisions weren't overriding. Were, for example, they to conclude that their advice isn't final, this conclusion would itself be final from the viewpoint of impartiality, thereby generating the same kind of paradox previously discussed.

Rawls' inclusion of finality as a formal constraint of the concept of right could be interpreted as the claim that a system of principles is not a moral one unless it constitutes the final court of appeal. Now, there are people in the real world who believe that the requirements of liberty are sometimes incompatible with those of happiness, for example, but that there is no system of principles that properly captures the proper demands of both values. If happiness is made overriding, as in utilitarianism it leads to unacceptable infringements of liberty. If liberty is made overriding, as in a theory such as Nozick's, then some people's happiness is unacceptably frustrated, or the proper demands of a value such as justice are ignored. They believe that a balance between these competing values can only be achieved by intuition (Brian Barry holds a view of this kind), that is, they do not believe that we act on a system of principles that is final or overriding but rather that we test the conclusions of various systems of

principles against our intuitions, sometimes doing as one system requires, sometimes as another does, but letting no single system be final. One can refuse to call such an intuitionist process a 'moral' one but this merely means, as we have seen earlier, that the parties behind the veil have to decide whether to act morally, as someone such as Rawls defines it, or non-morally. If they are required in advance of arguments behind the veil to argue morally in the sense of only following a system of principles that they regard as final and not using intuition to choose between systems of principles, then they are arguing as if they know in the real world that they favour 'moral' systems of principles, that is, that they are the kind of people who require the acts they perform to be derivable from a system of principles regarded as final. This however, is contrary to the requirements of impartiality.

Rawls' requirement that systems of principles be final or overriding might be a consequence of his rule-contractarianism. He not only wants the conclusions of the parties behind the veil to be final but he also wants certain systems of principles to be final. Impartiality certainly requires that the conclusions reached by the parties behind the veil be final, but, as we have seen, the parties behind the veil could conclude that intuition is the only way to decide between certain competing values. This conclusion is final in that it defines what an impartial person would accept as the conclusive solution to how to choose between competing values. Rawls, however, wants the finality of the decisions of the parties behind the veil to be applied only to the choice of the system of principles. Once the system is chosen, Rawls wants it to be final. This means that the system has to be such that it settles all disputes between principles within itself. It doesn't generate any conflict of principles that amounts to a conflict of systems between which the parties behind the veil have to arbitrate. Rawls, then, insists that the parties behind the veil be moral

monists not moral pluralists. They must have one overriding value, not two or three equally important ones, with intuition as the arbitor. The same conclusion applies here as has previously. The parties behind the veil choose as if they know they are rule-theorists rather than act-theorists in real life if they are required prior to their arguments behind the veil, to reach only rule-contractarian conclusions. This contravenes the requirements of impartiality.

8. Reasonableness versus Morality

Katzner claims that Rawls' list of formal constraints of the concept of right confuses two distinct elements.

- (1) Constraints that select some principles rather than others because these principles fit our concept of what a principle has to be to be a moral one.
- (2) Constraints that select some principles rather than others because these principles fit our concept of what a principle has to be to be a reasonable principle of action.

He claims that 'the arguments for generality, universality (in the weak sense) and finality as constraints of the concept of right are not arguments about the reasonableness of the theory of which these constraints are a part. They are arguments about the nature of morality and moral principles'. (1980, 51)

This, however, is contrary to Rawls' own declared procedure. He says,

'I do not claim that these conditions follow from the concept of right, much less from the meaning of morality ... These conditions do, however, exclude the various forms of egoism, as I note below, which shows that they are not without moral force. This makes it all the more necessary that the conditions are not justified by definition or the analyses of

concepts, but only by the reasonableness of the theory of which they are part.' (1972, 130-131)

Katzner comments,

'Generality, universality in the weak sense (i.e., a principle must hold for all moral persons: my parenthesis) and finality do seem to be constraints which apply to all moral principles. But this must be determined by analysis of what it is for something to be a moral principle, rather than as Rawls maintains by an assessment of a (sic) reasonableness of the theory of which it is a part.' (1980, 52)

Katzner earlier points out that the other constraints on the concept of right - 'publicity, ordering and universality in the strong sense (that is, principles must be judged in terms of the effects of everyone's complying with them: my parenthesis) should be assessed in terms of the reasonableness of the theory of which they are a part; yet it is illegitimate to insist that they are constraints which apply to the choice of all moral principles'. (1980, 51)

So, if Katzner is correct, publicity, ordering and strong universalizability are not formal constraints on the concept of right, that is, a theory can still be a theory about what is and isn't morally right yet not meet these three criteria. Alternatively, a theory of what is morally right that doesn't meet these criteria, while still a moral theory, is not a reasonable one. On the other hand, generality, weak universalizability and ordering are formal constraints on a theory's being a moral theory, yet a moral theory can meet these and still be unreasonable. In Katzner's view, Rawls has confused these two quite distinct issues. Katzner appears to accept that the formal constraints do apply to the reasoning of the parties behind the veil, that is, the parties behind the veil's reasoning must be general, weakly universal and final if it is to be moral. As we have seen, however, this merely causes the parties to reason as if they know they are

in the Group A that prefers moral theories of how to act rather than members of Group B who don't. This contravenes the requirements of impartiality, which means the parties behind the veil must debate whether the principles they should follow should be general, weakly universalizable and final. They may well so conclude, but they cannot, in all impartiality, be required to so conclude before they argue behind the veil.

Rawls claims that his six constraints are justified as constraints 'on the conceptions of justice that are to be allowed on the list to be presented to the parties'. (1972, 130) This is an odd claim. It includes the strange notion that the parties behind the veil choose a conception of justice behind the veil from a list of conceptions. It entails the further strange notion that someone has censored this list, that is, there are candidates for conceptions of justice that are not on it. In particular, any egoistical theory that poses as a theory of justice is excluded. Also, any tyrannical candidate for a theory of justice is excluded (by the generality constraint). Any theory of justice that only works if it is kept secret from some of the people is excluded. Any theory of justice that settles disputes by appeal to force and cunning is excluded, and so on. Now, Rawls might be able to exclude these on the grounds that such candidates for theories of justice are charlatans, false pretenders who are not theories of justice at all. He can exclude them by linguistic fiat. But this doesn't show that they are not theories about which actions we should prefer to do, so they all remain candidates for a theory of right.

Hare has similar misgivings. He says, 'Let us remember that the main object of these conditions is to secure impartiality'. He then asks, 'How much the POPs have to be ignorant of to secure impartiality?' He points out that 'much of the work is already done by the "formal constraint"

that the principles have to be "general".' Rawls himself says that the formal constraints rule out egoism (1972, 136); it might therefore be asked what there is left for the 'veil of ignorance' to do, since to abandon egoism (and for the same formal reasons the pursuit of the interests of any other particular person or set of them) is eo ipso to become impartial'. Hare himself doesn't think that this objection sticks because 'a POP, if he had full knowledge of his own role as a POL might adopt principles that were formally "general" or universal but were rigged to suit his own interest', that is, there is a role for the veil. Hare, however, is not prepared to allow Rawls to put the veil to this use. He says, 'Rawls, however, thinks (wrongly) that such rigged principles can be ruled out on the formal ground of lack of "generality" and so is open to the objection ad hominem. That is to say, he has left nothing for the veil of ignorance to do as regards impartiality'. (1975, 89-90)

Even if Hare is right about the role that Rawls has left the veil in his theory, the same ad hominem does not apply to its role in impartialism. As we have seen, the veil is one of the set of devices for identifying when a person's consent to a principle in real life is genuine. As it turns out, it rules out non-general principles but only because it combines with the position of equal power and the requirement of noncoercionism, not because it has had an ad hoc constraint placed upon it. On my account, the mere fact that J and K are behind the veil does NOT prevent them choosing between general and non-general principles. It is only the addition of the other two axioms that leads them to always choose general principles. On Rawls' account, the mere fact that they were behind the veil meant that they were not free to consider non-general principles in the first place.

Nonetheless, as we have seen, Rawls did have a reason for including the generality constraint. Hare is optimistic if he thinks that the veil itself will ensure impartiality without a generality constraint. As I showed in the section on 'Generality' the parties behind the veil could come up with principles such as 'All conflictants are to do X except whichever conflictant was J behind the veil', and this could be agreed to behind the veil depending on the relative power of the various velorians. A generality constraint, then, is required if this is not to occur, and it can't occur if the principles the parties behind the veil are to endorse are to meet the requirements of impartiality. Hare, however, is right that, if J knew he was the conflictant called A, then he could rig principles to favour A, and this is removed by the veil. Something like the generality constraint, then, is necessary to ensure that the relationships between the parties behind the veil are suitable to generate impartial principles, and the veil is necessary to ensure that the relationship between each velorian and any conflictant he might be is such as to ensure impartial principles. It is just the ad hoc nature of the generality constraint that is the problem. The same effect is achieved by the position of equal power and the requirement of noncoercionism.

Rawls, however, doesn't give his reasons for his constraints in these terms. He talks instead, as quoted earlier, of the constraints being 'reasonable'. This raises the question of 'reasonable for what?' The kinds of things that are said to be reasonable are not so described because of some innate property of reasonableness but because they are appropriate to some purpose. A constraint is reasonable if there is some purpose that is better fulfilled with the constraint imposed. Rawls indicates that he wishes the reasonableness of these constraints to be judged in terms of their propriety 'in adjusting the claims that persons make on their institutions and one another'. He sees these constraints as 'natural'

('reasonable'?) 'if the principles of justice are to play their role, that of assigning basic rights and duties and determining the division of advantages'. (1972, 731) The problem here is that there are many 'reasonable' ways of assigning basic rights and duties and determining the division of advantages. The word 'reasonable' can only do the job that Rawls wants it to if the constraints we place on the principles by which these things are assigned are reasonable from the point of view of 'justice'. After all, we can assign basic rights and duties and divide advantages from other points of view e.g., from that of a particular egoist or in terms of a particular ideal. It would be reasonable, then, to put in other constraints. In particular, as Katzner has pointed out, not all of Rawls' constraints on principles assigning rights, duties and division of advantages are reasonable from the utilitarian point of view. Rawls, then, appears to be constraining these principles in ways that are reasonable from the viewpoint of a just distribution. He is, therefore, appealing to us to agree with him that, if we want the principles contracted behind the veil to produce a just distribution, then it is reasonable to place his constraints on them. Why would it be reasonable? Because the principles the parties behind the veil could produce without these constraints might not be just, that is, they might not match with (in fact, Rawls can see that they will not match with) his intuitions of which distributions were just. The constraints, then, do not 'follow from the concept of right, much less from the meaning of morality'. (1972, 130) Rather, they follow from Rawls' concept of justice. They derive from what he feels is meant by the word 'justice'. He has not avoided basing his theory on a definition. He has merely avoided explicating the definition he has based his theory on, leaving the reader to piece it together from the various constraints and other restrictions he places on the parties behind the veil. Hare comments, in his review of a Theory of Justice, that Rawls 'as usual says that "it seems reasonable" to impose' the

constraints on the principles. He then rather tartly adds, 'He does not tell us what he would say to somebody to whom they did not "seem reasonable" '. (1972, 88) The answer to Hare's point, however, is clear enough. Rawls must say that his concept of what justice is differs from that of anyone who does not find the constraints reasonable. It is possible, therefore, for people to find different constraints reasonable, and still see their own constraints as appropriate for justice, because they have different concepts of justice. Rawls is, then, faced with the problem of either showing that his is the correct concept of justice, or else of showing that his distribution is better (from an agreed upon point of view) than the ones that come from different concepts of justice, even though his meaning for the word 'justice' is no more correct than anyone else's. Even if he can show that his meaning for the word 'justice', as embodied in his selection of constraints, is more correct (closer to common usage or whatever?) he would still have to show why this makes it the one to choose. He cannot say that it is more reasonable, if he means by this that it is more reasonable from the viewpoint of justice as he defines it, because this merely begs the question. Rawls, then, needs to define justice independently of the veil, the constraints, etc. He can then use this definition to identify those features of the veil that are consistent with, or required by, this definition, then use the veil to judge principles. He will still need to show however, why his sense of justice is to be the standard by which he determines whether or not an act is right or wrong, a principle or institution is to be preferred, etc.

Chapter 5

Impartialism and Non-Persons1. Introduction

The last chapter concluded the section of the thesis in which I showed that many of the criticisms of the veil of ignorance lose their force if the scope of the veil is not constrained in the way Rawls recommends. As a consequence, the range of conflicts of desires about which the veil has to arbitrate has greatly increased. This makes the task of coming up with a determinate theory of right more difficult but greatly minimises the likelihood that the theory begs the question against one side of any particular conflict of desires.

In this section I will look at claims that the normative conclusions derived from the veil are biased by its ethical or metaphysical assumptions. I will show that these criticisms fail. The veil requires only that the parties do not know which side of a conflict of desires they are on. It is the other two impartialist axioms that give any substantive content to impartialism as a theory of right. The veil serves only to exclude those who are egoistic in the sense that they want the desires of whichever of A or B they happen to be satisfied.

I will begin with the criticisms that Haksar makes of what he believes to be a metaphysical assumption of the veil.

Haksar (1979) argues that perfectionist egalitarianism produces a greater reflective equilibrium between our moral and political principles and our

moral intuitions than any of the alternatives, provided we also try to establish a harmony between these and our metaphysical principles. In particular, he believes that perfectionist egalitarianism produces a more comprehensive reflective equilibrium than contractarian egalitarianism of the kind proposed by John Rawls. It is my intention to show that impartialism is preferable to Haksar's perfectionist egalitarianism because it reduces the reliance on intuition unavoidable in Haksar's perfectionist approach.

Haksar's theory is egalitarian, he says, because it holds the following two principles:

- (1) there are certain rights held equally by all human beings, especially the right to equal respect and consideration;
- (2) these rights are owed to the individual for his own sake, not because his possession of these rights has desirable consequences of some kind. (1979, 65-66)

His theory is perfectionist because it holds the following three principles:

- (1) some forms of human life are intrinsically inferior to other forms of human life;
- (2) human beings, with the possible exception of some such as congenital idiots, have more intrinsic worth than animals;
- (3) human ways of life, at any rate those that are not anti-social such as Nazism, have more intrinsic worth than animal forms of life.

Haksar claims his perfectionist egalitarianism to be superior to Rawlsian contractarian egalitarianism because the particular moral judgements that Rawls favours cannot be derived from his theory unless perfectionist and metaphysical considerations are added.

Haksar argues that the following claims cannot be derived from contractarianism without appeal to independent metaphysical or perfectionist considerations

- (1) that congenital idiots are members of the egalitarian club;
- (2) that human fetuses are members of the egalitarian club;
- (3) that children are members of the egalitarian club;
- (4) that animals are not members of the egalitarian club;
- (5) that the senile and permanently comatose are members of the egalitarian club.

His claim raises two issues that I wish to consider.

- (1) Is it true that these positions can't be derived for contractarianism without appeal to independent metaphysical or perfectionist considerations?
- (2) If it is true, does this demonstrate that perfectionism is a better moral theory than contractarianism. It may be that an adequate moral theory does not need to be able to derive all of these positions.



I will be concerned to see whether impartialism rather than Rawls' version of contractarianism can answer these questions. The five claims I have listed above are all claims to the effect that a contractarian approach cannot derive equality of rights for non-persons without introducing non-contractarian considerations. Haksar regards it as a fault of a theory that it can't use its own premises to show that all these claims are true. In essence, he believes that humans have a value lacked by non-humans, even when these humans are not yet, nor never will be, rational agents. Before I look at these claims in more detail I will need to distinguish persons from non-persons.

2. Persons and Non-Persons

Singer (1979, 75) makes a distinction between persons and non-persons. A person is not necessarily a human, nor is a non-person necessarily non-human. Singer follows John Locke in treating rationality and self-consciousness as the defining characteristics of person-hood. The consequence of this definition of personhood is that most animals, foetuses, children, congenital idiots and senile adult humans are not persons. This means that they cannot be velorians, that is, they cannot be parties to the arguments about what principles should govern conflicts of desires. This, in turn, raises questions about whether only the desires of persons should be taken into account by the velorians in their reasoning.

The veil could be constructed in either of two ways. (1) Once behind the veil the parties do not know what kinds of beings they are. They only know that they are beings with desires. (2) Behind the veil the parties at least know that they are persons in real-life. The second strategy, however, just arbitrarily removes one possible set of conflict of desires from consideration, which means deciding against one of the parties to the conflict prior to any application to their desires of the impartialist

axioms. If this is done, then Haksar would appear to be right that considerations external to impartialism such as perfectionism have been introduced to arrive at the conclusion that the desires of non-persons are to have less weight than the desires of persons. If such external factors are not to determine the conclusions, then the impartialist axioms themselves would have to be sufficient to give differential weighting to the desires of persons and non-persons. This means that the parties J and K will have to take into account the desires of non-persons as well as persons. They will not be able to restrict their reasoning solely to conflicts of desires between persons but will also have to consider conflicts of desires between persons and non-persons.

J and K, then, have to consider conflicts between the desire of a human to trap or kill an animal and the animal's desire to escape, and between the desire of a human to experiment upon an animal and the animal's desire to escape from a painful situation. As the conflict is between the desires of humans and the desires of animals, they will have to try to formulate principles to govern the conflict of desires in terms of the desires as felt by the parties to the conflict. They would produce principles that were not germane to this particular conflict of desires if they reasoned about the desires they would have if they were the animal or the human rather than the desires the animal and the human actually have. If the animal, for example, has no desires about its own continued existence, then J and K cannot reason as if it had such a desire just because they would desire the animal's continued existence if they were to become that animal.

Whatever conclusions the velorians reach, the principles to resolve such conflicts of desires can only be binding on beings capable of entering into, and honouring, contracts. This is why I did not have the velorians

considering conflicts of desires between animals. A human, then, who is committed to acting in an impartialist manner will treat animals in accord with the principles that the utilitarians endorse for the relations between humans and animals even when he is the only human around and does not have to justify his treatment of animals to other humans. If two groups of humans A and B disagree over Group A's treatment of animals, then a member of A will first have to put a veil over which side he is on in the conflict between A and B. Then, in order to resolve that conflict, he will have to put a further veil over which party he is in the conflict between humans and animals. He then has to determine as accurately as he can what the actual desires of the parties to the conflict are, even when these may be radically unlike his own. He does NOT, however, have to do as Hare's theory requires, that is, in Taylor's words (1970, 55-56), 'assert in propria persona to an imperative prescribing that a certain action be done to him, given that he is in the same position as the victim of the action at present in question'. In impartialism, however, we are not asserting to imperatives in propria persona but reasoning about the desires of each party as if we did not know which party we were, so impartialism is not subject to Taylor's arguments about the logical impossibility of Hare's argument.

The utilitarians, then, have to argue as if they do not know which kind of sentient being they are but, whenever they put themselves in the place of one of the conflictants, they are to reason about the conflictant's desires as the conflictant conceptualises them.

Given these requirements, they would treat any desire as equally worthy of consideration as any other equally strongly felt desire. They will have to take into account, however, a difference between the desires of persons and the desires of non-persons. Persons have desires about the

satisfaction of their desires whereas non-persons do NOT. Singer (1979, 81) makes this distinction in terms of how these desires can be frustrated. There are two ways, he says, in which desires can be frustrated.

- (1) One could say that a desire is frustrated when the owner of the desire experiences the frustration of having his desire thwarted. An example is where someone wants to go to the movies, arrives at the ticket box but finds all the tickets sold or that he forgot his money. He feels frustrated. He is aware or conscious of not getting what he wants.
- (2) One could say that a desire is frustrated when the owner of the desire doesn't live to have the desire satisfied, or, perhaps, when the owner of the desire loses all consciousness of whether or not his desires have been satisfied, as when he is in a coma. In this case, he doesn't experience any frustration because he is either dead or unconscious but, nonetheless, he may well have had desires about what he wanted to happen tomorrow, or next year, or when he retired, which he will now never be able to satisfy.

Both animals and humans have desires but there is a crucial difference between the desires of most animals and the desires of most humans. Whereas most humans have desires about what they would like to happen to them in the future, most kinds of animals do not or, if they do, they have them in a very crude form.

This difference is clearly illustrated with the desire for life. A being only has a desire for life if it is able to comprehend what it is to be alive and what it is to be dead. Singer (1979, 81) argues that animals do not have the concepts of 'life' or 'death'. Cases in which they behave

as if they have the concept are only apparent. Singer says,

'A being might struggle against a situation in which its life is in danger, as a fish struggles to get free of the barbed hook in its mouth; but this indicates a preference for the cessation of a state of affairs that is perceived as painful or threatening. Struggle against danger and pain does not suggest that the fish is capable of preferring its own future existence to non-existence, the behaviour of a fish on a hook suggests a reason for not killing fish by that method, but does not suggest a preference - utilitarian reason against killing fish by a humane method.'

Conscious beings, then, can be divided into two classes:

- (1) those which have a concept of what it is for them to be the same being over a period of time, which have a concept of death, and which have desires for their own futures;
- (2) those which lack these.

The parties behind the veil of ignorance do not know whether they are human or non-human. They do know, however, that if they are persons they will have a concept of a continuing self, they will have desires about their own futures, that they will not want those desires frustrated and that the prospect of those desires being frustrated will fill them with alarm. They also know that, if they are non-persons, the frustration of their future desires will cause them no alarm in the present. They examine their possible frustration, not from their perspective as rational agents behind the veil of ignorance, but in terms of how they would feel once they had become a particular sentient being in real life. They know that, were they to be a non-person, they now would not want that non-person's desires frustrated, whether he was aware of the frustration or not but they also know that, once they are a non-person, the non-occurrence of future desires will lead to no sense of frustration to the non-person they had become. Clearly, only self-conscious rational agents can argue about the rights and

wrongs of killing non-persons and thereby eliminating the desires these non-persons might have had. As such, rational self-conscious agents will be inclined to treat the case of non-persons as if they were being experienced by persons. This, however, is not the case. The rational self-conscious agents behind the veil of ignorance must differentiate between the cases of persons and non-persons in real life in terms of the desires that persons and non-persons are capable of having in real life. The argument between the parties behind the veil of ignorance is not about whether animals should be members of the egalitarian club if they were persons but about whether they should be if they are not persons.

Behind the veil of ignorance, the parties will agree that, if they are self-conscious agents in real life, they will contract with each other to ensure that their future desires are not frustrated by their being killed while unconscious or while asleep. The contractarian theory is able to answer a problem that has concerned utilitarians since the 19th century, namely, what is wrong with killing people in their sleep or while they are unconscious other than the effects of this on their conscious family and friends?

Behind the veil of ignorance, the parties will agree that, if they are not persons in real life, then, provided they are killed painlessly, there is nothing wrong in their being killed. This, perhaps, is too strong. Rather, there is nothing directly wrong in their being killed. If there is a wrong, then it is either because other sentient creatures suffer as a result of the killing or else because the person who did the killing acted for motives of a vicious or otherwise undesirable kind, that is, he showed an undesirable character. What makes his character undesirable will, on the contractarian model, be something that the parties behind the veil unanimously agree to be in their interests to prevent.

The contractarian model, then, allows us to conclude that persons should be members of the egalitarian club but that non-persons shouldn't. It does this without resorting to perfectionist assumptions. One of Haksar's claims, therefore, is false. Insofar as animals are non-persons, impartialism as a form of contractarianism excludes them from an equal right to life with those humans who are persons.

This, however, still leaves Haksar's other challenges. Not all humans are persons. Foetuses aren't persons, nor are infants, nor congenital idiots, nor the senile, or those in permanent comas. I will now turn to the arguments Haksar uses to show that congenital idiots are members of the egalitarian club.

3. Congenital Idiots and Perfectionism

The congenital idiot is an interesting case because he lacks even the potential to be a person. There might be an argument based on potentiality that entitles us to include non-persons in the egalitarian club on the grounds that they will become persons in the normal course of events. Haksar argues this for foetuses and infants. But this is not available for congenital idiots who, by definition, have no such potential.

Haksar thinks congenital idiots should be treated as associate members of the egalitarian club, whereas animals shouldn't. He does this, even though he agrees that there are no relevant differences between congenital idiots and animals. His reason for this is as follows: defectives who began with the full potential but whose potential was destroyed are to be treated as members of the egalitarian club because, Haksar believes, once a member always a member. He calls this the 'doctrine of the transitivity of ends in themselves'.

I do not accept this doctrine but its rightness or wrongness is not really relevant to the present discussion. Congenital idiots are often indistinguishable from non-congenital defectives. If non-congenital defectives are members of the egalitarian club then, says Haksar, we have a 'pragmatic' argument for treating congenital defectives as members of the club. As he says, 'There is the danger that if you give inferior status to the congenital idiot, some ordinary human beings as well as some non-congenital idiots may be given inferior status by mistake or by malicious design'. (1979, 75)

This argument depends on the lack of clear demarcation between congenital and non-congenital idiots, and, to a lesser extent, on a similar possibility of 'confusion of similars' between defectives in general and normal humans. It also rests on there being clear demarcation between humans and animals. As Haksar says, 'So the practice of giving inferior status to animals is not likely to be abused in the way that the practice of giving inferior status to congenital idiots is'.

Haksar's pragmatic argument isn't really a perfectionist one. It does not establish that congenital idiots are members of the egalitarian club. Nonetheless, it attempts to show that a perfectionist is not merely being arbitrary in treating congenital idiots with greater consideration than he gives animals. A contractarian can adopt the same strategy.

4. Impartialism and Congenital Idiots

Behind the veil of ignorance the parties do not know what kind of sentient being they are in real life. For all they know, they may be congenital idiots or animals. The criteria established earlier for treatment of animals would then apply, i.e. as neither animals nor congenital idiots are

'persons', with life-plans, then neither will have their desires frustrated by painless killing. Painlessly killing either is therefore permissible. Neither, however, will want to suffer feelings of fear or frustration, therefore the parties will contract to minimise such conditions for either. So far, animals and congenital idiots would be treated as on a par. When, then, might the parties agree to give preferential treatment to congenital idiots? They would do so under the following circumstances:-

- (1) the usual 'indirect' classical utilitarian kinds of circumstances, e.g., where other people would be made unhappy by their not being treated differently;
- (2) the 'pragmatic' argument Haksar has given, stripped of its reference to confusion between congenital and non-congenital cases, i.e., the parties would accept the argument from 'confusion of similars'. Unless there are clear markers between congenital idiots and normal adults, then unscrupulous people might treat normal adults as if they were congenital idiots, i.e., treat them as they might animals by killing them painlessly. If this is not permissible with congenital idiots, then making such a mistake cannot be an acceptable excuse. Furthermore, if congenital idiots are allowed to be killed painlessly, then this could lead to people deliberately causing the birth of congenital idiots to serve their own purposes. It may be that the disposition that allows a person to turn what would otherwise have been a normal human into a congenital idiot would also allow him to turn a normal adult into a non-congenital idiot to serve his own purposes. If such a disposition is suppressed in the case of causing the existence of congenital idiots, then this may prevent its manifesting itself in 'idiotising' normal adults. The parties behind the

veil would not want to be turned into non-congenital idiots against their will if they were normal adults. They would contract to forbid this.

We see, then, that giving greater consideration to congenital idiots is derivable from the contract situation. It does not assume perfectionist presuppositions.

5. Abortion and Potentiality

Haksar argues that fetuses should be members of the 'egalitarian' club. This is because egalitarianism requires what he calls the 'dynamic' as opposed to the 'static' view of intrinsic worth.

The static view of intrinsic worth is that an individual only is entitled to be treated as a member of the egalitarian club if he possesses at the time the abilities that warrant membership of the club, e.g., self-consciousness, rationality. It is not enough that he has the potential for these abilities. Haksar attributes this view to Feinberg (1974). The dynamic view, which Haksar endorses, treats an individual as a member of the egalitarian club if he has or had the relevant potential. Whereas the static view would exclude fetuses, infants, the senile, the comatose, the psychopath, the catatonic schizophrenic and the dead from the egalitarian club, the dynamic view includes them.

My interest in these classifications is that Haksar claims that contract theory of the 'veil of ignorance' kind must also opt for either the static and the dynamic view of intrinsic worth, yet it cannot choose between these options unless it relies on non-contractualist considerations. It must, in Haksar's terminology, import 'perfectionist' assumptions.

I have already shown that the parties behind the veil of ignorance would contract that persons are not to use the same criteria when deciding how to treat persons as they use when deciding how to treat non-persons. Non-persons include foetuses, infants, the senile, the comatose, the catatonic schizoid, and the dead, in fact, all of Haksar's list except the psychopath. Whereas persons are automatically members of the egalitarian club, non-persons are not, as we saw with the case of the congenital idiot. If non-persons are to be treated as if they were members of the egalitarian club, then argument is needed as to why this should be so. A contractarian, then, is committed to the 'static' view of intrinsic worth unless there is a contractarian argument to show that present and/or past potentiality to be a person entitles an individual to membership of the egalitarian club.

Haksar clearly regards it as a defect in a moral theory that foetuses, infants, the senile, the comatose, the catatonic schizoid and the dead are not granted certain rights granted to persons. He believes that all of those groups, except the dead, have a right to life. Those who were once persons are entitled to have contracts or promises honoured if these were made about situations that might obtain if these individuals ceased to be persons. Haksar, therefore, seeks a reflective equilibrium between theory and particular moral judgements (intuitions) such that his theory generates at least these particular moral judgements. Contractarianism may not generate precisely these particular moral judgements. The mere fact that it does not do so is not necessarily a defect. Haksar needs independent arguments to show why a perfectionist reflective equilibrium is to be preferred to a contractarian one.

Haksar attributes the rights he does to the kinds of non-persons listed earlier because he treats 'potentiality' to be a person as a rights-conferring characteristic. It may well be that there are good contractarian

arguments for treating potentiality this way, although these contractarian arguments may not confer the relevant rights on all of the groups Haksar wants.

Behind the veil of ignorance the parties would contract that persons should give similar rights to life to the comatose, the senile, the catatonic schizoid as are given to normal persons. They would also contract that persons give priority to normal persons rather than to members of these abnormal groups in situations where one has to die to enable the other to live. They would certainly not treat the comatose, etc., as liable to the kind of painless killing permitted of animals. This is because the parties know that, if they were human persons, they would not want to be killed while in a coma or while a catatonic schizoid as long as they possessed the potential to become a person again. Potential, therefore, is clearly relevant in determining the moral rightness of an action if the beings under consideration are likely to suffer at the prospect of being killed while they retain a potential for recovery. This is not the case, however, with foetuses and infants, who will not suffer at the thought of their never realising their potential. If foetuses and infants are to be given the same right to life as persons, different arguments than the 'potential' argument will be needed. The senile, too, constitute a different case as they have lost their potential to be a person. Their case has become like the congenital idiot case. Nonetheless, persons will worry now about how they will be treated when senile. The parties behind the veil will contract to alleviate this worry by requiring persons to give the senile the same right to life as the non-senile, with the proviso stated earlier. Singer gives an example of how this might be done. He says,

'Elderly people, knowing that nonvoluntary euthanasia is sometimes applied to senile elderly patients, bed-ridden, suffering and lacking the capacity to accept or reject death, might fear that every injection or tablet will be lethal. This fear might be quite irrational, but it would be difficult to convince old people of this; particularly if old age really had affected their memory or powers of reasoning.

This objection might be met by a procedure allowing those who do not wish to be subjected to non-voluntary euthanasia under any circumstances to register their refusal. Perhaps this would suffice; but perhaps it would not provide enough reassurance. If not, non-voluntary euthanasia would be justifiable only for those never capable of choosing to live or die.' (1979, 139)

Contract theory of the kind I have been using, therefore, requires us to treat those with a potential for personhood as if they were persons, provided they have already been persons. Contract theory, therefore, agrees with Haksar that the comatose, and the catatonic schizoid, should have the same right to life as ordinary persons. Likewise, contract theory concludes that the senile should have this right, unless they voluntarily consented to waive it when they were persons. While this is not the place to spell out in detail a theory of euthanasia, the parties behind the veil would contract to build in certain safeguards against misuse of euthanasia, or its careless implementation.

Haksar's own 'pragmatic' justification will apply in those cases of catatonic schizoids who have no potentiality to become persons again. Until there are clear markers separating person from non-person categories, all catatonics would be treated as persons. This is because the parties behind the veil know that persons would worry that they might lose their right to life while catatonics, even though the catatonic had the potential to become a person again. This would worry them while they were persons. They would be less concerned about losing their right to life if they had become permanently non-person catatonics, although even this might worry

some of them. This could be overcome by prior consent to, or refusal of, euthanasia. The parties know that, once they are non-person catatonics, they will not care if they are painlessly killed, that is, the parties will not contract to require that non-person catatonics be killed. As non-persons they will live in the immediate present and the near past, with very little anticipation, if any, of the future. They will not be afraid of dying but nor will they resent staying alive.

It would appear, then, that the fact that fetuses have the potential to be persons does not provide a conclusive reason why they should be given the same rights as persons. It is not a defect, then, of contractarian theories that they cannot deduce that fetuses have these rights without importing perfectionist assumptions. I will take the implications of impartialism for the moral status of fetuses AND infants in more detail in the next chapter.

Chapter 6

Assumptions about Personal Identity1. Haksar's Criticisms

Haksar (1979, 106-107) claims

'On the reflective equilibrium model, according to Rawls, we adjust our moral (and political) principles and our moral intuitions until we get a harmonious fit. I should like to add one more complication to the reflective equilibrium model. The fit has to be not just between our intuitions and our moral and political principles, but also between these and our metaphysical principles. We must examine the metaphysical assumptions and presuppositions of our intuitions and of our moral and political principles and then see whether these metaphysical assumptions are acceptable. In cases where we cannot on reflection accept the metaphysical assumptions, we may have to abandon the intuitions or the moral and political principles whose presuppositions we cannot accept. Rawls is quite wrong to think we can bypass metaphysical (and perfectionist) considerations in constructing a theory of justice. For instance, we saw that his criticism of utilitarianism becomes less plausible if we abandon the metaphysics of the persistent self, which Rawls presupposes.'

Let us assume that Haksar is right that reflective equilibrium requires a fit between metaphysical principles, intuitions and moral and political principles.

What are the consequences for the impartialist theory I proposed earlier of the existence of two major metaphysical theories of personal identity? Haksar believes that the particular moral judgements derivable from Rawls' form of contract theory depend on his presupposition of the persistent self. If one adopts instead an alternative metaphysic which denies the existence of the persistent self then, Haksar argues, Rawls' anti-utilitarian particular moral judgements (intuitions) no longer follow.

Is this also true of the impartialist version of contract theory?

Before I try to answer this question I will need to spell out in sufficient detail the two contending metaphysical theories of personal identity.

2. Parfit's 'Complex' theory of Personal Identity

We will need a terminology in which to pose alternative theories of personal identity, without thereby begging the question against any particular theory.

Parfit uses the notion of a 'self' to explicate his theory (1973, 1976).

An indication of what he means by the term can be gained from the following,

'It would be even more accurate to abandon talk about 'selves' and to describe actions, thoughts, and experiences on a quite 'impersonal' way ... If these are not ascribed to any 'subject', their various interconnections can then be directly specified. But the concept of a 'subject of experience', like that of a nation, is an abbreviatory device of enormous convenience. If we remember that it is just this and nothing more, it can be safely used.' (1973, 163)

Haksar offers no objection to this terminology, so I will use it to pose the disagreement between him and Parfit.

A 'self', as described by Parfit, is a technical term bearing little relation to the ordinary language use of 'self'. It refers to the contents of consciousness of a particular individual over some period of time. It is, in fact, nothing more than a particular time-slice of the consciousness of a particular bodily individual.

Generally speaking, there will only be one self to any one body at any one time. I will refer to the self in the body B1 at the time T1 as the self B1 S1. This self will have visual, tactile, auditory, olfactory, gustatory kinaesthetic, pain etc. sensations from a perspective located within B1. I will, therefore, say that the self B1 S1 is "located" in B1. It is possible that B1 may contain more than one self at T1, or that there may be selves that are located at a particular spatial point where

there is no spatio-temporal body. The terminology of 'selves' does not rule out these possibilities. Whatever is said about the central case of one self in one body at one time can be modified to apply to these cases where they occur.

The terminology of 'selves' should not beg the question in favour of one metaphysical theory rather than another. This means that whatever makes the time slice of the consciousness of B1 at T1 the 'same' self as the time-slice of B1's consciousness at T2 must not depend on which metaphysical theory of personal identity is true. We must be able to pose any metaphysical theory about personal identity in 'self' terminology without thereby guaranteeing its truth or falsity by definition. This makes the duration of a 'self' a matter of convenience. Sometimes it will suit us to start S1 at the onset of consciousness in B1 for the day and conclude S1 when B1 goes into deep sleep for that day. Other times it will suit us to treat S1 as a year in length, or a mere thought in length.

Another technical term I wish to introduce is 'individual'. I shall say that B1 S20 believes itself to be the same bodily individual as B1 S1, when it believes that its body is the same body as B1 S1's body. B1 SN believes itself to be the same suffering individual as B1 S1 when it believes that B1 S1's pains were also its pains. B1 SN believes itself to be the same remembering individual as B1 S1 when it believes that the experiences that B1 S1 remembers are the same experiences that it remembers. There may, of course, be no such things as individuals. Whether they exist is one of the bones of contention between the competing metaphysical theories of personal identity. We could also have 'intending' individuals, 'feeling' individuals (in the sense of 'feeling' emotions), and so on.

If B1 S1 is the same individual as B1 S2 then I shall say that B1 S1 and B1 S2 are both instances of L1. This can be represented by L1 B1 S1 and L1 B1 S2. If however, they are not the same individual, this will be reflected by a change in the notation such as L1 B1 S1 to L2 B1 S1.

Parfit believes that S1 is the same individual L1 as S2 when (a) there is psychological connectedness between S1 and S2. 'Connectedness' is a 'direct' psychological relation, that is, it is the kind of relation that holds between a memory of an experience and the experience it is a memory of, or between the intention to perform some action and the action. Personal identity is a matter of degree. The greater the connectedness between S1's experiences and intentions and S2's memories and actions, then the more inclined we are, Parfit believes, to say that S1 and S2 are part of the same individual L1. The less these connections, the less inclined we are to identify them.

(b) psychological continuity between S1 and SN. S1 is psychologically continuous with SN if there is a chain of connections between S1 and SN, for example, SN remembers S3's experiences, S3 remembers S2's, S2 remembers S1's, but SN need not remember S2's or S1's and S3 need not remember S1's, and so on. The more continuities there are between S1's experiences and intentions and SN's memories and actions, the more inclined we are to say that S1 and SN are both instances of L1.

Whether S1 is the same individual as S2 or SN is a matter of degree. It is a convention we adopt because it suits our purposes in certain kinds of case. Some purposes require a greater number of connections than others. For Parfit, the identity of S1 and S2 is analagous to the identity of England before and after 1066. Was it still the same 'nation'? From the point of view of geography it was but from the point of view of political

control it is no longer the same nation.

4. Haksar's 'Simple' Theory of Personal Identity

Haksar (1979, 81, 83) claims that personal identity is not a matter of degrees. It is an all-or-nothing business. The identity of S1 and S2 is not analagous to the identity of a nation. Rather, it is as if there is an underlying 'metaphysical' substance called the 'individual' that endures from S1 to S2, linking them together through changes of memory and intention and, possible, changes of body.

5. What Hangs on Which Theory is Chosen?

Haksar believes that Rawls' contractarianism assumes the truth of the simple view of personal identity. If the complex view is true many of Rawls' particular moral judgements would not follow from his veil of ignorance conditions. In particular, he would not be able to avoid utilitarian moral judgements.

My concern is whether Haksar's criticisms of Rawls' contractarianism also apply to impartialist contractarianism. This depends on whether or not the impartialist contractarian must assume the truth of either one or the other of these two metaphysical theories of personal identity (or of any other such theories). I will argue that he does not need to assume the truth of any metaphysical theory of personal identity. To do this I will look at the particular moral judgements that Haksar claims do depend on the simple view and show either that impartialist contractarianism can generate these judgements without relying on any metaphysical view of personal identity or else show that impartialist contractarianism can reject these particular moral judgements without depending on any metaphysical view of personal identity.

6. Continuity of Body and Potentiality

Haksar (1979, 84-86) states that the human infant has the potential to lead a significant life, except for such cases as congenital idiots. He believes that infants with this potential have the right to develop it. Their having this right, he claims, depends on their being the same substance as the adult they eventually become. This right, then, depends on the simple view of personal identity being true. If the complex view is true, the infant's right to develop its potential can always be sacrificed for utilitarian ends. Haksar says,

'On the complex view the infant is not in any deep sense the same individual, the same substance, as the adult it will form into. And so the fact that it has this wonderful potential can at best be a reason for thinking that it has instrumental worth; on the complex view it is not easy to see why the newly-born human infant has any more intrinsic worth than an ant or a bee or a sperm.' (1979, 84)

The questions that Haksar's claims pose for impartialist contractarianism are

- (1) Does impartialist contractarianism generate the particular moral judgement that human infants with the potential to lead significant lives have the right to develop this potential?
- (2) If impartialist contractarianism accepts or denies this particular moral judgement does it do so because it either accepts or denies one of either the simple or the complex theories of personal identity?

We have already seen that people arguing as if behind a veil of ignorance must argue as if they didn't know what kind of sentient being they are. They can, however, hypothesise that they might be sentient beings of a particular kind. They can then discuss the contract they would make about how beings of this kind should treat each other. This contract would then be binding on all actually existing beings of this kind, including the

people arguing as if behind a veil of ignorance if they are human beings of this kind.

This type of contractarian exercise does not pre-empt decisions about how beings of this kind treat beings of a different kind. When this is raised, people arguing as if behind a veil of ignorance will have to consider any proposed principles of action from the viewpoint of all affected kinds of being.

Let us consider a particular kind of being X . Each member of X at some time T_N is a self, S_N , of some particular body B_N . Let us consider the case of S_N of the body B_1 . The body B_1 has changed over time, as do the bodies of all members of X . At some time T , body B_1 was in its infant stage. Now, twenty years later at T_{20} it is in its adult stage. It is, however, still properly identified as B_1 because there has been a spatio-temporal continuity to all its changes. The selves located in B_1 in its infant stage were not persons in the technical sense of "person" introduced earlier, that is, they were not rational self-conscious agents. Nonetheless, given normal conditions of development, the selves in the adult stage of B_1 will be persons. The infant B_1 then possesses the potential to become a person, but is not yet a person.

What contracts would be reached by people arguing as if behind a veil of ignorance about how such beings should treat each other at their various stages of development? They are to argue as if they don't know whether or not they are B_1 or B_2 or, if they are B_1 say, as if they don't know whether they are the infant $B_1 S_1$ or the adult $B_1 S_{20}$.

Also, as I pointed out earlier, each velorian is to argue as if his desires were the desires of B1 S1 or B1 S20. He is not to argue as if what matters is how he, the velorian, wants B1 S1 to be treated should he ever turn out to be B1 S1 but rather he is to argue as if what matters is how B1 S1 thinks B1 S1 should be treated in the circumstances under consideration.

The velorians, however, do know that, if they are B1 S20 then what they are able to do now at T20 to implement their plans will be greatly affected by what has been done to B1 at T1 and at the times in between. They know that B20 will not want his plans frustrated because injuries done to B1 at T1 make it impossible at T20 to do what he wants. They will, therefore, want to prevent people negligently or maliciously causing the infant B1 any injuries at T1. The velorians, therefore, will contract to penalise people who do cause such injuries in order to ensure that such injuries do not occur to them, whether they are B1 or B2 or BN. They will do this even if they only believe that the selves of B1 from T1 to T19 are selves of the same bodily individual as the self at T20. Whether S1 to S20 is the same individual in any other sense in no way affects the degree to which S20 has been handicapped at T20 by the injuries B1 has received between T1 and T19. People arguing as if behind a veil of ignorance, then, do not need to presuppose any particular theory of personal identity in order to reach a contract about how adult members of the species X should treat infant members of X. This contract, however, does assume that a particular self S20 of some body B of the species X wants handicaps to the implementation of its plans minimised but this is not an assumption about personal identity. It is, rather, a consequence of adult X's being persons, that is, rational self-conscious agents.

As the utilitarians argue from the viewpoint of the being whose interests are affected, they will be indifferent as to whether or not Bl Sl is painlessly killed because Bl Sl is not a person with the concepts of himself as continuing to exist over time with a future and a past. Bl Sl lacks the concepts of life and death, consequently he cannot desire to be alive rather than dead. No desire of his is frustrated by killing him painlessly. As Bl Sl, however, utilitarians will seek to avoid threatening and painful situations so they will wish to be protected from these. They will not be concerned to ensure that they develop their potential to lead significant lives because they lack the concepts necessary to possess such a desire. They will contract, therefore, to prevent cruelty to infant X's, but it is by no means obvious that they would contract to give an infant X the right to develop its potential to lead a significant life. If they do give it this right it will be for some "indirect" reason such as the distress caused to adult X's by the thought that infant X's lack this right. It will not be because infant X's have more intrinsic worth than an ant or a bee or a sperm. An infant X is not relevantly different to ants etc. in spite of its potential to lead a significant life. Like ants etc. its desires are NOT frustrated by its being killed painlessly, nor is the lack of frustration altered by the fact that it possesses a potential that ants etc. do not.

We have seen, then, that impartialist contractarianism can generate distinctions between kinds of beings relevant to conclusions about how they ought to be treated. I have not worked through completely where impartialist contractarianism would take us on the moral status of infants as this would require an investigation of a range of factual issues of no direct bearing on the present enquiry. Nonetheless, enough has been done to show that impartialist contractarianism can lead to particular moral judgements without presupposing the truth of either the simple or the

complex theory of personal identity. All that is needed is that the utilitarians consider the case of a rational self-conscious agent B1 S20 whose body B1 at T20 either is, or is believed by B1 S20 to be, spatiotemporally continuous with a particular infant body at T1 such that he believes any harm to B1 at T1 frustrates or hampers the implementation of his plans at T20. He does not need to believe, nor does it need to be the case, that he is an individual in any sense other than the same bodily individual L1 B1 S1 to L1 B1 SN.

7. Continuity of Suffering

Haksar (1979, 111) argues that any theory which requires the pursuit of utilitarian goals to be constrained by non-utilitarian rights must presuppose the simple view and reject the complex view. As contract theories are intended to generate such non-utilitarian moral constraints they must presuppose the simple theory. As Haksar says,

'If Parfit's complex theory is correct, if there are no persistent individuals (except in a trivial sense), why should we get so worked up about suffering in the world? Suffering would still be real, but how much worse it is when (intrinsically) the very same individual keeps suffering on and on. Such considerations make one suspect that the complex view is not only incompatible with a right-based approach ... but also with any kind of humane morality. Parfit rightly points out that on his view a person should not be frightened by the prospects of his own death, death is not as bad as it appears to the ignorant. He might have added that on his view murder is not such an evil as it appears to the old-fashioned.'

Does impartialist contract theory, then, agree with Haksar that it is much worse for the same individual to go suffering on and on than it is for the same amount of suffering to be divided among different individuals?

Whichever it is, is it the case that impartialist contract theory can only arrive at its answer by presupposing either the simple or the complex view of personal identity?

Earlier I said that the self B1 SN believes itself to be the same suffering individual as B1 S1 when it believes that B1 S1's suffering is the same as its suffering.

Any particular self B1 S1 may believe that he is the same suffering individual as the self B1 S20 or he may not. He may also believe that he is the same suffering individual as B2 S20, but I will limit my discussion for the moment to cases where the selves that regard themselves as belonging to the same individual are all located in the same body. The self B1 S1 may be wrong in believing that he is the same suffering individual as B1 S20. He will be wrong if the simple theory of personal identity is false and the complex theory true. Nonetheless, regardless of the truth or falsity of either theory, B1 S1 regards himself as the same suffering individual as the selves B1 S2 to B1 S20. This belief has certain consequences for how he regards the suffering of B1 S20. Granted that B1 S1 is a rational self-conscious agent, he will want to minimise his suffering, unless, in some way, greater suffering implements his plans better than lesser suffering. This exception, however, raises questions about what is to count as a rational agent rather than whether our judgements about the wrongness of suffering depend on one theory of personal identity rather than another. I will assume, then, that B1 S1 wants to minimise his suffering. As he regards himself as the same individual as the selves B1 S2 to B1 S20 he will also want to minimise the suffering of B1 S2 to B1 S20.

He doesn't need to believe that his later selves will remember his thoughts, experiences, emotions or intentions or even that these later selves will have relevantly similar thoughts, intentions, etc. While a self B1 S1 may believe it is part of some suffering individual L1 this doesn't mean that it is right. Contract theory would have to beg the metaphysical question

against the complex theory if it needed to assume that B1 S1's belief was true before contractarian moral conclusions could be derived from it.

The velorians, however, do not know whether they are a self that believes it is part of an suffering individual or not. This means they will have to consider a different contract for each of the following cases:

- (1) A contract governing how the selves of a particular suffering individual L1 should treat other suffering individuals L2, L3, etc.
- (2) A contract governing how suffering individuals should treat the selves of nonsuffering individuals, and vice versa.
- (3) A contract governing how the selves of a particular suffering individual LN should treat each other.
- (4) A contract governing how the selves located in the body of one nonsuffering individual should treat each other.
- (5) A contract governing how selves located in different nonsuffering individuals should treat each other.

In Case (3) each self will be out to minimise the suffering of each other self because it regards this suffering as its own and it wishes to minimise its own suffering. The various selves of a suffering individual can be treated as one self with respect to minimising suffering. This is how velorians would contract for the selves of the one suffering individual to treat each other, that is, the velorians would contract that a self that believed itself to be identical with respect to suffering as a number of other selves would minimise suffering across all of those selves considered as a unit.

I will deal now with the first case. In this case I can talk of how individual L1 should treat L2 to LN rather than use the terminology of selves and the BN SN notation. Individual L1 SN is understood to be located in body B1.

The velorians do not know which of L1 to LN they are, therefore they will want to enter into whichever contract offers the best chance that they will minimise suffering no matter who they are. This means that the velorians will contract to ensure that, as far as possible, suffering is distributed equally between L1 to LN. This means that they will contract for L1 to LN to be bound by laws that have this effect. Offenders will be penalised to discourage the offense, because the offense is a case where the offender has benefitted himself at the cost of a greater than equal suffering to someone else.

The velorians, then, can enter into a contract about how individuals of the kind L1 to LN should treat each other. This contract supports the particular moral judgement that it is worse for the one individual to go suffering on and on than for the same amount of suffering to be distributed among a number of individuals, provided the selves of each individual concerned have a concept of themselves as members of the same suffering individual, and provided inequalities in suffering are due to the deliberate acts of other suffering individuals or are remediable by the deliberate acts of suffering individuals.

This particular moral judgement has been reached, however, without presupposing the existence of individuals (in a deep sense) or without presupposing their non-existence. It has been arrived at instead by investigating what contracts the velorians would negotiate to cover the case of selves that regarded their own suffering as that of the same

suffering individual. It was not even assumed that there actually are such selves, i.e. selves who believe members identified with earlier suffering selves. As it happens, there are such selves, namely the selves of nearly every adult human capable of participating in moral discussions. Consequently, the particular moral judgements that follow from the hypothetical contract are binding on all such humans, provided we adopt contract theory as our moral basis. Even though there are such selves, it doesn't follow that the metaphysical theory that there really are suffering individuals is true. It may be that there is no metaphysical basis to this belief. The complex theory could be true instead, with the consequence that any self that believed its suffering to be that of a suffering individual would just be mistaken. Whether such selves are mistaken or not, until they have been convinced that they are mistaken, they will continue to behave as if they were selves of one suffering individual, therefore the utilitarians can negotiate contracts about how the individuals which such selves believe themselves to constitute should treat each other.

Nonetheless, it may seem a weakness of contract theory that the force of the contract depends on whether or not a particular self remains convinced that it is a member of a suffering individual, especially if this belief cannot be rationally justified or can be shown to be false. After all, if it can be shown to be false, then rational selves will abandon this belief, thereby exempting themselves from the contract governing the behaviour of selves who conceive of themselves as suffering individuals. This means that they are no longer required by this contract to regard it as worse that one individual goes suffering on and on than that the same amount of suffering be divided amongst a number of "bodies".

The availability of the complex theory of personal identity, then, enables any particular individual B1 to claim exemption from the contract and its penalties on the grounds that he does not regard the suffering of future selves of his body as his suffering. The velorians, however, will not be convinced that he is entitled to exemption merely because he claims to be a non-suffering individual. They know that suffering individuals will not want other suffering individuals to pretend to be nonsuffering in order to be exempt from the contract and its penalties. It will have to be the case that the claimant really doesn't regard himself as a suffering individual. There will need to be evidence of this. If he is exempt merely because he claims to be non-suffering without support of sufficient evidence then this will lead to pretenders bringing about a reduction in their fair share of the equal distribution of suffering at the cost of someone else's increase. Velorians will not consent to this, consequently they will require that the claim to be nonsuffering be consistent with the way that the claimant has behaved or will behave. The claimant, for example, the human B1, will have to have acted in the past as if he was indifferent to his likely suffering in the present. There can be no indications that his past selves continuously planned to minimise suffering for his later selves. Just as importantly, he cannot appeal in the present to the consequences for him in the future of his being treated as a suffering individual because, if he is truly nonsuffering, it shouldn't worry him that there is some future self of his body that is penalised for the actions of this present self.

It is possible, however, that a suffering individual at T1 will become nonsuffering in the future. Should this occur, the velorians will not want him penalised as if he had remained a suffering individual. This is because the point of the penalty is twofold. Firstly, to deter the offender and others from breaking the law. Secondly, to prevent the offender getting away with a benefit he has gained at someone else's

expense. He cannot be permitted to have created an imbalance in the distribution of suffering such that he preserves the benefit he has hurt others to get. There is no point, then, in penalising someone who will not be deterred and who, in a sense, does not gain a benefit at someone else's expense because he is not the same individual as the offender. The offender did not plan that the benefit be his benefit. Admittedly, the benefit was enjoyed by someone, and that someone had no right to it, but neither is that someone guilty of improperly obtaining this benefit. The velorians will not want beings of this kind to be a threat to the lives, safety, property, etc. of suffering individuals, but they also will not want such a being treated as if he were responsible for his acts in the way suffering individuals are. They will want him to be able to live his life in a manner that maximises the satisfactions of each of his selves, while minimising his risk to other people. This will mean that he will become a cost to society. Just in case they turn out to be a nonsuffering individual they will require suffering individuals to contribute something to his maintenance and, possible, cure. Velorians would seek to return a nonsuffering individual to his former suffering condition if the former suffering individual's expressed wishes or actual behaviour indicated that this is what he would have wanted. Nonsuffering individuals are indifferent to the suffering of their future selves so they will not oppose such a contract.

As nonsuffering individuals are indifferent to the suffering of their future selves, they will be indifferent about which contracts exist to regulate the behaviour of their future selves, or the treatment given to their future selves. There will be no point, therefore, in the velorians negotiating special contracts to govern the behaviour of nonsuffering individuals because nonsuffering individuals are no more capable of honouring contracts than animals or very young children. Unlike animals

and young children, nonsuffering individuals fully understand what a contract is, and how it governs future behaviour. They will not oppose contracts governing how suffering individuals treat nonsuffering individuals because they do not regard the suffering such contracts may bring their future selves as, in any sense, their suffering.

We have seen, then, that the velorians will be able to negotiate contracts which govern the behaviour of both suffering and nonsuffering individuals, even given that they must assess the acceptability of these contracts from the viewpoints of both kinds of individual. In negotiating these contracts, they did not need to assume the truth or falsity of either theory, although they were only able to negotiate these contracts because they took into account the viewpoint of beings that believed themselves to be suffering individuals. The applicability of the contract, then, depends on the actual existence of beings that believe themselves to be suffering individuals. As I indicated earlier, most humans regard themselves in this way. The contract negotiated by the velorians, then, applies to most humans, that is, to all who regard themselves as suffering individuals. We have also seen that those human beings who do not regard themselves as suffering individuals will not oppose this contract, regardless of the truth or falsity of the simple or complex theories of personal identity.

Let us suppose, however, that the simple theory is false. This merely shows that there is no metaphysical entity enduring from T1 to T2 in which the experiences of S1 and S2 both adhere, or which subtends these experiences, or in which these experiences subsist. Must B1 S1, who has regarded itself as part of a suffering individual, now cease to do so? It must certainly cease to believe that there is such an individual, but it can, without an error of reason, continue to value the experiences of B1 S2 to B1 SN in a way that it does not value the experiences of B2 S1 to B2 SN.

There are, however, factors that reduce the arbitrariness of B1 S1's valuing more the avoidance of B1 S2's suffering than B1 S2's, even if the complex theory is true. After all, B1 S2's suffering occurs against a background of contents of consciousness continuous with B1 S1's but not with B2 S2's. The self B1 S1's visual, tactile, etc. experiences, its pains, tickles and other sensations, are all experienced from a perspective located in B1 as are B1 S2's, whereas B2 S2's are in B2. If the complex theory is true, then the presence of all these factors provide the basis for adopting the conventions that B1 S1 and B1 S2 are the same suffering individual but the important point is that they, or something similar, have to be present to give the convention a foothold. The more of them that disappear the less the selves of B1 are likely to regard themselves as the 'same' individual along any of the axes of individuality, e.g. if B1 S2 remembers none of B1 S1's experiences then B1 S2 will not regard itself as the same remembering individual as B1 S1, if B1 S2 no longer has desires of the same kind as B1 S1 then it will not regard itself as the same intending individual as B1 S1, and so on. The fewer types of individual that B1 S2 believes it and B1 S1 are both parts of, then the less likely B1 S2 is to regard itself as the same as B1 S1. Williams (1973) however, argues that a particular self B1 S1 could believe that future selves B1 S2 to B1 SN could share no memories, intentions, emotions, etc. and yet B1 S1 still believe itself to be the same suffering individual as B1 S2 to B1 SN. Even if Williams is not right about all humans, he only needs to be right about some for it to be the case that contract theory applies, whatever the metaphysical truth.

What, then, of the person in an actual philosophical argument who either holds the complex theory of personal identity or who is agnostic about the truth of either theory. Can this person claim exemption from the contract? He is, after all, one of the real people involved in an argument about how

we should act, and he is unconvinced of the truth of the simple theory. Have the rest of us abrogated our impartialist contractarian responsibility if we proceed to negotiate contracts that will govern his behaviour as well as ours because we negotiate as if we were suffering individuals, even though we have not provided metaphysical proof that there are any such things? The answer to this question is much the same as the answer given to how suffering individuals should treat nonsuffering ones. The difference here is that the person in the real situation is not pretending to be a nonsuffering individual, nor is he claiming to be a nonsuffering individual. Instead, he has philosophical doubts about the existence of suffering individuals. His philosophical doubts, however, can be either serious ones or speculative ones. If they are serious, this means he is not prepared to regard himself as an suffering individual until they are resolved. He then becomes a nonsuffering individual and the velorians have already negotiated a contract about how his suffering conflictants should treat him. If his doubts are speculative, in that he still regards himself as an suffering individual but wonders how he can prove that he really is one, then he is bound by the contract governing the behaviour of all beings that regard themselves as suffering individuals.

9. Continuity of Memory

It may be that the moral judgements made by impartialist contractarians presuppose a particular metaphysical theory about whatever it is that has B1 S1's memories being the same as whatever has B1 S2's memories, that is, that B1 S1 and B1 S2 are parts of the same remembering individual. I will briefly consider whether or not this is so.

Contract theory would treat memory much as it treated suffering. The velorians need to consider two types of beings, namely, those that believe themselves to be remembering individuals and those that believe themselves not to be. A self B1 S1 believes itself to be part of the same remembering individual L1 B1 S1 as the self B1 S2 when it believes that the contents of consciousness that B1 S2 will remember will actually be those of B1 S1, that is, when B1 S1 believes that there is a 'direct' relationship, to use Parfit's term, between its contents of consciousness and B1 S2's memories.

Let us consider a case where B1 S2 believes that the thoughts it remembers are those of B1 S1. Suppose, for example, that B1 S5 believes it remembers an argument which has the steps A1, A2, A3. It could be the case that there really has been one remembering individual remembering all the steps A1 to A3 or it could be the case that there is no such individual but merely a series of selves such that S1 had A1, S2 had A2, S3 had A3. In this latter case, B1 S4 thinks it remembers A1 but it is mistaken. This means that S4 was just born ex nihilo with a knowledge of A1 to A3. There seems no logical impossibility in this. Nonetheless, when the velorians came to consider what contracts to conclude from the viewpoint, firstly, of beings with genuine memories of the thoughts of past selves and, secondly, of beings with pseudo-memories of thoughts of past selves, they would have to treat both kinds of beings as the same for all practical purposes.

What matters is not that S4 really does remember S1's production of A1 as opposed to S4's having a pseudo-memory of this, but rather that S4 knows of S1's production of A1 by some means or other. If S4 doesn't know of S1's production of A1, then the task of the velorians does get difficult, especially if there are lots of these gaps in lots of these beings the velorian had hoped to bind by a contract. This, however, is not a meta-

physical issue but an empirical one. If there are such beings, then they are bound by a particular contract. If there are no such beings, then they are either bound by a different contract, or by no contract at all. Contract theory can tell us what our obligations are if the facts are a certain way. The remaining problem is to determine whether the facts are this way or not. There can be no serious doubt that it is a fact that there are beings who believe themselves to be remembering individuals, whether this belief is due to their having genuine or pseudo memories of the contents of consciousness of previous selves. There can be no serious, as opposed to speculative, doubt about this because we have just been engaging in precisely the situation where each of us was an S1 considering A1, then S2 considering A2 in the light of A1, and so on. There is an ad hominem here against anyone who claims serious doubts. Anyone who is truly a non-remembering individual could not be a velorian. He would lack the abilities necessary to conduct an argument on any matter, therefore he would not be capable of arguing about which contract should govern his own behaviour. In fact, as he was not capable of deductive reasoning, he would not be able even to conduct his life in accord with a contract.

10. Continuity of Intention

It may be that the moral judgements made by impartialist contractarianism presuppose a particular metaphysical theory about whatever it is that has B1 S1's intentions being the same as whatever it is that has B1 S2's, that is, that B1 S1 and B1 S2 are parts of the same intending individual.

If the complex theory is true then, Haksar claims moral theories are committed to treating selves of the same body in the same way as they treat selves of different bodies. He says,

'Under the complex view, there is not a radical moral difference between punishing John Doe for the crimes that John Doe committed earlier, and punishing John Doe's son for the crimes that John Doe committed earlier. True, John Doe's son is innocent of the crimes that his father committed but on the complex view John Doe is innocent of the crime that was committed by his earlier self.'
(1979, 114)

Haksar admits, however, that there can be a contractarian justification for punishing a later bodily self for the actions of an earlier bodily self.

He says,

'Suppose John Doe is asked to pay a library fine of \$5 because some overdue books borrowed by his earlier self. Now in such a case if John Doe were to disclaim responsibility for this debt which he admits was incurred by his earlier self, we could retort "Look, you have inherited so much from your earlier self. For instance, the large house that belonged to your earlier self now belongs to you. It is quite fair that if you acquire all his capital, you should pay his debts out of that capital. You cannot claim to be the same as John Doe for the purposes of acquiring all his capital, and at the same time deny that it is you who owe the library money".'
(1979, 118)

Just as the son can't claim his father's wealth without accepting responsibility for paying out his father's debts, so a later self of the same bodily individual can't claim an earlier self's wealth without also incurring its debts.

While Haksar agrees that this strategy provides the contractarian with a justified basis to punish later selves of the same bodily individual for the actions of earlier ones for some cases, it is by no means adequate.

He says,

'But now suppose John Doe were to commit a murder. Is it fair to severely punish his later self, say, by life imprisonment? Is this not a case where the capital that John Doe inherits from the earlier self is less than the debt that he is alleged to have inherited from his earlier self? And so, if in such a case we were to punish him severely, we would be

punishing a person who was, morally speaking, innocent. We may, of course, decide that for pragmatic reasons we have to punish such people, in spite of their moral innocence. But if so, how can we rule out, on principle, a system which allows us to punish a son for the crimes of his father?' (1979, 118)

We have already seen one way in which the velorians could legitimately draw up a contract that required a later self of the same bodily individual to be punished for the crimes of an earlier self, that is, when the contract governs how suffering individuals should treat each other. If L1 murders L2 in order to avoid some suffering, then self L1 B1 S2 is legitimately punished for L1 B1 S1's attempts to avoid some suffering for L1 at L2's expense. This contract, however, only covers the case of beings each of whom regards himself as a suffering individual located in one body. The velorians would have to negotiate quite a different contract for beings each of whom regarded himself as a suffering individual located in two bodies, either simultaneously or consecutively. While there may be such cases, we do not need to investigate how the velorians would deal with such cases unless such cases can be shown to make a difference to the cases that do interest us. The case of punishing the son for his father's crimes is not a case of the same suffering individual located in two bodies. It is rather the case of two sets of selves B1 S1 to B1 SN and B2 S1 to B2 SN, where each of B1 S1 to B1 SN regards itself as the same suffering individual as the others, and each of B2 S1 to B2 SN regards itself as the same suffering individual as the others, but none of B1 S1 to B1 SN regards itself as the same suffering individual as B2 S1 to B2 SN.

Velorians dealing with beings of this kind will treat the crimes of B1 S1 as the crimes of a quite distinct individual to the crimes performed by any of the selves B2 S1 to B2 SN. Instead they will treat them as the crimes of the same individual as the crimes of the selves B1 S2 to B1 SN, at least when B1 S1's motivation for these crimes is the avoidance of

suffering to any of B1 S2 to B1 SN.

Avoidance of suffering is not the only motivation behind B1 S1's actions that can serve to identify B1 S1 as the same individual as B1 S2 to B1 SN. Any intention whatsoever will do, as long as B1 S1 believes that whatever intends its actions is the same as whatever will intend the actions of some other self B1 SN. If B1 S1 believes this then I will say that it believes itself to be part of the same 'intending' individual as B1 SN.

The velorians will draw up contracts to govern the behaviour of intending individuals. The intentions of no individual will be given superior treatment to those of anyone else when it came to the distribution of costs and benefits consequent upon these intentions. Anyone performing an act intended to gain a benefit beyond his entitlement or to avoid a cost he is required to bear by a fair distribution of costs will be appropriately penalised. The velorians will want the contract to ensure that individuals are NOT able to advantage themselves at the expense of other individuals. This means they will want to be able to distinguish between criminal individuals who really do not believe that their various selves are selves of the same intending individual, and those who merely pretend this.

The velorians will have to consider a number of possible relationships that can exist between the intentions of a self B1 S1 and the intentions of some later self B1 SN. Each of these different possibilities may require a different kind of contract. These relationships include

- (1) B1 S1 believes itself to be a member of the same intending individual as B1 SN and it matters to B1 S1 that B1 SN's intentions are not hampered but, rather, are facilitated in their achievement by the actions of B1 S1.

- (2) Bl Sl believes itself to be a member of the same intending individual as Bl SN but it doesn't matter to Bl Sl what effects its actions have on the achievement Bl SN's intentions.
- (3) Bl Sl doesn't believe itself to be a member of the same intending individual as Bl SN but behaves as if it did.
- (4) Bl Sl doesn't believe itself to be a member of the same intending individual as Bl SN and doesn't care what effects its actions have on the achievement of Bl SN's intentions.

The kind of contract the velorian would negotiate for Case (1) will be the same as the contract they negotiate for Case (3). In either case, Bl Sl will perform actions at T1 intended to facilitate the achievement of Bl SN's intentions at TN, whatever those intentions may be. In R.M. Hare's terminology, (1981, 101-102), Bl Sl will display signs of evaluating the relative merits of his 'now-for-now' preferences and Bl SN's 'then-for-then' preferences. This will manifest itself in his modifying actions based on his own now-for-now preferences in order to balance their demands against Bl SN's then-for-then preferences. As long as Bl Sl reveals this kind of thinking in his choice of the actions open to him, he will be covered by the contract that the velorians draw up for intending individuals, whether Bl Sl believes himself to be the same intending individual as Bl SN or not. The point is that he thinks and acts as if he and Bl SN were members of the same intending individual. If Bl SN claims that he should not be punished for the actions of Bl Sl because Bl Sl did not believe himself to be the same intending individual as Bl SN, then Bl SN must also be able to show that Bl Sl's actions were not intended to facilitate the achievement of Bl SN's intentions in just the way Bl SN would have done had he been present in Bl at T1.

The contract, then, is not affected by B1 S1's belief that he is not the same intending individual as B1 SN if B1 S1 behaves exactly as B1 SN would have if he had been present in B1 at T1. Likewise for each of B1 S2 up to B1 SN himself.

This, however, doesn't show that the contract is independent of the truth or falsity of either the simple or complex theory of personal identity because exactly the same analysis could be offered for B1 S1 and the selves of some other body B2. The problem Haksar has posed is why we shouldn't punish the son for his father's crimes. What if the father behaved towards his son as we have described B1 S1 behaving towards B1 SN? If there is no metaphysical basis for saying that the father is a quite distinct individual from the son then why can't we adopt a convention and treat the father and the son as the one individual, as Parfit's complex theory seems to allow us to do?

What we need to do here is distinguish the conventional from the contingent. It is a purely contingent matter that we do not regard different bodily individuals as the same intending individual. It is a contingent matter because, were the facts to be different, we would treat two bodily individuals as the one intending individual. This, however, need not be merely a matter of convention. It is, rather, that different facts mean we have to pursue the same kinds of interests and concerns in different ways. If the facts were such that the father behaved towards his son as if he and his son were the same intending individual, then it would not be a mere matter of convention that we treated them as the same intending individual. It would be a requirement of our desire to ensure a fair distribution of costs and benefits between individuals as applied to the particular facts of the case.

As it turns out, it would be most unlikely that the facts of the case would ever be such as to warrant our treating the two bodily individuals as the one intending individual, given the kind of universe we find ourselves in. Given a different universe, such treatment might be warranted. The behaviour of father and son would have to be much more co-ordinated and inter-dependent than it is in our universe. It would have to be rather like the situation that John Wyndham describes in his novel, 'the Midwich Cuckoos', where there are a group of children, all of whom act as if they were merely different bodily limbs to the one mind. Whatever happens to one is known to the others. They never disagree with each other. Their interests never conflict, and so on. Something similar would have to be the case with the father and son after the father committed the crime for the son to be treated as the same intending individual as the father. In our kind of world this rarely if ever occurs. Not only does the father at TN perceive the world from a location in B1, but the father's self at TN, that is B1 SN has no 'direct' relation with the perceptions or intentions of the son's self at TN, that is, B2 SN. The self B1 SN cannot act through the body B2. It cannot control B2's arms or legs etc. as it can B1's. It cannot make B2 a vehicle of its will, nor does it believe it can, nor does it believe that B2 SN can make B1 a vehicle of B2 SN's will. It is also the case the B1 SN remembers the intentions of B1 S1 to B1 SN. It is as if he had been an observer to the contents of consciousness of each of these selves whereas he had never been an observer to the contents of consciousness of B2's selves. His memories not only involve his observation of these contents of consciousness but they also are accompanied by the belief that he was not merely an observer. While B1 S1 in fact had a particular set of intentions at T1, the self B1 SN believes that these intentions would have been different if he, B1 SN, had so desired. The self B1 SN can only believe this if he regards himself as the same intending individual as B1 S1. He does not

however believe it of B2 S1 or even of B2 SN. Likewise for B2 SN.

These, then, are all contingent facts that, when they obtain, will lead the velorians to negotiate a particular kind of contract in which a son will not be liable to be punished for his father's crimes. This is so whether or not there is really a metaphysical substratum linking the son's various selves and a quite different substratum linking the father's various selves. It is not, however, merely a matter of convention. Given that these are the facts of the situation, the velorians could not regard themselves as free to adopt some radically different criterion of personal identity.

The irrelevance of the metaphysical theories to the calculations of the velorians can be seen by considering a universe in which the facts warranted the father and son being treated as the same intending individual. The velorians will still want to ensure that innocent parties are not punished for the crimes of the father-son, even if the only beings available for consideration are all multiple-bodied. For example, the velorians have to take into account that they could be the innocent mother-sister. They will, therefore, still want to establish procedures by which to determine guilt and innocence, enforce penalties, and so on. They will not regard as acceptable a contract which advantages the father-son at the expense of other members of the community by refusing to punish the son for the crimes of the father on the grounds that they are distinct intending individuals when the facts have each of them behaving as the one intending individual. In this sense, the velorians are not free to adopt whatever convention they like about what is to count as an intending individual.

Nonetheless, what is to count as an intending individual may still be conventional in Parfit's sense in that there may be no metaphysical substratum linking the various selves of the father-son. Even if the velorians believed there to be no such substratum, they would still insist that the empirical facts of the case required father and son to be treated as the one intending individual. Their metaphysical theory would make no difference to the contract. Similarly, even if they believed that personal identity depended on a unifying metaphysical substratum, the facts of the case would lead them to postulate that there was such a substratum linking father and son. The metaphysical theory would be tailored to accommodate the empirical facts. Contract theory, then, need not presuppose any particular metaphysical theory of personal identity in order to generate substantive moral judgements about whom it is proper to punish for the commission of a crime. There can be non-utilitarian right-based constraints on who is to be punished. It can be the case that B1 S1 and B1 SN are linked in such a way that B1 S1 licenses us to punish B1 SN at TN for the murder B1 S1 committed at T1, whereas there is no similar link between B1 S1 and the selves of his son or his wife. Contract theory can refuse to sacrifice the individual for utilitarian ends without thereby begging any metaphysical questions. Even if the individual is a construct out of separate experiences, the contract theory does not have to focus on the separate experiences but can meaningfully negotiate contracts as if there were individuals who had these experiences. What is more to the point, in the cases I have examined the velorians would show a blatant disregard for the facts if their contract governed separate selves rather than intending individuals.

We have looked so far at two of the four possibilities I listed earlier, that is, we have looked at

- (1) Bl Sl believes itself to be a member of the same intending individual as Bl SN and it matters to Bl Sl that Bl SN's intentions are not hampered but, rather, are facilitated in their achievement by the actions of Bl Sl.
- (3) Bl Sl doesn't believe itself to be a member of the same intending individual as Bl SN but behaves as if it did.

The other two cases were

- (2) Bl Sl believes itself to be a member of the same intending individual as Bl SN but it doesn't matter to Bl Sl what effects its actions have on the achievement of Bl SN's intentions.
- (4) Bl Sl doesn't believe itself to be a member of the same intending individual as Bl SN and doesn't care what effects its actions have on the achievement of Bl SN's intentions.

Case (2) is likely to occur where Bl Sl believes that Bl SN will remember its contents of consciousness and those of the selves from Bl Sl to Bl SN. It will also believe that Bl SN could have changed what Bl Sl thought and did. The self Bl Sl, then, is to be treated in much the same way as the case where Bl Sl doesn't believe itself to be the same suffering individual as Bl SN. Bodily individuals like Bl Sl will not want to oppose a contract that treats Bl Sl to Bl SN as the same intending individual, because he is indifferent about whether or not Bl SN is punished for his acts. If he does oppose such a contract, then this is prima facie evidence that he does not really fall under Case (2).

The velorians, however, would be disinclined to punish Bl SN for the crimes of Bl Sl because, from the viewpoint of Bl SN, they would not want

to be punished for the crimes of a self that acted with such a blatant disregard for their interests. Also, intending individuals who might become non-intending individuals would be concerned to ensure that their later selves of their non-intending phase were not punished for the crime of an earlier phase. The velorians, then, would contract to do much as we do, that is, isolate non-intending individuals so that they cannot harm intending individuals, providing them as far as possible with a non-punishing environment.

Case (4) would be treated much as Case (2). In this case Bl SN's sense of injustice at being held responsible for Bl SI's crimes would be even greater than in Case (2).

These cases, then, are consistent with the points made against Haksar in my discussion of Cases (1) and (3).

Chapter 7

Epistemological Assumptions

1. Dworkin claims that

'many of the substantive claims that emerge from hypothetical contractarianism follow only because implicit and controversial assumptions are built in a non-obvious fashion into the structure of the theory.' (1975, 139)

In particular, Dworkin argues that Rawls' theory entails religious tolerance only because it assumes that 'one cannot arrive at justified true belief in religious matters'. (1975, 139)

In this chapter I will argue that impartialism does not make this assumption, even if Rawls does. In addition I will look at how the three impartialist axioms can be applied to epistemological questions to produce resolutions to conflicts of desires over what to believe, that is, if A wants B's belief that P to be suppressed, what kinds of reasons would be unacceptable.

2. Dworkin's Criticisms

While Rawls' contractors know nothing which enables them to identify their real-life counterparts, they do have knowledge of the general laws of science and human behaviour. Under such circumstances Rawls believes that they will contract to prefer religious tolerance to intolerance. Dworkin claims that this preference can only be deduced from the original position if it is assumed that the parties know or believe that 'one cannot arrive at justified true belief in religious matters'.

To show that this is so, he postulates that the parties behind the veil of ignorance all believe that redheads know the truth in religious matters. If this is so, then, Dworkin thinks the parties behind the veil of ignorance may well contract to prefer religious intolerance over tolerance. He says,

'If there were a truth, and it could be ascertained, would those in the original position who contemplated the possibility that they were holders of false views regard their integrity harmed by choosing that it should be suppressed? For while it may be biased to put one's own interests and desires in a favoured position, it is not bias to give preference to the truth or the highly probable.' (1975, 138)

3. The Relevance of Dworkin's Criticisms to Impartialist Theory

3.1 Dworkin's Argument

Dworkin postulates that the parties behind the Rawlsian veil of ignorance know that redheads have religious truth. If this is so, then would it not be rational, he asks, 'to choose these people as authorities with the right to suppress false views'. (1975, 138) Dworkin believes it would be rational to do so .

I will now look in more detail at his argument as quoted above.

3.2 Dworkin's Non-Sequitur

Dworkin gives us the first premiss of an argument, namely,

(1) It is not biased to give preference to the true.

From this premiss he derives the conclusion

(2) Redheads have the right to suppress false views.

This conclusion only seems to be possible if Dworkin treats 'giving preference to the true' as identical to 'suppressing false views'.

If this identity is made, then the argument would go through given the minor premiss (2) below

- (1) It is not biased to give preference to the true (suppress false views).
- (2) Redheads have the right to suppress whatever it is NOT biased to suppress.
therefore
- (3) Redheads have the right to suppress false views (give preference to the true).

It is by no means obvious, however, that giving preference to the true and suppressing the false are identical. A utilitarian, for example, who only values truth as a means to happiness may feel that someone who insists without argument that happiness-producing falsity must be suppressed while unhappiness-producing truth need not be thereby shows distinct signs of bias. Using impartialist arguments we can see further difficulties with this supposed identity.

3.3 A Provisional Impartialist Argument

In Dworkin's opinion, people behind the veil of ignorance who know that redheads have the truth, would contract to suppress all views that disagreed with redhead views. We can express the principle by which they would act as

- (4) Suppress all false (non redhead) views.

From an impartialist point of view, however, Dworkin's rider that the people behind the veil of ignorance know that redheads have the truth is a problem. The impartialist position is unconditional. All principles are open to dispute. Only if this is so are we in a position to say that the principles eventually contracted for are really person-neutral, respecting the integrity of each moral agent, considering all their interests equally. Nonetheless, we can discover whether or not people are impartialists by seeing how they would argue given common assumptions. If all parties are presumed to know something to be true, they will reach certain contracts about which symmetrical principles they will obey. A person who refuses to obey these, even though he accepts the common assumption, will not be impartialist. Those who do obey them will also obey whatever new contracts would eventuate were the common assumption shown to be unfounded. We can therefore operate on the basis of a provisional impartialism, remembering that the conclusions we reach only hold as long as our conditional assumption isn't challenged.

Given Dworkin's condition, then, the grounds on which we oppose the suppression of false views can't be that the freedom to express all views, including false ones, maximises the likelihood of the truth emerging. The truth already has emerged as far as all the parties are concerned. What other kinds of grounds might we have for opposing the suppression of false views?

3.4 The 'Error' Principle

Suppose there is a conflict of desires between A and B such that A wanted to suppress B's views just because he thinks them false. He is therefore operating on the principle 'Suppress all views that A believes false'. Impartialism, however, will only accept this

principle if it can be meaningfully symmetrified. Meaningfully symmetrified, it becomes, 'A is to be free to suppress whichever of B's view that A thinks false if B is free to suppress whichever of A's view B thinks false'. A and B would only accept this symmetrified principle, however, if they were coercionists. If A wanted NOT to be coerced by B more than he wanted to coerce B, then he would rather that those of B's views he considered false were not suppressed than run the risk that those of his views that B considered false were suppressed.

The impartialist axioms, then lead us to the conclusion that it is NOT a reason to suppress someone else's views that you think them false. This I will call the 'error' principle.

3.5 The 'De Gustibus' Principle

Suppose A merely wants to suppress B's views because he doesn't like them. He is operating on the principle 'Suppress any views I don't like'. Meaningfully symmetrified, this becomes, 'If A is to be free to suppress views that he doesn't like, then so is B'. A and B would only accept this symmetrified principle if they were coercionists. As a noncoercionist A would rather be free to express views that B doesn't like than to be free to suppress B's views that he doesn't like at the risk of having his own views suppressed instead.

The impartialist axioms, then, lead us to the conclusion that it is NOT a reason to suppress someone else's views that you don't like them. In fact, this can be extended to the conclusion that it is not a reason in itself to suppress anything anyone else wants to do that you don't like it. I call this the 'degustibus' principle.

3.6 The 'Self-Protection' Argument

If rational agents risked their desires not being satisfied as a result of false views not being suppressed, then they would all contract to suppress false views. Is it true, however, that not suppressing false views poses this threat?

Often, in the real situation, we risk frustrating our desires by acting on a false view we believe to be true. Dworkin, however, has ruled out this way of frustrating desires because everyone knows the truth. Any rational agent, then, who knows the truth yet acts in accord with the false either has acted irrationally or else is satisfying his desires by choosing to do what he knows is based on a false claim.

Someone who is satisfying his desires by acting on a claim he knows to be false will not contract to be bound to suppress false views, so an argument is needed to show why it is not acceptable under impartialist conditions to satisfy one's desires by acting on what one knows are false claims. This is the utilitarian consideration raised in 3.2. I will not provide such an argument here but merely point out that one is needed if the false is to be suppressed on the 'self-protection' ground, given Dworkin's condition.

A rational agent only acts irrationally by acting in accord with the false if thereby he frustrates his desires. By acting irrationally he acts in contradiction to what is definitionally required for him to be a rational agent, i.e. a maximiser of desire - satisfaction. In the example being discussed here our rational agent, M, is presumably contemplating a future case where he ceases to be rational. He knows the truth, he knows that acting in accord with it will

maximise his desire-satisfaction, yet he suffers irrational urges to act in accord with the false. It is by no means clear that suppressing the false will remove this problem unless the availability of false views is the cause of his irrational urges. This would be a rare case. It is only in this situation that a rational agent might contract to suppress the false for self-protection. Even in this kind of situation, however, a rational agent would impose stringent conditions on the circumstances under which he would allow the false to be suppressed. Firstly, he would require that those who suppressed the false for his benefit always had his agreement that their idea of the true coincided with his own. After all, the case we are discussing is one where everyone knows the truth but some people are irrationally acting contrary to it. The whole case changes if there is disagreement about what is true or false. Given that everyone agrees on the true, and continues to agree throughout on what is the truth, then rational agents will agree in advance that should they be caused to act irrationally by exposure to the false then either they will have the false suppressed or else they will have treatment that eliminates their irrational urge to act on the false.

It is clear, however, that this agreement depends entirely on there being no disagreement on the true. In the real situation this is a highly implausible assumption. Moreover, from the impartialist point of view, the absence of actual disagreement is irrelevant. Someone always could disagree. An unprovisional impartialist argument would need to accommodate this possibility. The fact that a person would abide by the principles contracted for under Dworkin's condition shows that they are impartialists but does not show that the principles they have contracted for are properly person-neutral.

3.7 Aquinas' Argument

Rawls believes that his two principles of justice are incompatible with Aquinas' argument for the death penalty for heretics. Aquinas claims that as corrupting the soul is far graver than counterfeiting money, yet counterfeiters are punished by death, then likewise heretics, who corrupt the soul, should be punished by death. Rawls rejects this argument because 'it cannot be established by modes of argument commonly recognised'.

Dworkin comments, 'On this view one is saved the labor of investigating specific details of Aquinas' position such as whether a heretic does corrupt the soul or whether, if he does so, this is bad or whether, if it is bad, it deserves death as a penalty'. (1975, 136)

Dworkin wishes the arguments we have discussed earlier to show that Rawls cannot so easily dismiss Aquinas' view. The problem, however, is that once we accept Dworkin's condition that everyone knows the truth, there are no heretics. There may be heresies, in that people may postulate false views for discussion purposes, but no heretics in the sense of people who believe the heresies to be true. The corrupting influence, therefore, has to be the heresy. The heresy, however, corrupts by making someone cease to believe what is true, which immediately alters Dworkin's conditions. Once someone doubts the redhead view, a contract needs to be established on how doubts are to be removed. While Dworkin's conditions continue to obtain, there seems to be no reason for banning the heresy, other than a fear that the mere existence of the heresy will lead to a situation where Dworkin's conditions don't obtain. As rational agents, however, people will not come to believe the heresy unless they have good reasons for thinking they will thereby maximise their desire-

satisfaction. Consequently, a rational agent has nothing to fear by the availability of the heresy. While he knows that redheads have the truth, he will not frustrate his desires by acting contrary to the truth, therefore heresy doesn't need to be suppressed. The only exception here is the one mentioned in the previous section. This, however, can be met if he agrees to have his access to the heresy suppressed rather than suppressing everyone's. As rational agents, other people will want minimum restriction on their freedom. Restricting their access to heresies just to prevent his irrationality would only be acceptable if it protected the maximising of their values. On the other hand, if he comes to think that redheads' views are false, he will not have the supposed heresies suppressed at all.

3.8 The 'Disapproval of Goals' Problem

It may be that you know a view to be true, for example, let's assume that promiscuity leads to eternal damnation. Even knowing this, you may like being promiscuous so much that you are prepared to suffer damnation. In a case like this you are not corrupted by false views. Rather, you know the truth about how to get eternal salvation, but your values are such that you value eternal salvation less than present pleasures. Here you have not frustrated your desires by a false belief, so being exposed to false views to the effect that promiscuity is compatible with salvation does not lead you to behave as you do. Other people may dislike your preference but, as rational agents, none of you will contract to prevent such a preference solely on the grounds of dislike. This is the degustibus principle we discussed earlier. It will have to be shown that even you are committed to a high order principle which rules out this preference.

4. Impartialist Epistemology

4.1 Introduction

So far we have operated as if Dworkin's condition applied. We have assumed that everyone knows that redheads had religious truth. Even with this assumption we have seen that it is very difficult to find a case where 'Suppress all false (non redhead) views' is acceptable under impartialist conditions. While it may not be biased to prefer the truth, in the sense of your preferring to believe what is true, it can be biased to suppress the false. Forcing people to believe only the true is prima facie a case of bias in favour of valuing truth over all values that might conflict with it, such as happiness or equality. It is only justified if valuing truth over all other values is what rational agents would contract to do when equal on all advantaging characteristics. Although I will not argue for it here, I believe rational agents would contract to encourage people to develop a disposition to prefer the true to the false but they would not actually suppress the false. Be this as it may, Dworkin has not shown that Rawls deduction of religious tolerance depends on scepticism about religious truth. My argument has shown that, even if everyone agreed on religious truth, there would be no need for the universal suppression of the false. Of course, if the truth is that all false views are corrupting, this would no longer hold, and it is possibly this kind of case that Dworkin most has in mind. As I indicated earlier, this kind of claim has its problems, in that Dworkin's conditions only continue to hold while the claim remains false, i.e. once someone is corrupted by a false belief it is no longer the case that everyone knows that redheads have religious truth. If Dworkin's case is not invalidated by this apparent paradox, then he has shown that Rawls' deduction of religious tolerance is a

provisional one. It is sound only if religious knowledge is not attainable. To render it unprovisional, Rawls needs to show that his contractors would contract to regard religious knowledge as unattainable. Rather than use Rawls' system, I will show that impartialist theory produces an epistemology which is sceptical about religious belief.

4.2 Intuition versus Knowledge

Rational agents will, as I indicated before, want as few restrictions as possible on their freedom to satisfy their desires, so they will need to be convinced that their desire-satisfaction is maximised by their adopting true beliefs rather than false ones. This, however, is not difficult to do in the area of beliefs about what is causally efficacious in maximising their desire-satisfaction. A person is more likely to maximise his desire-satisfaction if he acts in accord with true rather than false beliefs about the means of satisfying those desires. Consequently, any rational agent whatsoever will want to have true beliefs about the means of satisfying his desires, even if he is indifferent about the truth of any other beliefs he might hold. Any rational agent, then, will want to maximise the likelihood that he lives in a society which educates him in such a way as to maximise his chance of recognising this truth when he sees it, as well as maximising his disposition to find it. When people propose criteria of truth, then, he will want to arrive at the best criteria, at least as far as the true means to maximising his desire-satisfaction are concerned. These criteria will have to be relevant to the purpose of having true beliefs, that is, they will have to indicate the likelihood that a belief about the nature of the world matches, or corresponds, to how the world actually is. Unless the criteria satisfactorily do this, a rational agent as such need have no special interest in them.

Someone might propose that the truth of a belief be determined by his having an 'intuition'.

The word 'intuition' here covers faith, revelation, visions, dreams, mystical experiences, in fact, any perceptions of non-public objects. A rational agent will only accept this criterion if he thought there was good reason to believe that how he felt reality to be corresponded with how reality actually was. (The only reality we are concerned with here is the reality of means to satisfying desires). This will be true even if the agent advocates the criterion himself. He will want his own intuitions to give him accurate guidance to the true means for satisfying his desires. If the agent is not the advocate of the criteria he will want some way of determining when its advocate's intuition is a reliable guide. Just because the advocate feels that reality is of a particular kind is no guarantee that it is. Both parties, then, as rational agents, will want to be able to check the advocate's feelings against the world itself in some way. There appears to be only one plausible candidate for such a checking procedure, and this is observation of the world. Observation, however, will not be enough because the causes at work in the world may not be directly observable. Claims about such causes will need also to be checked against the world in some way, presumably by predicting observable consequences of there being such unobserved entities. This will mean proposing theories. Theories, however, will only be able to be the basis of action if we act as if their past predictive successes are a reliable guide to future predictive successes. By doing so, we are not making the unjustified assumption that the future will be like the past. Rather, we are adopting the only strategy available to us if we are to plan our lives so as to maximise our desire-satisfaction. If the universe is totally random

in that its past behaviour bears no relation to future behaviour, then there can be no person - independent basis for preferring one causal hypothesis to another.

These criteria combined suggest that a rational agent will prefer that theory which correctly predicts how the widest range of things will behave, as determined by repeatable observation. His main interest in modifying a theory will be to increase its predictive success rate.

In so far, then, as a person is interested in the true causes of public phenomena, he is bound to use empirical truth-criteria of the kind just described. Of course, there is still considerable controversy about empirical truth-criteria, as the continuing debate about the views of Hume, Winch and Feyerabend shows. From the impartialist point of view, however, this is not a problem. If those who deny the empirical criteria just described at the level of philosophical debate in fact use them in their choice of means to ends, then their challenge to these criteria as a basis for action is purely speculative. A provisional impartialist argument will suffice against them, because they hold to these criteria for all practical purposes. It is not good enough, for example, for a creationist to make a last ditch assault on the truth of evolutionary theory by appealing to unresolved disputes in the philosophy of science if his own everyday actions are most adequately interpreted as based on the very same criteria by which evolution is to be preferred to creation. All he can do is investigate the empirical evidence to see whether new evidence changes the picture.

What then of the person who doesn't just philosophically challenge the status of these truth-criteria but who refuses to live his life in accord with them, like Pyrrho, the Greek sceptic who refused to draw inductive conclusions from falling down a cliff, or being hit by a cart or bitten by a dog? Even Pyrrho, presumably, was inductivist enough not to starve himself to death as a result of scepticism about the efficacy of eating. There may, however, be someone sufficiently serious about their scepticism to do so. Such a person could not meaningfully participate in an argument about whether some arrangement is causally more effective in favouring one person's values rather than another's. The parties in such an argument must at least have those beliefs in common that are necessary to get such an argument going at all. They must believe communication is possible, that commitments can be honoured, and so on. This doesn't show that the sceptic's view is false but merely that those who engage in the kind of argument that is the subject of this paper have thereby committed themselves to personally regarding scepticism of this kind as false. This is all that is needed to proceed with an evaluation of the arguments that they then produce about what ought to be done in the realm of values.

How, then, does all this relate to religious tolerance? Suppose that A proposes that it is a good enough reason to suppress B's belief that A has an intuition that it is false. This means he advocates the principle 'Suppress whichever of B's beliefs that A has an intuition is false'. This is similar to the case where B's belief was to be suppressed because A believed it to be false but A is claiming his intuition as a source of knowledge of greater authority than a mere belief. Symmetrified, this becomes, 'A is to be free to suppress any of B's beliefs he intuitively is false and B is to be free

to suppress any of A's beliefs he intuitively to be false'. The outcome is the same as the Error Principle. Impartialists will NOT accept intuition that someone's belief is false as a reason to suppress it.

Unless religious beliefs, then, can be tested by criteria that do NOT turn out merely to be some person's private opinion, that is, unless religious beliefs are testable by public repeatable criteria, then rational agents under impartialist conditions will not agree to adopt them. Impartialist epistemology, then, seems to require religious tolerance. It does so, however, without assuming that one cannot have justified true belief in religious matters.

Chapter 8

Autonomy, Rationality and Perfectionism1. Introduction

In my earlier discussion of Rawls' 'publicity' constraint I claimed that the velorians would never agree to conflictants coercively removing each other's autonomy. Haksar, however, thinks that Rawls can only derive such a conclusion from the veil if he adopts certain perfectionist assumptions. In this chapter I will show that such a conclusion can be derived from the impartialist axioms without perfectionist assumptions.

Haksar argues for his claim by imagining a Brave New World in which people are effectively conditioned from a very young age to have internally consistent but low-level desires. This system has eradicated all serious mental illness. Mental health, Haksar believes, would be seen by the velorians as a high-ranking primary good that is as much under the control of the basic structure as self-respect is. The velorians can't object to such a society on the grounds that it restricts freedoms, that is, that it contravenes Rawls First Principle of Liberty. As Haksar says, 'The condition of the worst-off, from the contractarian standpoint, can be improved not just by giving them what they want and by not preventing them from getting what they want, but also by bringing them up in such a way that their aims and aspirations are easy to satisfy, that they do NOT set their aims high, and do not have mental and emotional troubles'.

(1979, 177)

Haksar, then, has posed the question of the inviolability of autonomy in terms of what the velorians would regard as being the worst-off members of a society. It is, he believes, plausible to propose that the velorians

would choose a society with little or NO autonomy and little or no mental illness to one in which each individual had a high degree of autonomy and a high likelihood of mental illness. His argument is, essentially, that the worst-off person in a brave new world will be happier than the worst-off individual in an autonomy-oriented society. This can be granted him. The truth of his conclusion, then, depends on whether happiness is the standard the velorians will adopt to determine who is the worst-off. It may be that they can be shown to be committed to the view that not being autonomous is worse than not being happy. The way that Haksar has posed his Brave New World challenge to Rawls biases the argument in favour of a 'happiness' solution. This is because he adopts a particular 'hedonistic' conception of what it is to be rational. I will now look at how this 'hedonistic' understanding of rationality leads him to too swift a victory over Rawls.

2. Rationality

The velorians, by definition, do not know which conflictants they are. As rational agents they will want who ever they turn out to be to satisfy their desires. This in no way entails that they must, as rational agents, want their desires to be such that they can be easily achieved or that they want their desires to be such that the chances are small of getting mentally distressed because these desires are not satisfied. Just because they may, if their desire is to be a painter as brilliant as Van Gogh, want this to be achieved with no more effort or suffering than necessary, it does NOT follow that, if there are desires that involve less effort and suffering than this one that they will prefer them. Rationality, as I will use the term, merely requires that, whatever desire we have, we want to achieve it in a way that leaves maximum opportunity for the satisfaction of other desires compatible with the satisfaction of this one. One of the possible obstacles to the maximum satisfaction of other goals is that we take more

time satisfying this one than we need, thereby leaving us to satisfy less of our other desires. This doesn't mean that all desires need to be satisfied quickly. Some of our desires, of their very nature, may involve the casual use of time. Nonetheless, the point remains that, as rational agents, we will not take more time than the nature of our desire requires. This is quite different from the requirement that our desires be such that they can be satisfied with little effort, or at little cost, or with little suffering, or in a quick time.

A rational person, in Rawls' terms, has a life-plan. This means there are a number of desires he wants satisfied. There will be an order of priority between these different desires. It may be, for example, that he wants one particular desire satisfied regardless of whatever happens to the others he would also like to satisfy, that is, a person's life-plan may be such that, although he would like desires D, E and F satisfied, he ranks them such that if D and F can only be satisfied if E is not, then he would rather E be satisfied and D and F not. A political dissident, for example, may want a secure job, enjoy his work, be happily married, have a contented home-life and good friends, while at the same time wanting that there be freedom of speech and other civil liberties in his nation. It may be the case that his nation is ruled by an oppressive regime so that his attempt to satisfy the last of these desires means that he frustrates all the others. Nonetheless, he may want the last more than all the others. As a rational man, then, he will sacrifice these other desires to the pursuit of civil liberties within his nation. Any judgement to the effect that he should not do this is not a judgement of his rationality but a disagreement either with his choice of desires or with the lexical ordering he placed on his desire. This is a disagreement in values. An individual's decisions about which desires are to be preferred to which, or what the lexical ordering between desires is to be, are not decisions

whose rationality is open to question, except in so far as the lexical ordering is self-defeating. Rationality, as a quality of agents, is not of itself to beg the question in favour of the desires of some agents over those of others. Rather, it merely requires of any particular agent that his actions can be integrated as either consistently achieving a particular lexical ordering of desires or else that he has changed his mind and he is now acting so as to consistently achieve a new lexical ordering of desires. It is perfectly possible, then, for a rational agent to want both to bring about civil liberties in his nation and avoid mental distress, yet so order his desires that he pursue civil liberties even if this means mental distress. Other rational agents may make the reverse choice. There is nothing in rationality itself that requires the one rather than the other. If someone so defines rationality that it does require one rather than the other, then this merely raises the question of why rationality in this sense should be desired. Rationality, in the sense I have given it, is not itself one of a number of kinds of desires, but merely a specification of what it is to want something more than something else.

If we are psychological hedonists, then we are likely to regard a person as rational only if he seeks to maximise the kinds of feelings concomitant upon the satisfaction of desires. This is because, as Frankena puts it, 'pleasure is the good in itself, it is what we all, ultimately at least, desire or aim at'. (1973, 85) This means that, over a life-time, the more desires one has that result in these feelings the more 'rational'

one has been. I will call these feelings 'happy feelings'. (1) The dissident I described earlier would be judged irrational by these criteria. He should have calculated the happy feelings that would have come from keeping his job, seeing his family, etc., weighed these against the happy feelings he will get as a political prisoner in a labour camp, then chosen whichever maximised his happy feelings. As a political prisoner, he will at least have happy feelings associated with his preservation of his self-image as a man of integrity, etc., but, qua happy feelings, this will be far outweighed by the happy feelings he would have got as a hypocrite, or as a moral coward. Rationality defined this way, however, is merely one kind of desire among a number of kinds of desire. There can be conflicts of desire between Group A who want to resolve conflicts of desires 'rationally' in this sense of 'rational' and Group B who don't. The velorians, as such, have no commitment either way. This conflict of desires is just one among many. They do not, however, give any special treatment to Group A just because Group A tries to persuasively define 'rationality' to make its criterion for resolving conflicts of desires the only 'rational' one in the desire-neutral sense of 'rationality'. Our political prisoner is quite rational in this desire-neutral sense in that he wants to be man of integrity more than he wants to maximise the amount of happy feelings he has in his life. To him, the mental distress that is associated with his decision is worthwhile. Haksar, then, only shows that the velorians would choose the Brave New World over an autonomy-oriented

(1) Barrow (1980) argues that 'one cannot logically accept' that 'a man might be happy without having the sense of enmeshment (the feeling whatever form it may take), that is happiness'. 'Enmeshment' is the notion that there is a fit between how things are and how one desires them to be. This feeling of enmeshment, then, can be achieved either by changing the world to fit a person's desire or changing his desires to fit the world. If it is rational to maximise this feeling, then it may be more rational to change one's desires than to change the world. Similar problems arise if we talk of 'preferences' rather than 'desires'. We can maximise the extent to which the preferences the person has over his life are satisfied by changing his preferences to more easily satisfied ones.

society if the velorians accept that the criterion of being worse-off is being mentally-distressed rather than autonomous. The political prisoner, however, would regard himself as worse-off if he were to be less mentally-distressed and less autonomous. Haksar's argument, therefore, begs the point at issue. What he needs to do is to show that the velorians would believe a person to be worse-off mentally-distressed but autonomous than not be mentally-distressed or autonomous even if the velorians did not already value lack of mental distress more than autonomy. In other words, he cannot attribute to the velorians a concept of rationality that is really a value-judgement in disguise. As rational agents, the velorians are to be construed as out to maximise the satisfaction of their desires. This, however, does not mean that they are only interested in maximising the happy feelings that come from having desires satisfied. Rather, they are concerned with satisfying desires in the sense of having whatever state of affairs is the object of the desire come into existence. It may be that the state of affairs that they want to come into existence more than any other is the one in which they maximise their happy feelings. If so they are psychological hedonists of the kind I have described above. It may be, however, that the state of affairs they most want to see come into existence is a society with political and civil liberties, or an autonomous society or the destruction of their enemies regardless of the cost to themselves. Rationality merely requires that they take whatever course of action is most effective in achieving these desires and any other compatible with them. It does not require them to abandon these desires and replace them by others.

3. Autonomy

We have seen that rationality itself does not require the velorians to prefer mental health to autonomy. It could still be the case, however, that the impartialist axioms are not strong enough to generate the

conclusions that an autonomous society is to be preferred to a Brave New World. Haksar claims that Rawls' principles only become strong enough if they are combined with perfectionist assumptions. It could be the case that impartialism, too, needs perfectionist assumptions. I will now consider whether this is so.

Person A will be said to remove B's autonomy when A controls B's will in the sense that, whenever there is a conflict of desires between A and B, A can bring B to want, or to think he wants, whatever A wants. (1) If B is not autonomous, then B is not capable of resisting A's will. It is not just that A always wins in any conflict between A and B, but that B cannot control what he himself does. B may be aware of this. He may know that what he wants is to do X yet find his will controlled in such a way that he finds himself doing Y which he does NOT want to do, and does NOT intend to do. It is as if he watches himself doing Y rather than being the agent who does Y. Alternatively, B may not even be aware that his will is controlled by A. He may always think that he wants to do X, when the cause of his wanting to do X is that A wants him to.

If A had the means by which to deprive B of his autonomy, this would be a very satisfactory way, from A's point of view, of resolving conflicts of desire between himself and B. A, then, would be prescribing the principle that, when there is a conflict of desires between A and B, then B's will be under A's control. Symmetrified, this becomes the principle that A be

(1) Benn (1976, 116) distinguishes the inner-impelled or 'autarchic' person from the 'heterarchic' person. Heterarchic persons include those 'acting under hypnosis, or brain-washed, or unable to contemplate disobeying an authoritarian parent'. Haksar's challenge is that Rawls' principles don't rule out heterarchy.

free to coerce B by controlling his will and B be free to coerce A by controlling his will.

As the velorians J and K want not to be coerced more than they want to coerce, they will reject this proposal. We have, then, another impartialist principle, namely, the 'autonomy' principle. It is the principle that A NOT be free to coerce B by controlling his will and B not be free to coerce A by controlling his will. In other words, within impartialism it is wrong for either party to resolve a conflict of desires by depriving the other of his autonomy.

This still leaves it open for A to take someone who lacks autonomy but would later develop it naturally, for example, a new-born child, and to bring up the child so that its will is subservient to his. A has not deprived the child of its autonomy.

Nonetheless, he shows a lack of respect for autonomy that most contractarian philosophers would regard as totally reprehensible. Can this reprehensibility be deduced from the impartialist axioms?

The velorians, as we saw earlier, are to consider conflicts of desires between parties by putting themselves in the place of the participants of the conflict. In the kind of case I have proposed, A is an autonomous rational agent, that is, a person, whereas B is not. B has a potential for autonomy but, as we saw in discussing abortion, potential is not a factor the velorians would take into account as it doesn't affect the actual desires of the participant with the unrealised potential. As a consequence, the velorians would not be able to conclude that it would be wrong of A to deny B the opportunity to realise his potential for autonomy. This kind of conflict of desires, however, misrepresents the actual

argument in everyday life about how A should behave. The argument is NOT between the person A and the non-person B. By definition, if B is a non-person then B is incapable of argument. Instead, the argument would be between the persons A and B about how A might treat a non-person C or even about how A should treat B should B become a non-person. In this case, however, the only aspect of non-personhood we are concerned with is lack of autonomy.

I will consider first the case of an argument between A and B about how A should treat B if B were to lose his autonomy temporarily. While B lacks autonomy, he may lack certain kinds of desires such that A does NOT frustrate any of B's desires by making B permanently non-autonomous, even though A does this solely to make B a slave to A's will. A, then, wants to turn B into a permanent state of non-autonomy should ever the situation arise where B accidentally becomes temporarily non-autonomous. It could be that B does not value his autonomy enough to object. If so, there is no conflict of desires between A and B. It is far more probable, however, that B bitterly resents this attitude of A's. B would not only want A NOT to take advantage of his temporary loss of autonomy in this way but would want A to take steps to remove it.

How are J and K to deal with this conflict? A is prescribing the principle 'A is to be free to keep B permanently non-autonomous whenever B accidentally loses enough of his autonomy to no longer desire to retain it'.

Symmetrified, however, this means that B is to be free to do the same to A. In this particular case, the noncoercionist axiom does not apply as the situation is so set up that neither party can be said to be actually forcing the other. This means that J and K could adopt either of two principles, namely (1) that both A and B be free to keep the other permanently non-autonomous provided they didn't make him non-autonomous in the

first place or (2) that neither of them have this freedom. Which J and K choose will depend on the kinds of people the conflictants A and B are. If A and B each want not to be kept non-autonomous more than they want to keep the other non-autonomous, they will adopt (2). If, however, they both want to keep the other non-autonomous more than they want NOT to be kept autonomous, then they will go for (1).

Choosing (2), like choosing the noncoercionist position, gives a determinate answer as to who is to be kept in a non-autonomous state by whom. The answer, given (2), is no-one. If A and B, however, favour (1) then they are agreeing that, whichever of them is to be kept permanently non-autonomous is a matter of luck, that is, whichever of them is unlucky enough to be accidentally made non-autonomous first. If A and B both accept (1), then impartialism has no means directly deviable from its axioms for excluding (1).

If A accepts (1) and B accepts (2) then A just has to do whatever he can to ensure that he doesn't accidentally lose his autonomy first. It may even be the case that impartialism allows an even stronger conclusion to be drawn than this. It may, for example, be the case that a person A who wants to keep B non-autonomous when this happens accidentally is the kind of person who would deliberately make B non-autonomous. J and K have agreed already that this is not to occur. If allowing A to keep B non-autonomous increases the risk that A will force B to be non-autonomous or provides A with the opportunity to contrive accidents to B that put B in his power, then J and K will be loath to allow A to keep B non-autonomous even when this happens accidentally. They will try to ensure that B is returned to autonomy as soon as possible to eliminate any temptation A may have to render B non-autonomous on purpose. This, however, is what Haksar would call a 'pragmatic' argument against allowing A to keep B non-

autonomous. It is like the argument Haksar gives for treating congenital human idiots as members of the egalitarian club. Whether this pragmatic argument works or NOT is not the issue however. The issue is whether impartialism presupposes autonomy or whether it can only prefer autonomy to the Brave New World by imparting perfectionist assumptions. As we have seen, impartialism can derive the exclusion of coercing people into non-autonomy from its three axioms. It can also deduce the exclusion of non-autonomy from its axioms for any conflictants who want not to be kept non-autonomous more than they want to keep other people non-autonomous. In addition, there may be pragmatic arguments of the kind sketched above that exclude even those who want to keep other people non-autonomous more than they want NOT to be kept non-autonomous themselves.

This still leaves the case of the new-born infant. Should a baby with the potential to be an autonomous agent be denied the realisation of this autonomy in order to become the compliant tool of already autonomous agents. Would we ever be justified in operating on children to remove their capacity for autonomy or even of manipulating genetic codes to breed a race of willing slaves? Again, the argument here has to be between the two autonomous agents A and B about how infants should be treated. This case differs from a temporary loss of autonomy in that infants have never been autonomous. It is not a case of A's being concerned to ensure that he be returned to autonomy as soon as possible. As an already autonomous agent he has no concern to ensure that he becomes autonomous in the first place, and those who are not yet autonomous have no desire of theirs for autonomy frustrated if their potential is never realised.

Impartialism, I think, can only use pragmatic arguments in favour of the view that infants shouldn't be deliberately deprived of their autonomy. As we have seen, J and K would agree that neither A nor B should deprive

the other of his autonomy. J and K would agree to whatever non-coercive measures were available that were likely to reduce the chances of A's depriving B of his autonomy. One of these would be to create a climate in society in which autonomy in others was valued. If A values B's autonomy, then A is less likely to regret the fact that no accident has yet befallen B to deprive B of his autonomy so that A can take advantage of this to make B a mere instrument of A's will. In fact, it is this tendency to treat other people as a mere instrument of one's own will if one doesn't value their autonomy that would be of greatest concern to J and K given that they have agreed that neither A nor B is to coercively deprive the other of his autonomy. If A is hopeful that B will lose his autonomy accidentally then it is difficult to see how A can have a sincere desire not to deprive B of his autonomy coercively. It would seem that A has no real respect for B's autonomy but merely is constrained by laws punishing the coercive deprivation of autonomy. B's autonomy is best protected from coercive deprivation by A if A genuinely wants B to be autonomous. Such a genuine desire would carry over into the desire that B regain his autonomy if he accidentally loses it, that is, it carries over into a disposition to want B's potential for autonomy restored if the actuality is temporarily lost. A concern for the actualisation of B's potential autonomy shows that A doesn't regard B as a mere instrument of A's will but values B's independence of A's will. It also shows that A is concerned that B has desires and wishes of his own that come into occasional conflict with A's and which A wants to have resolved by impartialist processes. It shows that A treats B as an agent whose life-plan is worth equal consideration with his own. He sees B as someone with whom he must enter into argument about how to resolve conflicts of desires. He shows respect for B's personhood. All of these qualities of character are qualities J and K will want A and B to have towards each other as re-inforcements of the agreement between J and K that A and B are NOT to

coercively deprive each other of their autonomy. If A sees infants with this potential for autonomy as mere instruments of his will then he shows the kind of lack of respect for autonomy itself that is the cause of concern to J and K in the case where he seems to be anxious for B to lose his autonomy in an accident. A disposition to value actualisation of the potential for autonomy in people who have temporarily lost it will be difficult to separate from a disposition to value the actualisation of autonomy per se. The disposition to take advantage of an infant's helplessness to permanently frustrate its potential for autonomy so that it can be an instrument of one's will is difficult to distinguish from the disposition to permanently frustrate anyone's potential for autonomy to make them an instrument of one's will.

It would appear, then, that there is a theory of the virtues that can be derived from the impartialist axioms. I have only sketched here some of the dispositions the velorians would encourage the conflictants to develop in each other. What the velorians would regard as crucial, however, is that the conflictants develop attitudes that foster their commitment to the use of impartialism itself as the method for resolving conflicts of desires. One way of resolving conflicts of desires is by deprivation of autonomy. The velorians, however, agree this should NOT be done. It is important, then, that A and B want people to be autonomous, that they want the principles that govern conflicts of desires to be the ones that autonomous agents would agree to under impartialist conditions. They would discourage the tendency to turn others into mere instruments of one's own will. As a consequence, the potential for autonomy, as well as its actualisation, would be valued because those who did NOT value it would show a manipulative inclination even towards those in whom it was actualised in that they seem to be playing a waiting game for the moment when this actual autonomy is temporarily lost.

These, then, are the kinds of considerations that can be employed within impartialism to arrive at the conclusions that autonomy is preferable to the kind of Brave New World that Haksar describes. Perfectionist assumptions have not been necessary. Nonetheless, this whole case for the value of autonomy rests on there being two autonomous agents A and B in the first place to argue about these issues. Impartialism is unable to show that a person A who finds himself in the situation where he does not need to defend his position on autonomy to another autonomous agent must conclude that he should foster the autonomy of potentially autonomous beings rather than develop them all as non-autonomous instruments of his will.

Impartialism does assume that there is at least one conflict of desires between autonomous agents capable of argument about how to resolve this conflict. It does NOT set out to show that there should be such a situation but, rather, is concerned with how there can be an unforced resolution of the conflict when such a situation arises. The conclusion that there be at least two autonomous agents so that there are conflicts of desires that can be resolved by unforced agreement is not something that can be deduced from impartialism. To value this, then, may well be perfectionist in Haksar's sense. There is, I believe, no argument that would show that a person should value impartialism if the situation did NOT exist where he had to justify this view to some other autonomous agent. If A was the sole surviving autonomous agent, then I do NOT believe there is any argument he could put to himself that would require him rationally to want to bring other autonomous agents into existence. As it is, however, this situation does NOT obtain. We do have to justify our beliefs about what ought to be done to other autonomous agents. If we are to reach unforced agreement about such beliefs, we are thereby committed to valuing autonomy.

Chapter 9

The First Principle of Justice1. Introduction

I have shown that the veil of ignorance does not make the metaphysical end ethical assumptions that a number of writers have accused it of. I will now see whether it produces the same 'principles of justice' when combined with the other impartialist axioms as Rawls thinks it does when contained with the other demands of his theory. I will begin with Rawls' First Principle of Justice.

Rawls states this principle as follows:

'Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.'
(1972, 302)

2. 'Total System' of Liberties

Barry queries what Rawls means by a 'total system of liberties'. He notes that, in 'Justice and Fairness' Rawls had talked of 'equal liberties'.

His answer is that

'Rawls thinks of the various liberties as capable of occurring (within limits) independently of one another, and suggests that, if maximising one is inconsistent with maximising the others, different proportions of the various liberties should be combined in such a way as to make the "total system" one of as much liberty, equally distributed, as possible.'
(1972, 34)

Barry notes that Rawls 'gives little usable guidance about the way to aggregate the different liberties so as to arrive at an estimate of the total amount of liberty generated by alternative combinations of these different liberties'. (1972, 34)

Impartialism will treat liberties independently, as Rawls also wishes to do. What we have with impartialism is a machinery for looking at particular cases of where A might wish to coerce B to do X. As impartialism only deals with people for whom the power (or freedom) to coerce others is less important than the power (freedom) not to be coerced by others, then there will be cases where J will consent to whichever conflictant he turns out to be being coerced into not coercing others, or being denied the power to coerce others, or being punished when he does coerce others, that is, J will consent to whichever conflictant he is not being free to engage in certain kinds of coercion whereas other conflictants are free to engage in the kinds of coercion necessary to prevent other kinds of coercion. For example, if one group of conflictants wants to kill members of another without their consent, then, as A wants to kill B less than not be killed by B, and B wants to kill A less than not be killed by A, then, even if J turns out to be A who wants to kill B, J will have agreed that A be denied this power (freedom). This will generate a number of freedoms, i.e., A will be free from B's killing him without his consent, from B's taking his property without his consent, from libelling him, from suppressing his freedom of speech, and so on. It is unlikely that these freedoms need to be ranked in that it is not clear how they might conflict with each other. It is possible, however, that this machinery will produce duties and rights rather than just freedom, and that these will conflict with freedoms, e.g., the ability to pay taxes to support the poor unemployed. The obverse sides of those duties can be regarded as freedom, e.g., the freedom not to starve or whatever, i.e., freedom from fear of various kinds. If these are the outcome of the machinery, then there may need to be a ranking of freedoms.

3. Basic Liberties and Equal Liberty of Conscience

Barry raises the question of what exactly are the basic liberties. He looks at the candidates Rawls proposes. He first mentions political liberty but puts it aside for later discussion. He then looks at 'the Rule of Law' but merely notes that the content of the rule of law ('ought implies can', 'similar cases be treated similarly', 'no offence without a law', etc.) apply to the administration of the law, not to the content of the law, i.e., it is 'justice as regularity'. Religious freedoms, for example, could be denied yet this be compatible with the rule of law. He then moves onto how Rawls deals with the basic liberty that Rawls calls 'Equal Liberty of Conscience' which does give a 'liberal' content to laws. Barry notes that Rawls only deals with a small component of the liberal content of law, namely, freedom of conscience, i.e., doing what you believe is right. Rawls ignores the freedom (or right) to do what you want to do, even if it's not a matter of religious duty. (1972, Section 33) Rawls later (1972, 450) does assume that he has shown that his two principles enjoin freedom of action in sexual and other matters where these affect the states' interest in public order and security' (1972, 212). Barry, however, sees Rawls arguments as only concerned with 'conscience', what you think is right, rather than with non-conscientious freedoms to do as you want. In fact, Barry thinks Rawls' arguments for freedom of conscience can't be generalised to non-conscientious cases precisely because they rest on the conscientiousness of the act. Rawls, however, is concerned that there be sexual freedom, say, because this is where his theory triumphs over utilitarianism. Utilitarianism may deny the right to sexual preference for reasons other than state security if enough people are offended by a preference. Rawls doesn't want to allow this.

Barry thinks that Rawls can only extend his argument for liberty of conscience to 'the rest of the items in the liberal catalogue such as

sexual liberty' if the utilitarians can assume that 'any desire the fulfilment of which does not do any injury to others is liable to be felt so strongly that it would be irrational to permit the possibility that its fulfilment might be hampered if it was abhorred sufficiently strongly by enough people in the society'. (1972, 39)

Barry has two objections to Rawls assuming this.

- (1) It is factually implausible that conflictants are like this.
- (2) We do allow desires to be fulfilled that cause injury to others if those who desire it desire it strongly enough and those who are injured are only mildly injured. What Barry is getting at here is that Rawls wants to rule out A's stopping B's sexual practice just because A wants to stop it a lot more than B wants to keep it. Barry thinks this makes 'the detailed adjustment of conflicting demands' impossible and therefore rational men will not agree with Rawls in ruling it out.

Let's look at all this from an impartialist perspective. We need to look first at Rawls' reasons for liberty of conscience. His argument is that the utilitarians 'cannot take chances with their liberty by permitting the dominant religious or moral doctrines to persecute or to suppress others if it wishes'. This is because 'to gamble in this way would show that one did not take one's religious or moral convictions seriously, or highly value the liberty to examine one's beliefs'. (1972, 207)

Rawls' account of this is oddly put. The liberal conclusion is reached in a more straightforward manner by impartialism. If J doesn't know whether he is A or B, but does know that both A and B want to coerce less than they want not to be coerced, then what will he agree to when A wants

to coerce B with respect to B's religious beliefs or sexual practices? As A wants to coerce B less than he wants not to be coerced by B, and symmetrically for B, then J will agree to A's not having sufficient power to coerce B, or B to coerce A, and to there being a system which has sufficient power to prevent A's coercing B or vice versa in either of these areas. The fact that one is a matter of conscience and the other is not doesn't seem to affect the situation. Rawls is forced into his argument by his failure to recognise the difference between noncoercionists and coercionists. If J doesn't know whether he is A or B, but he knows that both A and B want to coerce the other more than not be coerced by him, then J will agree to take the risk that, if he's A he is coerced by B into B's religion just so that he can have the chance that he can coerce B. Now, Rawls wants to argue that this is a case of not taking your religious convictions seriously. The problem here is that he hasn't analysed closely enough what the conviction is. If my conviction is that I should practise my own religion and make others follow it, too, then, in order to be serious about the last bit I've got to risk (behind the veil) that others will suppress me, otherwise I can't get agreement to suppress them. But if I'm serious about suppressing them, then I have to agree to a rule which leaves me open to being suppressed. It's only if I'm serious about practising my religion and not serious about suppressing others that I will agree to religious tolerance. But there's nothing special about the religious case here. The same logic would apply if I was more serious about suppressing others' sexual practices than if I was about being guaranteed the freedom to indulge in my own.

What we have here is a generalised finding of impartialism that I have called the 'de gustibus' rule. Given impartialism, J will agree that A not have the power to stop B from doing X just because A doesn't like B's doing X, or anything that amounts to A's not liking B's doing X

such as A's having unfavourable attitudes to B's doing X.

Instead, J will always agree to A being free to do X unless A's doing X involves A's coercing B and J would rather that B be free of coercion than that A be free to do X.

Let's return now to Barry's specific criticisms of Rawls. Barry thinks that Rawls can only get equal liberty of sexual preference along the same line of argument as used for equal liberty of conscience if people wanted to engage in their own sexual preferences (without injury to others) so strongly that they wouldn't allow it to be restricted no matter how strongly other people abhorred it.

If this psychological generalisation held, then J would have to calculate as if

- (1) he were A when A does X to B
- (2) he were B when A does X to B
- (3) he were B when B does X to A
- (4) he were A when B does X to A.

Now, if X is 'suppressing any desire the fulfilment of which does not do any injury to others' then, if B desires not to be suppressed in (2) more strongly than A wishes to suppress in (1) then J will conclude that A shouldn't do X to B and, by symmetry, that B shouldn't do X to A in (3) and (4). This will only hold for all cases where the desire NOT to be suppressed is stronger than the desire to suppress, that is, it will only hold in those cases where the conflictants are what I have called 'noncoercionist' in their disposition. Not all conflictants, however, are noncoercionist. This is undoubtedly a problem for impartialism if Barry

is right that tolerance of sexual lifestyles requires that the desire to suppress in real life is weaker than the desire not to be suppressed. This, however, is a dubious empirical claim. It may well be that puritanism is often more powerfully felt than hedonism. Impartialism, however, does not rest on any factual findings that conflictants actually are noncoercionists. They may in fact all be coercionists. If this is the case then they can only settle their conflicts by force. Being a noncoercionist is a pre-requisite to settling disputes by non-coercive means. Barry's point, then, could not be used as a criticism of impartialism even if it could be of 'liberal' theories like Rawls contractarian one.

Barry's second point seems to be a case of affirming the consequent. He says that 'There seems no a priori reason to suppose that if someone does have a desire whose fulfilment is of absolutely central importance to his life, it will necessarily be one whose fulfilment cannot possibly cause any "injury" to others'. While this claim seems true enough, the impartialist case doesn't require that all 'desires whose fulfilment is of absolutely central importance' to a person's life be given the same status as conscientious beliefs, but merely those that do not cause injury to others have this status. Barry's claim may be that there are few, if any, such desires. Even if this is true, it doesn't change the logic of the impartialist argument. It merely means that there will be few desires of central importance to people's lives that have the same status as conscientious beliefs. In fact, the impartialist position doesn't allow conscientious beliefs to be given special constitutional protection if they cause 'sufficient' injury to others. As we have seen, however, it's not that the desire be one whose fulfilment causes no injury to others but, rather, that the desire be such that, if one wants not to be coerced more than one wants to coerce, then one will not agree to the fulfilment

of this kind of desire being suppressed. If one wants not to be coerced more than one wants to coerce, then one will want not to have others restrict one's sexual preferences more than one will want to restrict other people's sexual preferences. This conclusion doesn't depend on the sexual preference one has not harming others. Admittedly, if I am a sadist who wishes to kill other people for my sexual pleasure, then I clearly harm them, but my freedom to do this is ruled out by principles dealing with killing others, or physically mutilating others or deliberately causing others bodily pain, not by the principle of free sexual preference choice. Nonetheless, there is a lexical relationship between the principle of freedom of sexual preference and principles outlawing murder, torture etc. This means that the principle of sexual preference doesn't have to be explicit about the kinds of injuries it allows. It allows none that are ruled out by other principles higher than it in the total system of liberties. This finding, too, can be generalised. Any 'equal liberty of conscience' principle or 'no injury type of desire' freedom will be restricted by the limits placed on coercing others. Individual J will allow person A freedom of conscience provided A's freedom of conscience is not itself a coercion of B, and likewise with A's sexual preferences. Of course, some coercion of B is allowed, for example, A's freedom of conscience can entail the coercion of B's freedom of conscience if B's is such as to coerce A's, and so on. This, however, must follow the pattern established earlier. Individual J knows that A and B want not to be coerced more than they want to coerce. Consequently, if A wants to coerce B into a religion or a sexual preference, but this is only acceptable if B is free to coerce A, then J will agree that, even if as A he wants to coerce B, he not be free to do so, and likewise for B.

It is worthwhile exploring here in some detail the rest of Barry's remarks in this paragraph because they reveal very well how easy it is to mis-

understand what the liberal is doing.

Barry says, with respect to his two objections to Rawls' 'equal liberty of conscience', that 'the implications of both objections are, I think, the same: that the need for a "calculus of social interests" cannot be swept away, as Rawls supposes, by importing heroic assumptions about a universally valid hierarchy of human goals into the original position'. (1973, 39)

'Rational men will not be so willing to make impossible the detailed adjustment of conflicting demands. In judging between the desire of A to do something and the desire of B to have him prevented by law from doing it, they will not rule out in advance the relevance of the question "How much does A want to do it and how much does B want to stop him?" ' (1973, 39)

Now, there are a number of ways of making detailed adjustments of conflicting demands. One is by force. Barry presumably doesn't mean this. As we have seen, however, if A wants to coerce B with respect to X more than he wants not to be coerced by B with respect to some action Y, then force is the only way of adjusting conflicting demands that can be used with A. One, therefore, has to take into account the degree to which A wants to do something and the degree to which B wants to stop him doing it. Once, however, we are within the impartialist framework, what is relevant is not how much A wants to do X or how much B wants to stop him doing X. The relationship between the strength of A's desire and the strength of B's desire is not the issue at all. The issue is whether B wants to stop A's doing X more than he wants A not to stop him doing Y. Let us put numbers to these strengths of desire. It's not a question of A's desire to do X being 1000 units whereas B's desire to stop A's doing X is 2000 units. If it were merely this relative strength of desire then whether or not A is free to do X depends on whether he or B has the stronger desire. It is, instead, a question of B's desire to stop A's doing X

being 2000 but his desire that A not stop him doing Y is 4000. If it is, then the fact that his desire to stop A is twice the strength of A's desire to do X is not germane to whether or not he be free to stop A.

Hare illustrates how a utilitarian concerned with impartiality can reason in a similar manner to the impartialist and yet still miss the boat.

'I have to give a bun either to Jones or to Smith, and Jones wants it more than Smith. Let us call Jones-with-bun J1, Smith-without-bun S1, Jones-without-bun J2, and Smith-without-bun S2. Jones prefers J1 to J2 more than Smith prefers S2 to S1. Suppose then that I put myself in their two positions in the two outcomes, four positions in all. My choice is going to be between J1 plus S1 (Jones-with-bun plus Smith-without-bun) and J2 plus S1 (Jones-without-bun plus Smith-without-bun). It is not necessary to imagine myself occupying the four positions simultaneously; that would be asking too much. We might follow a suggestion of C.I. Lewis (1964, 546f) and suppose that I have a choice between occupying J1 and S1 in random order and occupying J2 and S2 in random order. Clearly, given the assumed strengths of Smith's and Jones' preferences. My own preference will be for the first of these alternatives.' (1981, 128-129)

The first thing to notice about Hare's example is that is not about coercion but merely about distribution. Suppose that Smith actually has the bun. Jones prefers J1 to J2 by a margin of 1000 whereas Smith prefers S1 to S2 by 500. However, Jones can only get the bun by taking it from Smith by force for, whereas Smith prefers S1 to S2 less than Jones prefers J1 to J2 it is still the case that Smith prefers S1 to S2. He is not going to give Jones the bun. In this situation, the fact that Jones prefers J1 to J2 more than Smith prefers S1 to S2 is irrelevant, morally. It is only relevant in Hare's example because Hare is playing God, or Father Christmas. While such situations do arise in real-life, they are not usually the morally interesting ones. Moral interest tends to be in cases where something is taken from one group and given to another. Now it may be the case that Jones desire to take the cake from Smith is 1000 units but his desire NOT to have Smith take things from him is 2000. The fact that he

prefers J1 to J2 twice as much as Smith prefers S1 to S2 no longer matters.

4. Equality

Barry raises the question of what Rawls means by 'equal' in the statement of the first principle. (1973, 39) He concludes that 'the first principle cannot be truly egalitarian, and in practice Rawls treats it as a maximin'. (1973, 41) To be 'truly egalitarian', in Barry's view, the first principle would have to require that A and B get exactly equal amounts so that it would be better that each got less rather than that the distribution was unequal, e.g., if a freedom is such that A can only have it at all if B has more of it then it is better that neither A nor B have it because neither having it is more equal, although less free, than B's having slightly more freedom than A. Rawls, however, is concerned that A can get as much freedom as he can, and that this be distributed as equally as is possible for A to get as much as possible even if this means that B gets more. What is crucial is that A doesn't get less freedom in order that B may get more. Rawls is not out to increase the total amount of freedom but to give both A and B as much freedom as possible as is compatible with NOT decreasing the freedom of the other. This is only possible if A and B are not envious.

How does impartialism deal with this? The individual J, behind the veil, will want both A and B to have as much freedom as possible. As J doesn't know whether he is A or B he will want no restrictions on either of their freedoms. J only has a problem when A wants to coerce B, i.e. when A's freedom to do what he wants to do is incompatible with B's freedom to do what B wants to do. Given that each wants not to be coerced more than he wants to coerce, then J will agree that his freedom to coerce the other be curtailed by force if necessary, whichever he is, in order to ensure that

the other is not coerced (as per example earlier). It would appear that the symmetry of this argument must ensure that A's and B's freedoms are equal. J cannot, on the symmetrification model, agree to any arrangement where A has a freedom to coerce B that does not also give B the freedom to coerce A. Admittedly, if such a freedom is granted then J is agreeing to there being actual inequalities of freedom in the real world. What would be such an inequality with respect to freedom to get what you want? It could only be a freedom which A is prepared to concede to B but deny to himself because A thereby gains a more important freedom. Nearly all the cases I can think of are cases where A and B would both be prepared to allow a third party C to have certain freedom denied A and B in order to ensure that A and B had more important freedoms, that is, it is essential that someone have this freedom whereas everyone else is denied it. Consider, for example, the freedom of police officers to carry guns, or the taxation department to have access to bank records, and so on. Now, in none of these cases would J agree that C be allowed these freedoms in order to pursue ends other than the preservation of freedom C has in common with A and B, nor would J agree to C's using this freedom to advance C's private interests as a citizen by employing his freedom to carry guns (or make arrests, or drive over the speed limit) when this freedom was given him solely to further public ends.

Rawls' principle requires that 'each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all'. (1972, 302) This is oddly worded. Why is it an equal right rather than an equal freedom? Why is it 'the most extensive total system' of equal basic liberties rather than equal liberties? Why are the equal liberties the basic ones? What makes a liberty basic?

Barry's answer to some of these questions is as follows: 'The answer is that Rawls thinks of the various liberties as capable of occurring (within limits) independently of one another, and suggests that, if maximising one is inconsistent with maximising the others, different portions of the various liberties should be combined in such a way as to make the "total system" one of as much liberty, equally distributed, as possible.'

(1973, 34)

Rawls himself gives the example of freedom of speech. 'To illustrate by an obvious example, certain rules of order are necessary for intelligent and profitable discussion. Without the acceptance of reasonable procedures of inquiry and debate, freedom of speech loses its value. It is essential in this case to distinguish between rules of order and rules restricting the content of speech. While rules of order limit our freedom, since we cannot speak whenever we please, they are required to give the benefit of this liberty'. (1972, 203)

It is to be noted here that it is not that A has one system of liberty and B another. This is contrary to the impression that Rawls gives when he words his first principle as 'each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all'. Wording it this way makes it sound as though A has system of liberties S whereas B has system of liberties T. It could be, then, that S is similar to T in either of two ways:

- (1) it has the same liberties as S, but with reference to B rather than A, OR
- (2) it may have quite different liberties but makes A as free as B.

Now, Rawls' freedom of speech example makes it sound like (1). If A is the chairman of the meeting, and B a speaker, then A's freedom to choose the next speaker, to rule when a speaker has spoken enough, etc. does NOT constitute a system of freedom that A has but which B lacks which B is compensated for by having the freedom to speak. All that the principle requires is that freedoms be organised into a system which is, essentially, the same system for all, not that there are lots of different systems all of which give different people equal totals of different freedoms. The only variation between A and B may be that B, say, has certain allowed powers that A lacks which enable both A's and B's freedoms to be enhanced. What J is faced with, then, is the choice whether the freedom to be heard is more important than the freedom to speak whenever you feel like it? J has to rank freedoms so that this ranking is a ranking for everyone. It will not be the case that J ranks being heard high for A but low for B, but makes other adjustments within S and T so that S and T give A and B equal freedom respectively. J decides between the freedom to be heard and the freedom to speak when you feel like it in the context of both A and B wanting not to be coerced more than they want to coerce. Neither A nor B, then, wants to risk not being heard just for the chance to prevent the other being heard. If A and B want to be heard more than they want to speak whenever they feel like it, then J will have no difficulty ranking being heard higher than speaking at will. Likewise, if both value speaking at will higher than being heard, J will rank the former higher than the latter. It is when A prefers one whereas B prefers the other that the problems arise. Rawls treats it as unproblematic that the parties will agree to give being heard priority over speaking at will. This is, presumably, because he regards it as the more important. What needs to be established is that the freedom to be heard contributes more to the preservation of other freedoms than the freedom to speak at will, even for those who want to speak at will more than they want to be heard, that

is, that even those who want to speak at will more than they want to be heard will find that there is more freedom for them overall in areas of their life other than this if they live in a system of freedom which ranks the freedom to be heard higher than the freedom to speak at will. We cannot assume that this is easily shown but it is the kind of thing that needs to be shown. Impartialism gives the framework in which even this much is possible, i.e. impartialism has eliminated a lot of argumentative options that would have made agreement impossible. Nonetheless, the hard, normative, nitty gritty of showing that the freedom to be heard is more important than the freedom to speak at will is yet to be done. While Rawls is right that freedom of speech loses its value for those who see its prime function as the freedom to be heard if there are no rules of order or debate it may still have other values. The case needs to be made that freedom to be heard is more important.

The point, however, has been established that what is under discussion here is the ranking of freedoms in a system that is common to all members of the community, not different, but equally free, systems of freedom for different members of the community. This is not to deny that the use people make of the common system will be such as each appears to have a different system. How the novelist and the politician use freedom of speech is quite different but they may still be part of a system that ranks the freedom to be heard higher than the freedom to speak at will.

The system, however, will give some people freedoms (although 'powers' is perhaps the better word) that are denied to others, for the express purpose of making the system work. If we call these powers 'delegated powers' then the system will give some people powers, or freedoms, that are denied to others in order to maximise the overall system of freedoms. As J doesn't know whether he is A or B or C, and he doesn't know which

individual has these delegated powers that are denied to others, then he will seek to ensure that these powers are both minimal and not open to abuse. It is possible, for example, that police may use their powers of faulting vehicles or arresting drunk drivers to persecute their personal enemies, and so forth.

J, then, may rank the freedom to be heard higher than the freedom to speak at will. This will then become the system of freedom which will apply to A and to B and to C. If he had ranked these freedoms in the reverse order, then this would have been a different system (say T rather than S). S is to be preferred to T if its ranking of freedoms actually gives more freedom to A, B and C than T does. In order for S to work, someone will have to police situations where the freedom to be heard can be frustrated by those who want to speak at will more than they want to be heard. If this task is given to C, then C does not thereby gain a greater right to be heard in his capacity as private citizen. His greater right to be heard, if any, must be only with respect to what he needs to do to facilitate everyone's (his included) equal right to be heard. To do otherwise is to abuse his power as chairman and is grounds for the invocation of removal procedures. It may even be that he has a lesser right to be heard when he acts as chairman. If he wishes to be heard as an equal with others, he must relinquish the chair. This suggests that there needs to be further procedures to ensure the chairperson has an equal right to be heard, i.e. that the selection of a chairperson cannot be used to gag a troublesome individual. There would therefore need to be 'fair' selection procedures. And so on.

To return to the dispute between those who value the freedom to be heard (Freedom E) higher than the freedom to speak at will (Freedom F) E can only be ranked higher than F if those who prefer F to E can see that

the system S which ranks E higher gives them more freedom overall than T which ranks F higher. This parallels the kind of case where those who wish to kill surrender this freedom because they prefer not to be killed, and therefore are prepared to give someone the power to prevent them killing even when they want to kill. Those who want F, then, have to be persuaded that they should give someone the power to prevent them speaking at will, even though they want this more at the time than they want to be heard. This can only be done if it is the case that it is more important to them in their lives to be heard at various times than it is to speak at will. This is a different case from where they want to prevent others being heard more than they want to be heard themselves. This has already been ruled out as coercionist. What we have instead is a conflict within the individual, i.e. which matters most to him - being heard but at the cost of sometimes not speaking or speaking but sometimes at the cost of not being heard. In fact, the situation is more complex than this as there is a continuum of cases. The person who prefers to speak at the cost of sometimes not being heard might never be heard, i.e. this is a risk he takes. The person who prefers to be heard at the cost of sometimes not speaking doesn't run the same kind of risk. He can only sometimes be heard if it is the case that he sometimes speaks. Whichever is chosen, symmetrification still applies. If the parties are both ones who prefer to be heard at the cost of sometimes not being able to speak when they want to, then they will be prepared to empower someone to act as a chairperson in the manner Rawls describes; that is, if J considers the case of people who want to be heard more than they want to speak at will then he will not risk A's speaking at will when B wants to be heard (under specified conditions) should he turn out to be B, and vice versa if he turns out to be A. On the other hand, if J considers the case of people who want to speak at will more than they want to be heard, then he will risk A's speaking at will when B wants to be heard (even under specified conditions)

should he turn out to be B, just to ensure he, as B, can speak at will when A wants to be heard, and vice versa should he turn out to be A. J, then, has no problems with rules of order and debate when A and B both prefer E to F. When they both prefer F to E, or one prefers F to E, then further argument is necessary to show that ranking F higher than E puts a whole range of freedoms at risk which, either individually or together, are more important than F and can only be protected if E is ranked higher than F. Part of this may involve an appeal to majoritarian decision-making as a device for conflict-resolution. J still has problems, though, when either A or B prefers F to E, and prefers F to any other freedom, individually or as a total. This will be a recurring problem for impartialism. What can be said, at this point, is that there appears to be little profit in argument between someone like this, and anyone who doesn't place an absolute value on F. We have here what appears to be a recurrence of the "fanatic" case in a new guise. Resolution of an argument between A and B over whether X is to be done is only possible without recourse to force, that is, is only resolvable behind the veil, if both A and B do not place an absolute value on X. This means that there must be other values besides doing (or stopping) X that each of A and B has such that the desirability of X can be judged in terms of the degree to which it facilitates or impedes these other values. A monomania makes this impossible. Impartialism, then, is only possible as a conflict resolver if each party to the conflict has a plurality or diversity of values or goals such that no one goal is weighted in importance more than all the others put together. This could be called the "monomania" axiom of impartialism.

The upshot of all this is that Rawls' first principle would be one adopted by an impartialist, at least with respect to each person having the same freedoms as any other person in a system of freedoms which ordered all freedoms so as to achieve the maximum freedom possible when everyone has

the same freedom. If Barry is right, and Rawls' first principle chooses between systems of freedoms on a maximin basis rather than on a strict egalitarian basis, then it is in accord with impartialism in this respect, too. The impartialist would NOT reject a system which gave police, or chairpersons, or politicians (parliamentary privilege, for example) freedoms not possessed by the rest of society merely because this meant freedoms were unequal. Provided these freedoms are really delegated powers, there can be an inequality of freedoms. Nonetheless, J will choose that system which maximizes freedoms with minimum delegated powers. However, if S has more pure freedoms than T but T has fewer delegated powers, J will still choose S, provided these delegated powers can operate in such a way as to ensure that the freedoms in S remain greater than those in T.

While this is an impartialist interpretation of Rawls' 'First Principle', it may not be Rawls' own. Rawls does seem to think, for example, that there are circumstances where it is acceptable for some people to have greater political freedom than others, provided this is a temporary step towards the universal possession of such freedoms. Political freedom, however, is NOT a delegated power. It is, rather, what I will call a 'freedom-preserving right'. No matter who you are, you will need the freedom to form political parties, to publicise political views, to put up candidates in order to preserve your freedom to do the other things you might wish to do, such as practise your religion, or your own sexual preferences, or read what you like, etc. In order to guarantee these freedom-preserving rights, J and K will agree to set up rights-protecting institutions such as parliaments, or the police force, but the persons given special powers in these institutions are given only delegated powers. These powers are not intended to give them advantages over other people in securing the freedom to do what they want by denying freedoms to those over whom they have power. Their role is to ensure the equal freedom of all.

Chapter 10

The Priority of Liberty1. Introduction

Our concern so far has been with whether or not A wants B stopped from doing X to A against A's will more than he wants not to be stopped from doing X to B against B's will, i.e. does he want to be free to do X to B against B's will more than he wants B not to be free to do X to him against his will? Our concern, then, has been with when coercion is permissible.

To decide when coercion is permissible we have gone behind the veil of ignorance. This involves adopting the viewpoint of J, a randomly selected velorian who doesn't know whether he's A or B. X could be any act that A might want to be free to do to B against B's will that he would rather that B was not free to do to him against his will. Such acts would include killing, torturing, lying, cheating, robbing.

Within impartialism we have reached a number of general conclusions.

- (1) J will only agree with other velorians that A can be stopped from doing X to B against B's will if X is such that A wants B not to do X to A against A's will more than A wants to do X to B against B's will. J is, in fact, agreeing that, if he turns out to be A who wants to do X to B against B's will, then he, A, should be stopped from doing it. Whenever J agrees that A should be stopped from doing X to B against B's will, he is agreeing that B have a particular freedom, e.g., freedom from the fear of being killed, or, perhaps, freedom from being killed.

- (2) By the process described in (1), a list of freedoms will be established. Some of these will not be compatible with each other, e.g., the freedom to speak at will (E) and the freedom to be heard (F). If A is free to speak at will, then he is free to speak when B wants to be heard and, as A's speaking while B speaks means that B won't be heard (at least in the sense of 'paid attention to and understood'), A's freedom effectively demolishes B's. Both these freedoms are impartialist freedoms, i.e., they both survive the test in (1). When this occurs, E and F have to be ranked. J does not know whether he's the kind of person who would rather be stopped from speaking at will in order to guarantee being heard, or who would rather be stopped from being heard in order to speak at will. He therefore has to calculate which of these freedoms, with their obverse restrictions, best protects him from being coerced in other areas of his life, regardless of who he is as a conflictant.
- (3) J will conclude that unless a case is made to the contrary, A and B should be free to do whatever they want, provided that any freedom that A has B has also.
- (4) J will conclude that, unless a case is made to the contrary, if A is to be stopped from doing some class of actions X, then so is B.
- (5) J will only agree with other utilitarians that A can have freedoms denied B if these freedoms are effectively just delegated powers that protect the equal freedoms of A and B. In fact, delegated powers, which are the unequal freedoms, are only justified when they guarantee, or bolster, defend etc. the equal freedoms.

All of the above is 'formal'. It doesn't of itself give the content to which acts A or B should be stopped from doing. (1) limits the argument to those who share the attitude that they would rather be stopped from doing X if it is the case that they want to do X to others less than they want others to do X to them. The point of this limitation is that argument, as such, cannot resolve the conflict between people who do not have this attitude. Such conflicts can only be resolved by the application of force, i.e., they are to be resolved in favour of those with the greater power. Admittedly, a type of argument may go on among coercionists, but it is an argument about who actually has the force. It is a form of bluff, or a form of coercion that helps one party get its way without its actually having to physically coerce the other. (2) doesn't make such a clear-cut distinction between types of persons and the conflict between those who prefer E to F and those who prefer F to E cannot be resolved by any neat formula, but depends on an assessment of the facts of the situation and on the desires of the parties. Nonetheless, any solution, e.g., that E is to be preferred to F, is to be adopted on the understanding by J that, if he turns out to be A who wants to do X which is a case of F for A but a denial of E for B, then he, A, should be denied F, i.e., he should be stopped from doing X.

With these conclusions in mind, I will now turn to the derivation of the priority of liberty.

2. The Priority of Liberty

Rawls' final statement of the priority relationship, as Barry notes, is as follows: 'First Priority Rule (The Priority of Liberty). The principles of justice are to be ranked in lexical order and therefore liberty can be only restricted for the sake of liberty'. (1972, 302) Earlier Rawls had spelt it out a little more fully, 'By the priority of

liberty I mean the precedence of the principle of equal liberty over the second principle of justice. The two principles are in lexical order, and therefore the claims of liberty are to be satisfied first. Until this is achieved, no other principle comes into play'. (1972, 244)

He also expresses it as follows:

'The principles of justice are to be ranked in lexical order and therefore liberty can be restricted only for the sake of liberty. There are two cases: (a) a less extensive liberty must strengthen the total system of liberty shared by all, and (b) a less than equal liberty must be acceptable to those citizens with the lesser liberty.' (1972, 250)

Note: (a) reinforces the interpretation given previously that everyone shares a single system of liberty, not that each person has his own, equal system.

To fully appreciate the principle of the priority of liberty we need to remind ourselves of the two principles of justice.

(1) The First Principle of Justice (The Principle of Greatest Equal Liberty).

Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.

(2) The Second Principle of Justice (The Difference Principle and the Principle of Fair Equality of Opportunity).

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.

Putting the Two Principles of Justice together with the Priority Rule, it becomes clear that it is each person's right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all that has priority over the Second Principle of Justice. This means that no social or economic inequality is permitted, even if it is to the greatest benefit of the least advantaged, (to restrict ourselves to the Difference Principle) if it means that each person does not have an equal right to the most extensive total system of equal basic liberties. The force of this priority, then, will depend on what is understood by having an equal right to the most extensive total system of equal basic liberties, etc. Some of this has been discussed previously. What is to be noted now is that it appears that the priority doesn't necessarily hold for non-basic liberties. To understand exactly what the Priority Rule requires, then, we need to look at what Rawls means by 'basic liberties'.

Rawls list of basic liberties is as follows.

- (1) Equal participation in the political procedure defined by the constitution that is, 'all citizens are to have an equal right to take part in, and to determine the outcome of, the constitutional process that establishes the laws with which they are to comply ... the principle of participation also holds that all citizens have equal access, at least in the formal sense, to public office. Each is eligible to join political parties, to run for elective positions, and to hold places of authority'. (1972, 221-224) It also includes freedom of speech and assembly. (1972, 61)

(2) The Rule of Law. Barry summarises this as follows: 'It includes (1) the principle that "ought implies can", and therefore that laws should require only possible behaviour, (2) "the precept that similar cases be treated similarly", (3) "the precept that there is no offence without a law" which "demands that laws be known and expressly promulgated", etc., and finally (4) "precepts defining the notion of natural justice" which require, for example, that judges be fair and impartial, and no man may judge his own case".

(3) Equal Liberty of Conscience. This includes freedom of religious belief, freedom of thought, freedom of conscience, freedom of life-style when this doesn't harm others, freedom of person, freedom to hold personal property, freedom from arbitrary arrest and seizure. (1972, 61, 450)

Rawls summarises these as follows, 'The basic liberties of citizens are, roughly speaking, political liberty (the right to vote and to be eligible for public office) together with freedom of speech and assembly; liberty of conscience and freedom of thought, freedom of the person along with the right to hold (personal) property; and freedom from arbitrary arrest and seizure as defined by the concept of law'. (1972, 61)

It is these liberties then whose total cannot be reduced or which cannot be rendered unequal in order to achieve social or economic inequalities no matter how great a benefit these social and economic inequalities bring to the worst-off.

Why is this so? In justification of it, Rawls says,

'Now the basis for the priority of liberty is roughly as follows: as the conditions for civilization improve, the marginal significance for our good of further economic and social advantages diminishes relative to the interests of liberty, which become stronger as the conditions for the exercise of the equal freedoms are more

fully exercised. Beyond some point it becomes and then remains irrational from the standpoint of the original position to acknowledge a lesser liberty for the sake of greater material means and amenities of office. Let us note why this should be so. First of all, as the general level of well-being rises (as indicated by the index of primary goods the less favoured can expect) only the less urgent wants remain to be satisfied by further advances, at least insofar as men's wants (are) not largely created by institutions and social forms. At the same time the obstacles to the exercise of the equal liberties decline and a growing insistence upon the right to pursue our spiritual and cultural interests asserts itself. Increasingly it becomes more important to secure the free internal life of the various communities of interests in which persons and groups seek to achieve, in modes of social union consistent with equal liberty, the ends and excellences to which they are drawn. In addition, men come to aspire to some control over the laws and rules that regulate their association, either by directly taking part themselves in its affairs or indirectly through representatives with whom they are affiliated by ties of culture and social situation.' (1972, 542-43)

Rawls, then, does not believe that the First Principle of Justice always has priority over the second. It only has priority once a certain threshold is reached in terms of the satisfaction of urgent basic wants. This means that when the desire to satisfy urgent basic wants is compelling enough, then it is no longer the case that each person has an equal right to the most extensive total system of equal basic liberties in the sense of either

- (1) each person has an equal right to the same total system of equal basic liberties but it is no longer the most extensive one possible at the time; OR
- (2) the system of liberty is such that not everyone has an equal right to equal basic liberties.

It might be that Rawls believes it rational of the velorians to waive the priority of liberty if the threshold hasn't been reached for only one of these options. This does not, however, seem to be the case. He says, for example, that, if the basic wants are not compelling, then it is not rational 'for the persons in the original position to agree to satisfy them by accepting a less than equal freedom', (1972, 543) which does suggest that it is rational to accept a less than equal freedom when the wants are urgent. This shows he is committed to (2) above. His general position also seems to commit him to (1).

There are, then, a number of issues that need to be considered in Rawls' account of the priority of liberty before we can determine whether impartialism agrees with it or not.

3. The 'Civilization' Theshold

As we have seen, Rawls believes that there is a threshold below which the principle does not apply because it is rational of the velorians to prefer inequalities of the basic liberties in order to satisfy urgent basic wants. Rawls (1972, 543) says that it is rational for the velorians to agree to this when

- (1) there are obstacles to the exercise of the equal liberties
- (2) urgent wants need to be satisfied
- (3) it is not important under these conditions to secure the free internal life of the various communities of interests
- (4) people do not aspire to some control over the laws and rules that regulate their association.

Before we can assess these claims, we need to consider exactly what it is the velorians agree to when they agree to give up the priority of liberty,

beginning with Case (2) looked at earlier, namely, where people have unequal rights to the basic liberties.

4. Less than Equal Liberties

In what ways can the basic liberties be unequal? Only by one person, or group of people having more of them than another. I will limit myself to the equal participation and rule of law cases for the moment. The principle of equal participation is that each person 'have an equal right to take part in, and to determine the outcome of, the constitutional process that establishes the laws with which they are to comply'. I will look at this first. Suppose A is tempted to give up some, or all of this right, for some economic gain. This means that someone else, B say, has promised A that, if B is given a greater right to take part in, and determine the outcome of, the law-making process, then B will give A more money, or whatever, than A would have got if his participation was equal to B's.

J, not knowing whether he is A or B, will want A to determine the outcome of any law-making process as much as B. If A doesn't do so, then the outcome of this process is likely to favour B over A. J will not agree to B having a greater say in the determination of law-making than A if this means that B gains the power to break his promises to A, or to make A do as B wants without recourse to remedy. The principle of equal participation has to do with the balance of power in the society. J will not agree to one party having a balance of power over that of another such that it can ignore the interests of the other, even if in the first instance, or always, this benefits the other. The risk is too great. There can be no guarantee that this power won't be abused or even that the promise can be honoured. If A agrees to certain sacrifices in order to gain economic benefits he will only do so if he can opt out of the arrange-

ment if the promises are broken. He can only do this if he retains sufficient power to do so. If, however, he has surrendered his right of equal participation, then he has lost his power to cancel the deal. He cannot participate in the decision-making process in order to reverse decisions contrary to his interests. J, then, will not agree to surrender his right of equal participation (although it's more a right of equal determination) just for an economic benefit, for the reasons just given.

Rawls would agree that this is so after the civilization threshold has been reached but does not believe that it automatically holds prior to the threshold. He is probably correct that people will not concern themselves greatly with spiritual and cultural interests, with the free internal life of the various communities of interests in which persons and groups seek to achieve, or even that people will not aspire to control the laws and rules which regulate their association, while their most urgent and basic wants, presumably food, drink and shelter, remain unsatisfied. All this shows, however, is that certain ideals of life won't rate very high if the continued existence of life itself is at stake, including ideals that are intimately associated with equal liberty. Nonetheless, while 'modes of social union consistent with equal liberty, and the ends and excellences to which people are drawn', may not rank high in conditions of scarcity, it does not follow that the equal basic liberties themselves are not still paramount. Why should J want to let A surrender his equal determination of laws to B just because B promises to improve the supply of food, drink or shelter as a consequence? There are many subsistence cultures which do not hand over political power to their economic entrepreneurs in this way. J would have to be sure that B would keep his promise, and that B would not then use his extra power to make A's lot even worse than it was before. The only way this latter can be ensured is if A doesn't surrender his equal rights in determining decisions. J can adopt either of two models

here. He can insist on equal direct determination in decision-making for whichever of A or B he turns out to be, or he can adopt some form of equal indirect determination. In either case, he will insist that his direct determination of decisions be equal to anyone else's, or that his indirect determination be equal. J's first preference will always be that A's or B's determination of decision-making be direct, as this is the most effective way for A or B to ensure that their interests are represented to others as A or B themselves perceive them. Any indirect process will distort A's or B's preferences. Nonetheless, J may well be prepared to accept that direct representation has more disadvantages than indirect representation as a method for making most decisions. J may still want to insist on direct determination for the most important decisions, whatever he thinks these are. It is not appropriate to go into details here, but constitutional questions are obvious ones. There may be other cases, too, where an issue is so important it warrants a referendum. J, then, will insist on equal determination for A and B, whether it be direct or indirect, because this is the only way for him to ensure that, whoever he is, he has some control over the decisions that affect his interests. This will mean that he will insist on freedom of speech and freedom of assembly because these, too, are essential if he is to be able to marshal support so that decisions are determined as he wishes them to be, given that his structural power to determine decisions is to be no more than anyone else's. This leads to questions about how to resolve disputes between large bodies of conflictants about which decision is made. Rawls suggests this be done by a constitutionally controlled majoritarianism for societies above the civilization threshold. The kinds of arguments I have adduced so far indicate that a similar system would apply below the civilization threshold, at least with respect to equal determination. The rule of law will generally obtain below the civilization threshold, except, perhaps in cases of national survival such as war, although there may well be good

reasons why they should apply especially then. Be that as it may, the crucial point is that J would not agree, as Rawls believes, to surrendering equal determination for economic gain because this means giving oneself over to the power of others in a way that is not retrievable within the system, but is only retrievable by the forceful overthrow of the system. Giving oneself over to the power of others in this way is irrational because it means that one cannot even ensure that the benefits for which such a transfer was made will be received, so that one may be even worse off after the transfer than before. As there is no way that J can be guaranteed this will not happen, J will not agree to it.

5. Indirect Equal Determination of Decisions

J can, however, agree to something like it, and perhaps this is what Rawls has in mind. Suppose J decides that, under certain circumstances, no matter who he is he has more to gain under indirect participation in decision-making at the national level than he has in direct participation, provided there are certain constitutional safeguards. In particular, suppose he believes he has more to gain economically no matter who he is under indirect participation, that is, he hands the power to make decisions about the allocation of resources to some group, for example, a parliament, or some individual, for example, a President. This means he surrenders a freedom. He is no longer equally free to determine such decisions as anyone else because he has set up a group of people, of which he may not be a member, to make decisions in this area unrestrained by any direct input from him. J will still insist on indirect input, i.e., he will insist that there be some mechanism whereby A or B can stop, or contribute to stopping, the activities of this group if A or B see it as acting contrary to their interests, i.e., it is not producing for them the extra wealth for which they set it up. In most systems, this mechanism of

indirect determination of economic decisions is the election, but there may be other methods (e.g., the Gallup Poll). Let us call this decision-making body 'R'. Most conflictants such as A and B will not be members of R, therefore they will have only an indirect determination of decisions but some conflictants such as C and D will be members of R and therefore will have a direct determination of economic decisions. This means that C and D have a freedom lacked by A and B. The structure of which A, B, C and D are a part is such that C and D are free to directly determine decisions whereas A and B aren't.

Now, J will be prepared to give C and D powers or rights or freedoms not possessed by A or B provided these powers or freedoms are not irretrievable powers, and provided that they are delegated powers in the sense discussed earlier. The system will be a two-stage system, that is, J will only agree to only C and D having direct powers whereas A, B, C and D have indirect powers if the indirect powers of C and D are equal to those of A and B. If this is not so, then the retrievability of these powers by A and B is weakened, helping C and D entrench themselves in R with these additional powers in ways which A and B cannot affect even when C and D no longer make decisions that take A's and B's interests as equal to C's and D's. Moreover, J will only allow C and D these extra powers if they are used in a way that benefits the interests of A, B, C and D equally, i.e., J will not give C and D these powers under systems whose effect is to benefit C and D at the expense of A and B.

So, even below the civilization threshold, J will not agree to A surrendering his equal right to determine decisions just because B (or C or D) offers him greater wealth, if A's surrendering this right is either permanent or for such a long period of time that A can be irredeemably harmed by B's actions while powerless to do anything about it. At the

most, J will allow A to surrender his direct freedom to determine decisions to B provided A retains the same indirect freedom to determine freedoms as B, and provided A retains the same direct freedom (and indirect freedom) as anyone else to retrieve from B the delegated power, or to replace B with someone more responsive to his interests. As a rational agent, J cannot agree to A surrendering to B for the promise, or even the receipt of wealth, that power that enables A to prevent B reneging on his promise, or to prevent B from making A's situation even worse than it was prior to the transfer of power even though B has honoured his promise.

In surrendering power to the members of R (C and D), J does so only on the condition that C and D treat their own interests as private citizens as equal to, not superior to, those of A and B. J will insist that there be machinery built into the operation of R that prevents C and D misusing their additional power. As the point of J allowing R to be set up is to better his financial position if he is the worst off, J must give R power to compel A and B to do certain things, e.g., to surrender part of their moneys as tax, or whatever. What J will not give R is the power for members of R to benefit themselves at the expense of A and B, or give them the right to deny A and B the power to dissolve R, or to deny them a less than equal indirect determination of the policies or the membership of R, or a less than equal direct determination of the nature of R. There appears to be as much reason for J to insist on this below the civilization threshold as above it. In fact, given that the consequences of misallocation of power below the civilization threshold are the more serious because survival itself is more at risk than there is probably more reason for ensuring the maintenance of the equal right to determination.

It might be argued that this conclusion is empirically implausible. In feudal times, peasants put themselves unreservedly under the power of a

lord in order to gain protection from marauding hordes, etc. This is true, but the rationality of an action when the veil is lifted and people are in superior and inferior power relations does not necessarily carry across to the position behind the veil. It may be in real life, when one already lacks power, that it is rational to submit unconditionally to one ruler rather than another as the lesser of two evils. This in no way shows that it is rational to agree to let these rulers have this unconditional power in the first place, which is the issue that J is concerned with behind the veil. Similarly, it could be argued that, throughout history, women have surrendered an almost absolute power over their lives to their husbands. To say that the actions of half humanity are not rational is to put one's definition of rationality into question. Once again, what it is rational to do given that you already are in a situation with a particular power distribution is distinct from the question of whether it's rational to agree to that power distribution in the first place, or to the power distribution that caused it to come about. It may be that in any age when pregnancy, breast-feeding etc. prevent women being trained as soldiers and thereby makes them vulnerable to the attacks of the above marauding hordes that it is rational to put themselves under the protection of a single man even when this gives him the power to treat them badly should he so choose. They can only hope he doesn't so choose. Behind the veil, however, they insist on some arrangement that gave them as much power as the men in how they or the men were treated.

What has been shown so far, then, is that J wouldn't agree to A having a lesser liberty than B in the matter of determination of decisions, even if this lesser liberty offered, or delivered a large economic benefit, and even if this occurred below the civilization threshold. Certainly, A is more vulnerable to B's misuse of power below the civilization threshold than he is above it.

I will now turn to a consideration of the two other basic liberties, namely, the rule of law and the equal liberty of conscience. I will also look at the case of J considering a less extensive equal liberty rather than a less than equal liberty.

6. The Rule of Law

Essentially the same considerations apply here as applied with equal determination. While J may be prepared to hand over power to judges because there are many advantages in doing so, J will not agree to less than equal liberty with respect to the rule of law just to get an economic gain. If judges (C and D) have powers the ordinary citizen doesn't have (A and B), it is solely so they can enforce an equal liberty for all of A, B, C, D. J will not agree that A should give power to a Judge C to judge A's case if C is merely going to benefit C, or C and D at A's expense. Nor will J agree that C should judge A's case if C is the other party to the case, even if high financial rewards are offered, because this as an accepted practice can lead to C (or D) stripping A of the benefit he gained this time by a decision of a later case. The point about each of the rules of law is that they are procedures designed to ensure that the parties who make, and administer the law, do not use the power given them to do so to benefit themselves consistently at the cost of those not so placed. To this extent, the rule of law is a rights-protecting institution, i.e., those who administer the law still only have an equal freedom with everyone else in the areas that it exists to enforce. Fruit fly inspectors, for example, can enter anyone's yard to see if their trees have fruit fly. This is to prevent anyone having fruit fly. It does not mean that fruit fly inspectors are to be free to have fruit fly ridden trees but no-one else is.

Now, the rule of law is particularly important in those situations where survival is at stake, i.e., where those who are the stronger are likely to advantage their own survival at the expense of others. J is just as unlikely to allow A to suspend his right to the rule of law on a permanent basis below as above the civilization threshold. This is so even if, were B to have total power, B could meet A's urgent survival needs but, if B didn't have total power, then A will starve, or some such. The most that J could concede to A here is that A give B a temporary but retrievable absolute power. The only case in which J would agree to a permanent and irretrievable power for B is if A wanted nothing else but to live, that is, if A didn't care what his life was like as long as he was still alive. The case, however, is hardly one of concern, because J, even in these circumstances, will only agree to B's absolute power if A's survival does depend on B having absolute, permanent power. It is difficult to imagine what such a case would be. Any remotely plausible case requires, at the most, a temporary absolute power for B even if J knows (and presumably A knows) that B can deliver the goods, and this is A's only hope.

7. Equal Liberty of Conscience

This is not the same as freedom of speech and freedom of assembly. These are necessary parts of A being able to maintain an equal right of participation and determination with B. Equal liberty of conscience is not a liberty needed to preserve other liberties but one of the other liberties to be preserved. Rawls seems to operate a distinction between types of desires. Firstly, there are desires like the desire for food, the desire for drink, etc. i.e., basic desires. These can be satisfied by someone supplying the necessary materials. Secondly, there are desires for freedoms. These freedoms aren't necessary to meet the basic desires, nor are they necessary to meet any desires but rather they are just

another kind of desire. They cover such things as the freedom of speech, and the freedom to assemble, etc. Actually, Rawls misnames these. On his analysis they are not desires for freedoms at all, but desires for the things, e.g., the desire to speak, the desire to assemble. It is almost as if he would call the desire to eat the desire to be free to eat. For Rawls, the civilized life is one where you can speak to people, exchange ideas, write poems, read novels, go to plays, where you can assemble with others to listen to music, or play sports, etc. They are, in a way, ends in themselves. They are manifestations of the Aristotelian principle, of the complexities of human interaction that provide us with satisfaction. As such, they only become desires that figure largely in our lives when urgent and basic desires are met. It is not so much that we will only want to have freedom of speech or freedom of assembly above the civilization threshold as that we will only want to engage in certain kinds of speaking or certain kinds of assembling above the threshold (but then we will only want to engage in certain kinds of eating, drinking, shelter, etc. above the threshold). As you will only want to engage in these activities at a certain level, you will only want the freedom to engage in them then, too. (Nonetheless, while lots of people will be free to engage in these activities, they won't want to). But this is just as true of the freedom to eat, or drink, etc. in certain ways. It will only be wanted when the desire to eat and drink in these ways arises, and this will not occur until a certain level of civilization. However, even at levels below the civilization threshold, each of us will want to eat in order to survive, and will want the freedom to eat in order to survive, that is, we will not want people interfering with the eating we engage in in order to survive. Now, it may well be that we can only stop people interfering with our freedom to eat in order to survive, if we are free to speak about how to stop this. This means that we will want to be able to have freedom of speech, not for any fancy civilized reason, but as a tool to enable us

to survive. It is a freedom-preserving right. Likewise, with the freedom to assemble. It may be that only if we are free to assemble that we can prevent interference to our freedom to eat, or drink or have shelter. Freedom of conscience, and freedom of religious belief in particular, is unlikely to be a freedom that helps us secure our freedom to eat, or drink or build shelter, unlike all the other freedoms listed earlier (determination, rule of law, freedom of speech and assembly, the right to a trial, etc.). Freedom of religious belief (or practice, more so) is one of those cases of wanting to be free to do something for its own sake, not a case of wanting to be free to prevent interference to something being done for its own sake. We will want freedom of speech and assembly in order to protect our freedom to practise our religion, but we do not want the freedom to practise our religion to protect our freedom to speak or assemble. Practising one's religion doesn't help bring about greater freedom of speech or assembly unless prayer is more efficacious than it appears to be, or unless our practices invoke magical powers that have a causal effect upon those who would suppress our speaking or assembling.

The upshot of this is that freedom of conscience is not a basic freedom if a basic freedom is one that frees us to perform activities that are essential to our protecting our interests on an equal basis with others. Freedom of speech is a basic freedom because it frees us to speak, and speaking is essential to our protecting our interests against those who would alter the balance of power so as to advantage their interests at the cost of ours. One of our interests may be practising our religion. Freedom of speech helps protect this. Another of our interests may be speaking itself (of various kinds), or assembling itself (of various kinds), or participating in making the laws and rules that regulate our association for its own sake. Being free to do these, however, even for their own

sake, means that we can use these very freedoms to protect other freedoms. Not all activities that we do for their own sake, if we are free to do them, have the same usefulness. Liberty of conscience is one such. It is not, therefore, a basic freedom. It is, therefore, possible that J would allow A to give B power to deny freedom of conscience for economic gains. J, however, would not do so just because B wanted to prevent A from having freedom of conscience more than B wanted not to be prevented from having it himself. This is ruled out within impartialism. J, therefore, has to find some acceptable reason why B should want A prevented from having freedom of conscience given that he would rather have it himself than prevent A from having it, but, that if A is to be prevented from having it, so is he. It may be that, further developed, this gives liberty of conscience the kind of priority over the second principle that Rawls hopes for it. A fuller explanation here, however is not appropriate.

It should also be noted that the case that concerns Rawls is really a coercion case. The case is that B has the means to supply A the essentials of life but A doesn't have the means to supply these for himself. B, however, will only supply them to A on the condition that A agrees to give B permanent total power over A, i.e., that A renounce his basic freedoms. J will only agree to this if, were he B, he would be prepared to surrender his freedoms to A if A rather than B had the only access to the means of survival. The cases between A and B must be symmetrically acceptable when reversed. Now, if J is A, he wants to coerce B less than he wants to be coerced by B, and if he is B, he wants to coerce A less than he wants to be coerced by A. This means that J would not agree to A making this offer to B, or B making it to A. For either to do so would be to show themselves to be coercionists. I argued earlier that coercionists could get J to agree to a principle of action but could not get J to determine which particular act should be done unless J knew which conflictant he

was. In this case, J knows that one of A or B has the only access to the means of survival, whereas the other doesn't. The question here is whether the one with this access should deny it to the other on condition of surrender of the basic freedoms. Once again, the coercionist can only answer this question when he knows who has the power, and who doesn't, i.e., if he turns out to be A, and A has the power he will give one answer. If he doesn't have the power, he will give another. The impartialist's answer, however, doesn't depend on knowing which conflictant he is. He can answer that whoever has this power shouldn't use it to deny the other person the basic freedoms.

8. A Less Extensive Liberty

We have seen, then, that the utilitarians will not agree to there being a lesser than equal liberty even if, were B to have total power, B could meet A's urgent survival needs but, if B didn't have total power, then A would starve, or some such. At least J would only agree to this if A only wanted to live for the sake of living, rather than that A wanted to live to achieve anything of his own life plans. Even this case is weakened, however, by the fact that there would be no situation where A could eat only if B had permanent total power. The most that could happen is that A could eat only if B had total power, i.e., if A lost his basic freedoms. In these circumstances A may agree to surrender his basic liberties only for as long as was necessary to get beyond the bare survival stage. Moreover, he would insist on retaining the right to assess B's performance so that he could replace B, or adopt alternative strategies, if B wasn't supplying him with the means of bare survival or appears to be less than conscientious in bringing about a better than mere survival situation. A, then, is in the paradoxical situation that he might, if his survival depended on it, surrender as many of his basic freedoms as necessary to ensure survival provided he retained enough of these basic

freedoms to ensure that his surrender of the others paid dividends and to regain them when the grounds for surrendering them had passed.

The other strategy that Rawls suggests might be adopted below the civilization threshold is to have a less extensive total system of equal basic liberties. This assumes that, at the time, a more extensive total system of basic liberties was possible but only at the cost of economic hardships. The distinction made earlier between the political liberties and the rule of law as freedom-preserving rights and the liberty of conscience still applies here, where liberty of conscience is not the same as freedom of speech or assembly but rather the liberty of religious practice. Freedom of speech will include freedom to speak about religious matters as these relate to the control of power in the society. It is not clear how a less extensive equal basic liberty would increase wealth. Perhaps the idea is that political liberty and the rule of law is a costly business, i.e., people need to be taxed to support it. Below the civilization threshold, people are more concerned with having their money in order to survive than they are to ensure a fair system of justice and a participatory political system. While there may be some truth in this, it cannot be the case that they will agree to no political participation because this means an unequal liberty, i.e., giving power to some group without retrievability. Likewise, for no rule of law. The scenario that is possible is that they are prepared to bypass the benefits of certain grand systems of participation, etc. in order to secure survival, that is, they may be prepared to live in clans rather than tribes, or tribes rather than states, if this reduces their financial burden. This, however, will only be so as long as the relationship between clans is such that their wars with each other do not put life and property at so much risk that it is worth becoming a tribe or a state with greater overheads in administration that requires greater taxes. Nonetheless, at the clan

level, J will insist on equal basic liberties for all members of the clan. Lifewise, once the clans become tribes and the tribes states. As the costs of remaining at a less sophisticated level of organization increase, so the motivation to amalgamate and have less face-to-face, more bureaucratic systems of participation and justice increases. These will only be possible at an economic cost, that is, taxes. It is not clear, however, that there really is a civilization threshold here. The nature of participation will vary. The smaller the group, the more direct participation there is likely to be. Unfortunately, the smaller the group the more vulnerable to attack it is, i.e., the more the lives of its members are at risk if its environment is accessible to the ravaging hordes, etc. As groups become bigger, as co-operative effort increases wealth, the concepts that people have as to what their lives are about, that is, what they value, will change. What was not possible in the small society may become possible in the bigger one. In a larger society people will be free to do things they could not do in the smaller. In cities, where there is no longer the close scrutiny of family and neighbours of one's activities, there is a freedom to be private that does not exist in the smaller society. This may well come to be valued by many. There is also, as Durkheim and others have pointed out, the dangers of anomie and alienation. There is also likely to be a freedom of life-style, that is, an individuality or a lack of conformity, that is missing in the smaller society. It is possible to interpret these as greater liberties, but they are not greater basic liberties, in the sense of freedom-preserving rights. Rather, they are freedoms that may be obtained and maintained by the exercise of the freedom-preserving rights. It may well be that, once a certain economic level is reached, people will prefer more privacy, more diversity, more opportunity for cultural activities like the arts, and so on than they will prefer to possess material goods. This, however, is something they will hand over to a social and economic policy determining

group such as a parliament (i.e., the 'R' of previous sections) to sort out by procedures that arbitrate between the interests of conflicting groups in the society. Below the civilization threshold they may not want their possession of material goods put into competition with liberties such as privacy, individuality, etc. Physical security, physical comfort may matter more. Even here, however, R will be the determinant body. It's just that the preferences of those it represents so overwhelmingly favour the material over the cultural, or whatever, that R will not dare deprive its electorate of one for the other. The less surplus wealth, above and beyond that needed for material comfort, the less likely people are to be prepared to be taxed to fund other people's demands for support in cultural, etc. activities. None of this, however, affects everyone's interests in ensuring an equal say in the decisions that are made, at least indirectly. In fact, if people want to keep their material goods rather than fund cultural activities, they will need their basic liberties to give them the power to control the activities of R. To the extent, then, that being below the civilization threshold affects people surrendering liberties for economic gains it is not the basic liberties they will be prepared to surrender nor the equality of these liberties.

To what extent will the basic equal liberties be surrendered for economic gain, without losing their equal nature? The problem here is the choice between losing your wealth and weakening the guarantees you have that you will be free to use your wealth. It's a question of sacrificing some of your wealth in order to guarantee the free exercise of the rest of it. The less wealth one has, the less of it one can afford to lose and the less of it one has to spend freely. Nonetheless, one can lose one's life, or one's freedom of movement (being jailed, etc.) and one will be prepared to spend as much as is necessary to give reasonable security against these eventualities. The poorer the society, the less sophisticated

will be its method of securing justice, i.e., it will not be able to afford an apparatus of lawyers, courts, appeals, etc. Nonetheless, its members (or J, not knowing which of them he is) will still want the innocent cleared and the guilty punished and will be prepared to spend a reasonable amount of their money to ensure this. It is doubtful that how much this is can be reduced to a formula. Likewise, they will want their government to make decisions responsive to the will of the majority but they may not be able to afford all the modern apparatus associated with this. The best that they can do here is to ensure that enough is spent on the parliament and the justice system to ensure that it is able to respond to demands that more be spent in these areas when the majority of people are unhappy with their performance. How much this is will have to be a matter of judgement, but it applies even when there are few civilized liberties as when there are many.

9. Barry's Treatment of Rawls' Derivation of the Priority of Liberty

Barry claims that Rawls' official doctrine is 'so outlandishly extreme that it is scarcely worth devoting any space to its discussion. It can be accepted only if wealth is assigned a value that is literally infinitesimally small in relation to liberty, so that it would be judged worth dropping from general affluence to general poverty in order to score a minute gain on the 'liberty' criterion, if such a choice were presented to a society'. (1973, 60)

I have argued that, understood as I have described it, the priority of the first over the second principle is not outlandishly extreme. This has involved showing that the basic liberties only include the freedom of participation and the rule of law but not freedom of conscience in any sense additional to freedom of speech and assembly. ('Conscientious

objection' to a war, for example, may be a non-basic liberty, as may freedom of sexual preference, and a lot of the standard civil liberties). This removes some of the problems Barry finds with Rawls' account. It still remains possible that it is 'outlandishly extreme' to sacrifice any amount of wealth for either an increase in the total system of equal basic liberties or for an increase in the equality of the total system of basic liberties. Barry is probably right here. It is, once again, going to be a matter of judgement. Suppose J considers a case where a drop from general affluence to general poverty means one innocent person will not be found guilty who otherwise would have been. Clearly, the more we spend on our system of justice, the less miscarriages of justice are likely to occur, e.g., the quicker cases can be heard, the more thoroughly evidence can be collected, and so on. There will need to be a balance reached here that J is prepared to accept given the wealth of the community, given that he may be the innocent person and given that he may be one of those taxed. (This is a problem Nozick tackles in 'Anarchy, Utopia and the State'). Similar points obtain for participation. In a modern democracy we could hold referenda on all issues. J would have to calculate whether the added participation is worth the cost, both financial and in terms of time out of one's life. Once again, he would have to settle on a balance. Nonetheless, the balance is to be struck in such a way as preserves the effectiveness of his basic freedoms for, without them, his wealth is of dubious value, being at risk from those who control the decision-making and judicial processes. If anything, J would err in the direction of more money going on guaranteeing basic freedoms that is strictly necessary.

The rest of Barry's discussion concerns itself with taking seriously Rawls' 'civilization threshold' for the priority of liberty. The arguments I have adduced so far have been designed to show that this threshold is not to be taken seriously as far as the basic liberties are concerned. Barry's

points still hold with respect to the non-basic liberties.

What is crucial, I think, to a correct understanding of the relation between wealth and liberty is the distinction between freedom-preserving rights and "freedoms." Rawls has freedom as an ideal. He wants people not to sacrifice for increased wealth any of political liberty, rule of law, the tolerance of diversity, the freedom to engage in the arts, the freedom to be an individual, to not conform, the freedom to be private, etc. He has an ideal of 'civilized man' that he doesn't want people behind the veil sacrificing for greater wealth, so he wants to build constraints into the veil that prevent this. He has, however, failed to distinguish those elements of freedom that may be an ideal, and those elements that aren't, that is, the freedom-preserving rights. People may choose all sorts of freedoms behind the veil but they must, as rational agents, choose the freedom-preserving rights. Rawls doesn't need to worry about his ideal freedoms because, in general, the two requirements of (1) symmetrification and (2) not wanting to coerce more than not be coerced, will guarantee that the velorians don't swap them for wealth. Consider the freedom of privacy. J is not considering cases where A wants to coerce B into lack of privacy more than he wants not to be coerced into it himself. Rather, A wants not to be coerced out of privacy more than he wants B coerced. He will, therefore, have a prima facie preference for the freedom to be private. Certainly, he will want some greater good to accrue for his sacrifice of privacy. Can this greater good only be a greater freedom? We have seen that this notion is not a clear one. The more kinds of desires a person can come to have the greater need for freedom (in a sense) follows, because the more areas of his life can be obstructed by others. So an increase in material wealth, which creates greater capacity to satisfy desires, also creates greater opportunity for obstruction. As long as people aren't obstructing these freedoms just for obstruction's sake

(i.e., they are not coercionist) then the relation between Rawls' ideal freedoms and wealth is negotiable through the parliament, provided it really can be shown that there is economic gain from the denial of these freedoms. Liberty of sexual preference is a case in point. It may be the case that homosexuality being illegal is of economic benefit to sufficient people to justify it. (The benefit might not be strictly 'economic'. For example, the military effectiveness of an army may be affected by its allowing homosexuals in its ranks). What is crucial, however, is that J agrees that he is prepared to ban the sexual preference of A in order to gain the economic benefit if he is B, and of B, in order to gain the benefit for A. If, for example, soldiers fight better when they are not being sexually distracted at the front, therefore all front-line troops have to be heterosexuals without the opposite sex present, then it is an impartially acceptable policy to ban homosexuals because not doing so is unfair to the non-homosexuals in the front. Either no-one can have sex, or they all can. Rawls, then, has even less cause for worry about his ideal freedoms because J will only accept that A be coerced into losing an ideal freedom for B's economic gain if he is prepared for B to lose it for A's. This means J has to be prepared to ban B's heterosexuality for A's economic gains as well as A's homosexuality for B's. People, however, want the economic gain precisely to indulge their own versions of the luxury freedoms. J has to accept that the gain is worthwhile regardless of how many people there are in A or B. He doesn't know whether he's in the minority or not and cannot calculate on the odds that he isn't. As a consequence, there will be strong arguments against surrendering a freedom just for an increase in wealth.

Chapter 11

The Difference Principle1. Introduction

In the previous two chapters we have seen that impartialism reaches similar conclusions to Rawls about the distribution of liberty, namely, that everyone has an equal right to the equal basic liberties, although we had to modify Rawls' list of equal basic liberties slightly. We saw, however, that the parties behind the veil would not always agree to the system of equal basic liberties being the most extensive one possible given the economic circumstances. They would have to calculate on a cost-benefit basis the degree to which their wealth or its enjoyment was put at risk by the extent of the equal basic liberties they possessed. The more money spent on protecting the equal basic liberties, the less likely they were to be denied the enjoyment of their wealth but also the less wealth they would have to enjoy. At some point, there would have to be a diminishing return on increased expenditure on the enforcement of the equal liberties. At this point, it would be rational for the veilers to refuse to extend the total system of equal basic liberties even though it was possible to extend it.

Rawls' Second Principle of Justice deals with social and economic inequalities, rather than with inequalities of liberty. I will only concern myself with the first part of the principle, which Rawls calls 'The Difference Principle'. In its simplified form, it states that -

'Social and economic inequalities are to be arranged so that they are reasonably expected to be to everyone's advantage.' (1972, 60)

Rawls later glosses this principle so that it reads that these inequalities

are to be 'the greatest benefit of the least advantaged'. (1972, 83)

The parties behind the veil do not know who they are. They have to decide on the distribution of wealth in a position of uncertainty. Rawls argues that the rational principle for them to use in distributing wealth is maximin, that is, they are to opt for whichever principles guarantees them the best minimum. My task in this chapter is to see whether Rawls' arguments for maximin hold for impartialism.

2. Average Utility or Maximin?

The parties J and K behind the veil are to determine how to solve conflicts of desire between the conflictants A and B over the arrangement of social and economic inequalities. Barry argues that Rawls has made a 'monumental confusion' in his case for maximin over the principle of average utility, namely, that Rawls casts J's and K's calculations in terms of the choice of which society to enter rather than the choice of which institutions within a particular society to choose. The situation, however, is not one where J and K are arguing whether Society 1 is better than Society 2.

Barry concedes that 'if it were true that the choice were a choice among societies ... it is at any rate arguably rational to pick the society with the highest minimum provided (a) that the alternative minima available in the different societies are known and, (b) at the same time the proportions of people in each 'representative position' are not known'. (1973, 92) Instead, Barry claims, J and K have to decide which criteria are to be used to judge the institutions of whichever society it is that A and B are in. J and K know 'neither the proportion of the population in different "representative positions" nor what the levels of representative positions' (including the lowest) are in the society they belong to'. In this case, he says, 'There is complete symmetry between

the information bearing on the maximin criteria and the criteria (which is the only other one that Rawls seriously considers) of maximising average utility. The people in the original position know that if they choose the criterion of maximising the average each of them will gain the highest possible obtainable average individual expectation, while if they choose the criterion of maximizing the minimum each will guarantee himself the highest possible minimum individual expectation. The question then is, of course, whether it is rational to go for the highest average expectation or the highest minimum expectation.' As neither J nor K knows his attitude to risk, then, Barry concludes, all information 'which could provide a basis for making a rational choice of a criterion' has been removed.

(1973, 96)

Is this so?

J and K are to try to resolve conflicts of desire between A and B over arrangements of social and economic inequalities. A and B are in the same society. J and K do not know which of A or B they are. Suppose J and K are deciding which of two types of institution they wish to have in the society. Let's call these institutions V and W. Let's suppose that V and W have the same utility but V has the lower minimum. Let us suppose that, within the institution V, we can divide the representative positions into two, namely, the position occupied by A, and any other position which we will say is occupied by B. A's position has a higher utility than B's. Suppose, now, that A's utility in V is at B's expense, that is, that there could be an institution W where B would have a higher utility but A a lower one, while the average utility remains the same. The question is whether J and K will prefer V to W? We can allow J and K knowledge of the utility levels and minima of these institutions without biasing their decision by ensuring that they do not know which of A and B they are. We

can also allow J and K to know which representative role A and B occupy as long as we turn the situation into a two-party conflict where neither J nor K can calculate their chances of being either A or B.

As rational agents, J and K will go for whichever of V or W maximises their chances of getting what they want. Barry's problem remains, however. Without knowing their own attitude to risk there does not seem to be enough information for J and K to make a rational choice between V and W. This is because, as far as J or K know, A and B could both be high-risk takers who would prefer a system where they could either have great economic inferiority or great superiority to a system where they would have economic equality at a level much less than the economically successful in the other system. If J and K are low risk-takers, then they will maximin and most of Rawls' conclusions about social and economic management of inequality go through. If they are high risk-takers, then other arrangements for social and economic inequalities will be more rational than maximin. Whether the most rational one will be the principle of average utility is outside the scope of this thesis.

Rawls does not impose the different principle on the parties behind the veil but, rather, tries to derive it from the veil conditions. As we have seen, Barry believes he fails. I will now see whether impartialism is better able to derive the difference principle from its axioms, even given that J and K do not know whether or not they are high risk takers.

3. Power Relativities

As we have seen, J and K would agree to A and B having equal freedom-preserving rights. This is because both J and K want not be coerced more than they want to coerce. As maximisers of desire - satisfaction J and K

want A and B each to be free from coercion by the other to pursue the satisfaction of whatever desires they wish by whatever means they wish, provided that J and K have agreed that those desires and the means to these are ones that A and B should be allowed to choose. In order to ensure that B doesn't coerce A into maximising B's desire - satisfaction at the expense of his own, J and K will agree to the establishment of certain rights-protecting institutions. These institutions will themselves have coercive powers. These powers are intended to prevent B from coercing A or else to coerce B into restoring to A what is rightfully his if B has wrongly denied A his rights by B's greater coercive powers. It is essential if A and B are to have equal freedom-preserving rights in practice as well as in theory that the rights protecting institutions not be open to manipulation by one of the conflictants. If B, for example, is able to bribe the judiciary so that the judiciary uses the coercive power of the law to preserve B's wrongful possession of what J and K would have agreed was rightfully A's, then the rights-protecting institutions become rights denying institutions. As J and K do NOT know whether they are A or B, but they do know that they want NOT to be coerced more than they want to coerce, they will agree that B not be so much more powerful than A that he can subvert the course of justice. This will mean that B is not to be so much more powerful than A that he can permanently control the legislature and executive of the society because these bodies ultimately control what occurs in the courts and in the police force. Not only can they manipulate the law-enforcement agencies in favour of entrenched groups but they can consolidate the power and wealth of some groups by discriminating against others in legislation. J and K will agree that B is not to have this kind of power over A either. The wealthier B becomes relative to A, however, the greater his power to bias the rights-protecting institutions in his favour. When talking of 'power' and 'wealth' I will use Barry's account. He defines power 'as the capacity to get other people to do what you want'

and wealth 'as the ability to obtain goods and services that you want'.
(1973, 30)

This means that J and K will agree that the relative difference in wealth between A and B not be such as to enable B to bias the rights-protecting institutions in his own favour to any significant extent. This bias doesn't need to be an illegal bias such as the case mentioned earlier of bribing judges. J and K would be concerned about a legal system in which those court cases where one side would afford a QC and one couldn't regularly were decided in favour of the side with the QC. Wealth, however, is not the only method by which a group could bias the rights-protecting system in its own favour. Another could be by all the positions of power in the rights-protecting system being occupied by members of this group. This would only be a problem if this group either had a common interest in favouring its members against non-members, as it might if they were all part of the same family, or the same religion, or the same race, or whatever. It is this kind of possibility that Rawls' Principle of the Fair Equality of Opportunity is supposed to frustrate. J and K would agree with Rawls that neither A nor B be able to bias the rights-protecting system against the other by having a monopoly on the positions of power within the system. The rights-protecting system, then, can be subverted by methods other than inequalities in wealth. J's and K's reasoning about how to prevent such subversion by inequalities of wealth also applies to subversion by other sources of power.

J and K, then, will agree that B not be so much wealthier than A that he can use the rights-protecting system to coerce A to do what satisfied B's desires rather than what J and K have agreed he should be free to do to satisfy his own desires. One way of minimising B's capacity to do this is for A and B to be equally wealthy. J and K, however, would NOT require

that A's and B's freedom-preserving rights were protected by this particular strategy. They might be prepared to let B earn more money than A but only if this did not significantly increase B's power over A vis a vis the rights-protecting system.

This gives us a version of the difference principle, namely, that social and economic inequalities are to be arranged so that any improvement of the conditions of the better-off be accompanied by a betterment of the condition of the worse-off sufficient to ensure that the better-off have no significantly greater power over the rights protecting system than the worse-off. I will call this the 'relativities' principle.

This principle could be met in at least two ways. Firstly, by actually increasing A's wealth whenever B gets an increase in wealth so that B has no significantly greater power over the rights-protecting system than A. Secondly, by modifying the rights-protecting system so that B has no significantly greater power over it than A even though B is so much wealthier than A that had these modifications not been made to the rights-protecting system he would have had significantly more power over it than A. Possible ways of implementing this kind of strategy are by subsidising the poor when they employ lawyers so that they get as good representation as the rich, subsidising political parties so that the parties of the poor compete for votes on an equal footing with those of the rich, forbidding media monopolies, subsidising access radio and television, educating the poor in the processes by which they can protect their rights, and so on.

J and K, then, will at least require social and economic inequalities to satisfy the relativities principle.

B, however, can coerce A by means other than having greater power than A over the rights-protecting system. If B has greater wealth than A, B may be able to control A's means of livelihood so that A is dependent for his own, or his family's existence on B. This means that B can coerce A into not seeking redress through the rights-preserving system for B's infringement of A's rights by threatening to cut off A's means of livelihood. Likewise, B could threaten to cut off A's means of livelihood unless A uses his freedom-preserving rights such as the right to vote to hand over to B greater power over the rights-protecting system. As J and K want not to be coerced more than they want to coerce they will be more concerned to protect A from this kind of coercion by B than they will be to maintain B's superior wealth. There are two ways they can protect A from being coerced by B in this kind of way. Firstly, they can require that, if the society is wealthy enough, then it ensures that no-one is dependent in this way on anyone else, that is, it provides either a minimum wage or unemployment benefits or some such that is sufficiently high to ensure that B has little coercive power over A to get A either to surrender his right to use the rights-protecting system or to get A to use the rights-protecting system to confer greater power on B. Secondly, if the society is not wealthy enough to do this, they will insist that A and B be more or less equal in their financial insecurity, that is, that the difference in wealth between A and B is not such as to lead to A's becoming dependent on B. J and K, then, would adopt what I will call the 'independence' principle, namely, that social and economic inequalities be arranged so that A is not so dependent on B economically that he will surrender his equality with respect to the freedom-preserving rights or his recourse to the rights-protecting system in order to preserve his income.

The most, then, that can be derived from the impartialist axioms are the 'relativities' or 'independence' principles.⁽¹⁾

4. Maximising Income

For Rawls, J and K must prefer a situation where A has \$200 and B has \$300 to one where A and B have \$199 each, given that the spending power of each dollar is the same. The relativities principle does not require this. J and K are prepared to let A and B decide by some conflict-resolving process such as majoritarianism whether they will stay on \$199 each where neither has more power than the other to use the rights-protecting system to their own advantage or whether they will allow an inequality of wealth, and therefore of power, because this inequality of wealth generates greater scope for desire satisfaction for all. J and K will allow A and B to decide that the freedom-preserving rights are best protected by economic equality, even if this is at a low economic level. Admittedly, A and B may only be able to obtain the higher freedoms such as privacy, etc. if A and B go for economic inequality but J and K do not require A and B to prefer the higher freedoms to the lower ones. Provided that A and B have political equality in a tribal society J and K do not require that they introduce economic inequalities if this is necessary to move them to one that provides a wider scope for individuality, or a greater variety of consumer or 'cultured' satisfaction. Nonetheless, it may be the case that the move to greater scope for individuality is difficult to avoid once societies can no longer close themselves off from

(1) Impartialism, therefore, is not subject to those criticisms that are directed solely at the MAXIMIN ASPECT OF RAWLS' THEORY. Barry (1967) criticises MAXI-MIN as the rational strategy under the conditions Rawls specifies. Wolff (1977) gives a detailed criticism of the various versions of MAXI-MIN that Rawls proposes. MAXI-MIN is defended in Corrado (1980).

other societies. J and K would then have to devise principles governing the relationships between societies that resolved conflicts of desires between societies by non-coercive means. This may mean an educative process within each society that undermines the conformity of the culture and leads to pressure from politically equal citizens for the goods, or freedoms, available to citizens of other societies. While this may be required by J and K once there is a conflict of desires between societies, impartialism does not require it of a self-contained society provided all its citizens are political equals in that their equal freedom-preserving rights are equally protected by the rights-protecting system.

5. Liberty and Equal Worth of Liberty

Rawls makes a distinction between 'liberty' and 'the worth of liberty'. He says, 'Liberty is represented by the complete system of the liberties of equal citizenship, while the worth of liberty to persons and groups is proportional to their capacity to advance their ends within the framework the system defines. Freedom as equal liberty is the same for all; the question of compensating for a lesser than equal liberty does not arise. But the worth of liberty is not the same for everyone. Some have greater authority and wealth, and therefore greater means to achieve other aims.' One has a lesser worth of liberty, for example, when one has a greater 'inability' to take advantage of one's rights and opportunities as a result of poverty and ignorance, and a lack of means generally.

(1972, 204)

Impartialism, however, sees the worth of liberty and liberty itself as inextricably linked when it comes to the freedom-preserving rights and the rights protecting system. The rights-protecting system must ensure that A's and B's freedom-preserving rights are of equal worth. J and K

will not agree to any social or economic arrangements that give B greater power over the rights-protecting system than A. Where B has unequal rights within the rights-protecting system, e.g. the rights of police, or parliamentarians, etc., these need to be stringently controlled so that abuses of them are quickly detected and severely dealt with.

While the rights-preserving system must ensure that A and B have equal worth of freedom-preserving rights, impartialism does not require that there be a system that ensures that A and B have equal worth of freedom if this means that A is able to get as great a proportion of his desires satisfied as B. As we have seen, J and K may allow B to earn more than A provided (1) B does so by non-coercive means, (2) B's extra wealth is within the range permitted by the relativities and independence principles. If B's extra wealth meets these conditions then B has greater worth of freedom than A. He is no freer than A. This must be so if the relativities and independence principles have been met. He is no freer than A because he is no more able to coerce A into doing what satisfies B's desires than A is able to do this to him. Nonetheless, he is able to satisfy more desires than A. It is possible, then, that J and K accept the notion that B deserves to satisfy more desires than A. Many critics of Rawls have been concerned at his apparent rejection of 'desert'. I will take this up in more detail in a later chapter.

What we have, then, is that J and K will agree (1) that social and economic inequalities between A and B not be the consequence of coercion, (2) that the freedom-preserving rights of A not be rendered of unequal worth due to either B's being comparatively so much wealthier than A that he can control the rights-protecting system in a way that A cannot neutralise OR that A is so dependent economically on B that A cannot afford to antagonize B by making full use of the system to protect his rights.

J and K, however, are mutually disinterested desire-satisfaction maximisers. Within the limits imposed by (1) and (2) above, they will want as many of their desires satisfied as possible. As we have seen, Rawls believes this commits J and K to a kind of economic progressivism, that is, that J and K will always agree to B's becoming wealthier if he thereby also makes A wealthier, even if A is less wealthy than B. This aspect of Rawls' view is a consequence of his notion of the primary goods. In the next chapter, I will examine whether or not impartialism needs a theory of primary goods or whether it only needs the notion of maximising desire-satisfaction as restricted by the impartialism axioms.

7. The Lexical Ordering

I have argued so far that J and K would limit social and economic inequalities in ways that maintained the equality of freedom-preserving rights or maintained the equal worth to each of them of the rights-protecting system. I have, then, agreed with Rawls that liberty has priority over wealth. I have not, however, considered the case where B offers A enormous wealth to get A to voluntarily surrender his equality in either the worth of equal rights or the rights themselves. I have argued so far only about the degree to which J and K would insist on an equality of wealth between A and B. It is possible, however, that A values wealth more highly than his rights. I have been arguing from the contrary assumption that he values his rights higher than his wealth. Rawls believes that his two principles of justice rule out A's swapping his freedom (right) for more wealth. He says -

'Now it is possible, at least theoretically, that by giving up some of their fundamental liberties, men are sufficiently compensated by the resulting social and economic gains ... Imagine ... that men forego certain political rights when the economic returns are significant and their capacity to influence policy by the exercise of those rights would be marginal in any case. It is this kind

of case which the principles as stated rule out;
 being in serial order they do not permit exchanges
 between basic liberties and economic gains.'
 (1972, 62-3)

There are two ways in which there can be a gain in liberty or rights. The rights protecting institutions (1) can protect freedom more effectively, e.g. they could become more efficient at detecting and convicting the guilty, (2) they can protect freedom no more effectively but can protect it more equally. J and K will be primarily concerned that the rights-protecting system protects the freedom-preserving right equally between A and B. They will be concerned only secondarily that the rights-protecting system also protects rights more effectively. They will treat wealth as having a value infinitesimally small when compared with that of the equality of the freedom-preserving rights.

To show why this is so, I will consider the kind of case this issue poses for our velorian J and K. The situation as described by Rawls is one where B has offered A social and economic gains if A will give up some of his fundamental liberties, or what I have called the freedom-preserving rights. If A accepts this offer, then B becomes more able than A to use the coercive powers of the rights-protecting system to resolve conflicts of desire in his own favour. In offering A such a deal, B presumably is after precisely this kind of power over A. A would be handing B the power to go back on the deal, thereby depriving A of his economic gain. A would have no court of appeal. J and K would not agree to A's accepting this offer in any situation where the only plausible reason for B's making the offer is to enable B to advantage himself at A's expense. J and K, therefore, would agree that wealth is of infinitesimal importance compared to the freedom to utilise wealth once you have it.

If the society has to drop from general affluence to general poverty to get an effective equality of worth of rights, then J and K would choose to do so, given that the alternative is the case where B offers A general affluence in order to gain irreversible power over A. It is irrational of A to try to maximise his desire satisfaction by giving B the power to thwart this when B makes A an offer whose only purpose is to maximise B's own desire-satisfaction at A's expense.

When it comes to protecting the equal rights more efficiently, however, J and K will have to weigh up the kinds of factors discussed in the chapter on the derivation of the priority of liberty. They will have to balance the cost of a more effective system against the desirability of its increased effectiveness. In this kind of case, they may well conclude that a minute gain in freedom from the risk of being unjustly treated, a risk shared equally by all, is not worth the drop from general affluence to general poverty.

Impartialism, then, is not subject to the kinds of criticism that Barry makes of Rawls, even though it shares Rawls' view that liberty has special priority over wealth, at least in the sense that equality of liberty is worth the drop from general affluence to general poverty. It is difficult to imagine, however, a real case in which such a demand would be made of a society.

Chapter 12

The Primary Goods1. Introduction

Behind the veil, the velorians J and K do not know which of the conflictants A and B they are, therefore they do not know which desires they have.

Conflict 1 could be between A who wants to practice homosexuality and B, who doesn't want him to. Conflict 2 could be between A2 who wants to own the same piece of land that B2 wants to own, and so on.

If it is part of A's conception of the good that he likes vanilla icecream whereas it is part of B's that he likes strawberry icecream, this doesn't constitute a conflict of desires. It is only when A wants to force B to eat vanilla icecream, or when resources are limited in such a way as only one flavour icecream can be made, that there is a conflict of desires.

When there is such a conflict of desires, however, neither J nor K knows which side of the conflict he is on. This makes the range of possible conflicts of desires enormous. For any desire that a person A has, there could be a person B who either wants A not to satisfy their desire, or who has some other desire whose satisfaction is incompatible such as A's. It would simplify J's and K's tasks of finding principles to resolve conflicts of desire if this range of desires could be reduced.

Rawls suggests that the range be reduced in two ways. Firstly, that J and K only propose principles that resolve conflicts of desire over primary goods. Secondly, that J and K resolve conflicts of desire in accord with the principle that the desire that exercises a person's capacities more, or exercises more complex capacities, is to be preferred. This automatically resolves many conflicts of desires, leaving J and K only those conflicts of

desire to resolve where either side exercises capacities equally or exercises equally complex capacities. Rawls calls this principle the 'Aristotelian' principle. Both these ways of reducing the range of conflicts of desires that J and K are to consider do not result from using the veil of ignorance, however. They are not general conclusions about the kinds of principles that resolve conflicts of desires that have been reached as a result of applying the veil to J's and K's reasoning. Rather, they are limitations externally exposed on J's and K's reasoning. As a consequence, there will be two kinds of conflict of desire which they do not resolve, namely (1) the conflict between those who want conflicts of desire resolved in terms of primary goods and those who don't and (2) the conflict between those who want conflicts of desire resolved by using the Aristotelian Principle, and those who don't. Rawls sets up his theory so it favours one side over the other in each of those conflicts but, as his preference is not itself derived from the veil, it may be a partialist preference.

I will now look at the primary goods, then at the Aristotelian Principle, to see whether they are consistent with impartialism and, if they are not, what consequences this has for the kinds of principles J and K would choose to resolve conflicts of desires.

2. Primary Goods

Rawls defines the primary goods as 'things that every rational man is presumed to want'. (1972, 62) He asks us to assume, for simplicity, that 'the chief primary goods at the disposal of society are rights and liberties, powers and opportunities, income and wealth' and self respect. These are the 'social' primary goods. He also lists a number of 'natural' primary goods, namely, 'health and vigor, intelligence and imagination'. These are 'natural' because, 'although their possession is influenced by

the basic structure, they are not so directly under its control'.

The question for impartialist theory, then, is whether or not conflicts of desires are to be resolved only by principles that deal with these primary goods, or whether there is something other than primary goods in terms of which principles should be formulated.

Barry (1973, 22) suggests that Rawls has introduced the primary goods into his theory because, otherwise, the parties behind the veil would have to formulate these principles in terms of want - satisfactions. Rawls, he says, is in an awkward dilemma. 'On the one hand, he does not like the implication of the want-regarding view, and wishes, for example, to say that the desire of someone to practice his religion freely should have priority over the desire of another man to stop him, even if the second man's desire is more intense than the first's, or if those who want to worship are outnumbered by those who want to suppress and each person has the same intensity of desire for what he wants.' I will call this the first horn of Rawls' dilemma. Barry continues, 'Yet at the same time he wishes to derive principles of justice from an original position, which, by denying the actors specific information about themselves, seems to lead inexorably towards the formulation of principles in want-regarding terms.' Barry adds, 'In my view the connection is inexorable.'

Impartialism, however, is able to avoid the first horn of the dilemma without introducing the primary goods. It can remain a purely 'want-regarding' theory. This is because, as we have seen earlier, the principles J and K are to adopt have to be symmetrified ones, that is, the principle that A's religion is to be suppressed because B's desire to suppress it is stronger than A's desire to practice it can only be agreed to by J and K if B is prepared to allow B's religion to be suppressed if A's desire to

suppress it is stronger than B's desire to practice it (or some desire of equivalent strength). What matters, then, is not that B's desire to suppress A's religion is stronger than A's desire to practice it but, rather, that B's desire to coerce A is stronger than B's desire not to be coerced by A. Even if B meets this requirement, however, his desire to suppress A's religion is ruled out by the noncoercionist requirement that J and K reason as if they wanted not to be coerced more than they want to coerce. Impartialism, then, achieves the result that Rawls wants without bringing in primary goods, and it does so without weakening or abandoning the apparatus of the veil of ignorance.

The second horn of Rawls' dilemma is not a problem for the kind of theory that I have advanced so far, as I have formulated its principles in want-regarding terms.

3. The Aristotelian Principle

Rawls put the Aristotelian Principle as follows: 'Other things equal, human beings enjoy the exercise of their realised capacities (their innate or trained abilities), and this enjoyment increases the more the capacity is realised, or the greater its complexity. The intuitive idea here is that human beings take more pleasure in doing something as they become more proficient at it, and of two activities they do equally well, they prefer the one calling on a larger repertoire of more intricate and subtle discrimination'. (1972, 426) For example, 'some one who can play both chess and checkers 'generally prefers 'chess to checkers'.

Barry sees this as 'Rawls' most important attempt to set substantive limitations on the possible patterns of human desire.' (1973, 27)

Barry then proceeds to show that either the Aristotelian principle is an empirical generalisation or else a partially constitutive definition of

of 'rationality'. If it is the latter, then, Barry shows that, 'contrary to Rawls' professed intention, a substantive idea of human excellence is being advanced under cover of a neutral-appearing concept of rationality.' (1973, 28)

I will not repeat Barry's arguments here but will merely note that impartialism requires no more of the term 'rationality' than Barry gives it, namely, 'that a man's plan of life is rational if it results in his getting more rather less of whatever things it is that he happens to want'. (1973, 27)

Whether the Aristotelian principle is true of human beings or not doesn't really matter from the viewpoint of impartialism. What does matter is that the parties behind the veil not be required to reason as if it is true. If it is true, then the velorians will take this into account in their reasoning about how A and B should act. If, however, it is not true of either A or B, then J and K should not reason as if it is true. If the Aristotelian principle is imposed on the veil rather than being an actual condition of the state of affairs about which J and K are reasoning, then J and K would be reasoning as if they know they favoured the more complex over the less complex, etc.

While this produces the result Rawls is after, it doesn't show that those who reject the Aristotelian principle are partialists. In fact, Rawls' own result could be the partialist one. To show that it is not, he would need investigate whether J and K would agree on his result if they didn't know whether they were A who accepted the Aristotelian principle or B who didn't. This means they need to arbitrate between A and B from behind the veil, and do so without begging the truth of the Aristotelian principle.

Haksar claims that, without the Aristotelian principle, Rawls would not be able to show why the parties behind the veil would value autonomy or why they would give liberty priority over income. While this may be true of Rawls' theory, it is not true of impartialism. We have already seen that autonomy can be derived from the three impartialist axioms. We saw in the last chapter that utilitarians will give the freedom-preserving rights priority over income. Their reason for doing so, however, is not because it is part of their ideal that, in Hart's words, each person be "a public spirited citizen who prizes political activity and service to others as among the chief goods of life and could not contemplate as tolerable an exchange of the opportunities for such activity for more material goods or contentment" (1975, 252). Rather, they merely want these freedoms as guarantees that they will be able to satisfy whatever wants they happen to have, including their desire to live in accordance with a public spirited ideal if they have it. Nonetheless, it remains the case that they do give the equal basic freedoms priority over such other goods as wealth, so Rawls' major claim remains intact.

4. Wants as the Units of Social Evaluation

Barry comments that, 'To the best of my knowledge, the case against treating wants as the units of social evaluation has never been set forth systematically'. (1973, 21) He suggests three reasons why so many people are opposed to a want-regarding view. These are

- (1) 'It may be felt that there is something fundamentally wrong in treating alike for the purposes of calculation such things as wants for personal gratification of oneself, the desire to give others pleasure, ambitions to contribute to the world's stock of truth and beauty, aspirations towards spiritual improvement and enlightenment, and so on.'

- (2) 'It may be considered inconceivable that any set of principles which does not discriminate among wants (except of course in terms of relative intensity) could give rise to implications that could be conscientiously endorsed.'
- (3) 'One might actually carry on the exercise of constructing the most hopeful-looking set of want-regarding principles one can think of and discover that when this set of principles is applied to actual or hypothetical situations the implications are indeed morally unacceptable.'

I will look at each of these in turn to see whether they constitute reasons for rejecting impartialism.

5. Immoral Desires

The first objection that Barry mentions is really the view that there are some desires that shouldn't even be taken into account in calculating what should be done. As Rawls says, 'desires for things that are inherently unjust, or that cannot be satisfied except by the violation of just arrangements have no weight'. (1972, 261)

Rawls' concern here is with a theory like utilitarianism which admits into its calculations as to which act is the right one any desire whatever, including desires to be cruel, vicious, dishonest and so on. Utilitarianism has a procedure for determining which acts are right or wrong. It performs calculations in terms of the consequence of the various desires from the viewpoint of maximising happiness, or whatever particular interpretation is given to utility. It does not exclude any desires from this procedure. It can be contrasted with a theory which has a procedure for deciding

which acts are right or wrong but only applies this procedure to a certain domain of acts. The procedure itself does not identify this domain. The domain is determined antecedently to any application of the procedures in accordance with criteria quite independent of the procedure. Rawls sees his own hypothetical contractarianism as a theory of this kind. As I have shown in the early chapters of this thesis, impartialism is unlike Rawls' theory and like utilitarianism in this request. It does not say in advance that there is any desire whose conflict with other desires it will not consider. The impartialist axioms are to be applied to any conflict of desires whatsoever, not just to ones selected by some independent criterion. Nonetheless, the effect of the impartialist axioms is to eliminate egoistic and coercionist desires from any calculation as to what is to be done. In this way, impartialism differs from utilitarianism. Utilitarianism not only applies its procedure to all desires, it actually takes all desires into account in its calculations about what should be done. Impartialism, on the other hand, applies its procedures to all desires but the effect of these procedures is to isolate out those desires suitable for the calculation of what is to be done. Unlike Rawls' theory, however, it does not isolate these desires by a criterion independent of its own procedure. Rawls' theory poses two problems of justification. Firstly, justification of its procedure for deciding right and wrong. Secondly, justification for its limiting the scope of these procedures to some desires rather than others. Impartialism has only the first of these problems. Impartialism, then, has the virtue of utilitarianism in that it deals directly with desires rather than needing some other, possibly suspect, alternative such as primary goods. It has also the virtue of Rawls' hypothetical contractarianism in that it does not admit morally reprehensible desires into its calculations of right and wrong.

6. Relative Intensity

Impartialism, like utilitarianism, only discriminates between wants in terms of relative intensity. Unlike Rawls' theory, it does not reject certain principles because they fail to meet the formal constraints of the concept of right. Nonetheless, impartialism differs radically from utilitarianism in how it deals with the relative intensity of desires. It is concerned with the relative intensity of the desires to coerce and NOT to be coerced but it doesn't measure A's desire NOT to be coerced against B's desire to coerce. It then requires that principles which rely on A's desire to coerce being greater than A's desire NOT to be coerced be eliminated from the calculations. Whether or not the principles it eventually arrives at could not 'give rise to implications that could conscientiously be endorsed' is best tested by seeing what implications for action its principles actually do have. This test would also meet Barry's third point.

The objections, then, to wants as the basis for calculating which acts to do have to be stronger than the ones Barry lists before they constitute grounds for rejecting impartialism merely because it does use wants rather than primary goods as its basis for calculation. I haven't so much rebutted these objections to wants as the basis of calculation of the right as shown that impartialism bypasses them. They are objections that are met by producing a theory that does the kinds of things the objections suggest can't be done.

It is also the case that impartialism bypasses the problems utilitarianism faces with the 'interpersonal commensurability of utilities'. It judges principles in terms of the intensity of A's desire NOT to be coerced against the intensity of the desire to coerce of the same person. If the

latter is greater than the former, then there is no problem of comparing A's preference to coerce with B's preference not to be coerced because impartialism has excluded A's preference from any consideration in the determination of principles for resolving conflicts of desires.

I have now concluded my discussion of the major substantive conclusions that Rawls derives from the veil of ignorance. Before I look at the meta-ethical questions raised by a contractarian theory such as Rawls', or my own impartialism, I wish to consider the charge that the veil of ignorance of its very nature falsifies the crucial moral notions of 'desert' and 'entitlement'. I will look at Flew's arguments to this effect.

Chapter 13

Impartialism, Entitlement and Desert1. Introduction

In his article, 'Who Are the Equals?' Antony Flew charges that John Rawls is committed to the following theses:

Firstly, that 'justice prescribes universal equalisation'.

Secondly, that 'justice presupposes that in our deserts and entitlements we all are (normatively) equal'.

Thirdly, that our 'all being normatively equal presupposes that we all either have been or are, in whatever the relevant respects may be, equal (in fact)'.

Fourthly, that 'the characteristics in respect of which we can rely on all human beings always being in fact equal must be precisely and only our defining characteristics as human beings', thereby 'dismissing all our individual and differentiating characteristics as morally irrelevant'. (1980, 136)

Flew takes exception to these theses because he thinks that the term 'justice', 'as traditionally understood, demands that everyone should have their own, their due'. Further, he thinks it to be a correspondingly conceptual point that 'all deserts and entitlements have to be grounded in some facts about the people so endowed'. (1980, 136)

In order to defend 'justice' as traditionally understood, Flew sets out to show that the four theses listed above rest on two assumptions, both of which are wrong. The first assumption is that 'no one can be entitled to anything they have not deserved'. The second is that 'nothing can be either earned or deserved unless everything which makes the earning or the deserving possible was itself earned or deserved'. (1980, 140)

Flew's enterprise raises the following questions for impartialism:

- (1) Is impartialism committed to these four theses? (2) Are these theses false? (3) Do these theses make the assumptions that Flew thinks they do? (4) Are these assumptions false?

2. The Argument From Constitutive Bodily Parts

Flew believes that these theses are false because they rest on the false assumption that no one can be entitled to anything they haven't deserved. Actually, although Flew believes this assumption to be wrong, he admits that 'it is possible to argue only ad hominem' against it. Rawls, he believes, cannot base any conclusions on the assumption that a person can only be entitled to something if he deserves it because Rawls' own theory presupposes that the contractors behind the veil of ignorance are at least entitled to their bodily parts, even though they have done nothing to deserve them. Rawls, then, 'is in no position to deny unearned and undeserved entitlements', if Flew is correct that Rawls presupposes the contractors to have a right to their bodily parts. But does Rawls presuppose anything of the kind?

Flew's argument that he does is as follows:

'...it will take a very far-gone collectivist to deny to individuals their rights to their own constitutive bodily parts. Suppose that half the population is born with two normal eyes, and half with empty sockets. Suppose too that eye transplants became possible, safe and easy. Now, is it a matter of simple justice that all the two-eyed must yield up one eye each to the transplant surgeon, or would any such transfer be a deed of supremely generous, overflowing charity?'
(1980, 141)

Although my concern is to see whether impartialism succumbs to Flew's objections, I will look in detail first at how Rawls' form of contractarianism would deal with it. Flew's essential objection to Rawls' theory is the veil of ignorance which, he believes, turns moral argument into a debate between 'epistemological ignoramuses'. As my theory, like Rawls', utilises the veil then, if the veil must presuppose rights to bodily parts, my theory must do so as much as Rawls does.

Flew's point is that contractarian theories must presuppose some rights that are not themselves derivable from the contract conditions. His actual argument does NOT show this but, rather, is not so much an argument as an expression of disbelief that anyone could believe that a right to bodily parts might be the outcome of a contract, or that a contract could be the basis for a denial of such rights. His abhorrence of the forcible transplanting of bodily parts, however is likely to be shared by many people. If it is, this will make them reluctant to accept a moral theory that denies any such right. This raises two questions. Does impartialism really deny such a right? Secondly, if it does, then is this denial sufficient grounds for rejecting it.

I will now turn to the first of these questions.

3. What is a Constitutive Bodily Part?

Presumably Flew's use of the adjective 'constitutive' is meant to differentiate constitutive bodily parts from those bodily parts that are not constitutive. In one sense, of course, any bodily part is constitutive in that, while you have it, it constitutes part of you. You are less a complete person for the loss of it in that you are now less efficient with respect to those functions it performed for you at its healthy best. In another sense, however, very few parts of you are constitutive in that

you would no longer be you without them. After all, you retain your personal identity even after a heart-transplant. The only real candidate for a constitutive part in this sense is your brain.

Flew's argument for constitutive bodily parts is meant to show that the parties to the contract have a right to these bodily parts prior to the contract. Perhaps constitutive bodily parts are those bodily parts a contractor must have if he is to be capable of making a contract at all. If this is so, then Rawls needs to concede the contractors very few 'constitutive' bodily parts. The contractors need be no more than brains in bottles with telepathic powers or wired up to a sufficiently sophisticated communication system. Even if the contractors need somewhat more serviceable bodily equipment, this only shows that the contractors must have certain bodily parts prior to the contract. It doesn't show that they must have a right to these parts. Being 'constitutive', then, does not seem to make a bodily part into one to which you have an especial right.

Flew's abhorrence of enforced bodily transplants, therefore, would appear to be merely an intuition, as he has supplied no adequate argument in its support. This abhorrence is perfectly understandable. Most of us are utterly appalled at the idea of losing the sight of one of our eyes. We are protected from having guilty consciences about those who lack the sight of both eyes by the fact that modern medical technology cannot safely, painlessly and reliably perform such transplants. While these considerations may explain our aversion to enforced bodily transplants, they do nothing to show that such a sentiment is morally commendable.

Purely, as an intuition, it is no better and no worse than an intuition to the contrary. If ethics is to be more than a matter of personal taste, we need to find some way of deciding which of these two intuitions is to

be preferred in those cases where they conflict.

Rawls tries to provide such a decision-procedure. It at least gives us a basis on which to judge his intuitions. We can see what underlies his conclusions and what modifications have to be made to arrive at a better theory. If he had just given us his intuitions, or discrete arguments for various of his intuitions, this would NOT have been possible.

4. Rawlsian Rights to Bodily Parts

Even Rawls' theory does not lead to a complete denial of rights to bodily parts. In order to determine whether or not the Difference Principle would give people a right to their constitutive bodily parts, the worst outcomes for each option need to be calculated from the viewpoint of the contractors behind the veil. When dealing with the distribution of unearned goods, the contractors' application of maximin will require an equal distribution. This can be seen from the classic case of dividing up a cake. If each person wants at least his equal share of the cake, then the least bad of the worst outcome of any option is to get an equal-sized slice. It may not be as much as you wanted but, as you don't know who would get any bigger than equal slices if these were any, then it is rational to insist on an equal distribution. The parties behind the veil do not know whether they will have two eyes, one eye or no eyes in real life. They must decide between institutions which leave them with their natural allocation, or which may require them to donate one or both eyes to someone else if they are lucky enough to be born with two. The contractors will agree that, in general, they each would be best off with two eyes and worst off with none. It seems then that it is a worse minimum to have no eyes than it is to have one. This means that, if A has two eyes and B has none, the contractors behind the veil would require A to donate an eye to B, provided

all the conditions Flew gave with regard to guaranteed success, painlessness etc., were met.

This argument, however, has dealt only with shareable bodily parts, such as eyes, and possibly kidneys, lungs and other bits that are in pairs. Can it be extended to unshareable bodily parts like hearts or livers? Who it is that possesses a good heart is determined by the natural lottery. If A has a good heart and B has one that is about to fail, would the contractors behind the veil of ignorance be prepared to make A donate his heart to B under any circumstances? So far, the contract has not shown that A has any right to his heart. He no more deserves to have a good heart than B deserves to have bad one. He is not entitled to keep it, he has not merited it or earned it. How could the contractors decide who should have the heart? Any of them could be either A or B so they have no reason to favour one rather than the other. Say they decided to settle who should have it by a lottery, as the only fair means available. Such a lottery, however, is redundant. The heart was already allocated in a purely chance way, purely chance, that is, in the sense that the allocation of hearts in the natural lottery was not the result of any deliberate favouring of one person over another by some Father Christmas with a spare heart or two to give away to his chums. If a new lottery is set up to overrule the original distribution, then why not a third, and a fourth? Where you want to stop the regress of lotteries will depend on whether you happen to be in possession of the heart or not at the end of the last one. As rational agents, the contractors will not start such a regress. Instead, they will contract to obey the principle that everyone has a right to the unshareable bodily parts with which he is born. This kind of consideration will also prevent those born with one eye having the right to force those born with two eyes into a new redistribution by lottery.

At the most, then, Rawls' theory denies people a right to their shareable bodily parts. This is, however, merely denied in principle. In practice, the considerations involved in determining when such transplants were justifiable would be much more complex than the simple maximin situation I sketched out earlier. The crucial decision the contractors have to make is whether it really is a better minimum to remain blind than to allow the transplant of one eye from those with two good ones. How good, for example, does the success rate need to be? If you are totally blind, it would be good to gain an eye but if you had two eyes it is not only bad to lose one, it is far worse to lose one for no purpose because the transplant fails. Would you prefer to stay totally blind rather than have a 9 in 10 chance of losing a good eye for no benefit to anyone? What risk would you be prepared to tolerate? The actual maximin calculations across all these factors would not be an easy one, making it very likely that the contractors would be rather conservative about tampering with the natural lottery. It may even be that these factors weigh sufficiently for the parties to contract to give you a right to your shareable bodily parts, but this can't be determined without doing the detailed calculations. On the face of it, however, the benefit of ceasing to be totally blind is so great that it is unlikely that the countervailing disadvantages will be sufficient to rule transplants out in principle. For those who find the possibility of bodily transplants utterly repugnant, the fact that Rawls' theory probably allows them in principle will constitute an overwhelming objection to his theory, just as those who find slavery utterly repugnant reject utilitarianism because it appears to allow slavery in principle, if not in practice. This repugnance, however, hardly constitutes an argument against bodily transplants. It may be the case, however, that Rawls' theory may not even merit this repugnance. We have considered the Maximin solution for distribution of bodily parts solely from the viewpoint of the distribution of goods other than liberty. It may be, however, that

this distribution leads to an inequality of liberty which is not able to be justified by the increase in liberty it produces for those with the least liberty.

5. Liberty

Rawls claims that liberty can always be explained by reference to

- (1) the agents who are free
- (2) the restrictions or limitation which they are free from and
- (3) what it is they are free to do or not to do (1972, 202).

Rawls requires that the basic liberties must be assessed as a whole, as one system. With respect to the transplant of bodily parts, the contractors have to consider the effect on the liberties of real people of any institutions requiring that parts be transplanted, forceably if necessary. If one begins with the case of adults, those with two eyes will resist any attempts to deprive them of one. Blind adults, while realising the benefits of sight, may be more reconciled to continuing in a blind condition than people with two eyes are to losing one. The system of coercion needed to implement the transplants may have to be so severe as to not in fact increase the freedom of the least free, if not being able to see is to count as a lack of freedom and they are regarded as the least free. That is, the freedom made possible by gaining at least the sight of one eye may not compensate for the freedom lost by a social system that can only provide the freedom to see by denying many other freedoms.

There are other costs, too. A society with no compulsory transplants is likely to be one in which people are free from being regarded as walking organ banks. It is likely to be free from criminal trades in organs, and

so on. These would need to be taken into account in the contractors' maximin calculations.

Some of these problems, however, may be removed if transplants occur as far as possible at birth. A baby with two eyes who loses an eye to a baby with no eyes will become accustomed to his condition as an adult. He will feel his loss less than an adult who knows from experience what it is like to have two eyes. There will still be trouble from parents trying to stop their child being the one to lose an eye. Preventing this will require coercive systems that may restrict more liberties than are gained. Not all cases of where a transplant is needed, however, will be solved by using the bodily parts of babies. If an adult loses both eyes, the transplant will have to come from an adult, so the kinds of coercion mentioned earlier may still be necessary, possibly leading to less liberty overall for the worst-off. This could be overcome if all babies lose an eye soon after birth, these eyes being cultivated until needed for transplant. This, however, contravenes the requirements of the Principle of the Greatest Equal Liberty in that societies where most people have two eyes and some have one will create more liberty for those with only one eye than societies where everyone has only one eye. So much cannot be done as well by one-eyed as by two-eyed people. All sorts of modern freedoms dependent on technology would be at risk - freedom from disease, freedom from drudgery, freedom to travel, and so on, would be lost. As the contractors are not envious, they would not sacrifice these freedoms merely to gain equality.

Rawls' theory, then, might well give us a right to our bodily parts depending on whether such a right gives more freedom to the worst-off. As it turns out, this decision-procedure does not result in an emphatic endorsement of Flew's intuition, whereas another decision procedure may. The obvious alternative, utilitarianism, seems to me to be unlikely to

give Flew's intuition emphatic endorsement either. Flew offers no decision-procedure for arbitrating between competing intuitions. His approach, therefore, is not very helpful when what Hare calls Level 2, or critical, thinking needs to be done. We would need to know the ramifications that sticking to Flew's intuition is seen by him to have for other moral issues before we could think critically about whether to prefer his theory to Rawls.

Nonetheless, it may be case that it doesn't require very much critical thinking to see that a theory must be wrong if it assumes that a person is only entitled to something if he first deserves it, e.g. one would never be entitled to anything.

We have seen that Rawls does not make the particular assumption that we are entitled to our bodily parts but it may just be that Flew has chosen an unfortunate example. In fact, Flew set out to establish this right to our bodily parts in order to show that we have an entitlement to our undeserved natural talents. He argued from bodily parts to natural talents as follows -

'Yet if once we allow to individuals, as we surely should, their rights to their own constitutively bodily parts, then it must be excessively hard to deny to them parallel rights in their own, and similarly constitutive, natural talents and temperaments.' (1980, 141)

As Rawls is not committed to the existence of rights to bodily parts the above argument loses its ad hominem force. Even if he were committed to such rights, it's by no means obvious that bodily parts and natural talents are sufficiently similar to allow the parallel to be drawn. Bodily parts and talents, etc. are not alike in that if I have one you can't. If A can play the piano and B cannot, A is not able to gain this ability by depriving B of it, as he can gain the ability to see by depriving B of an

eye. Rawls' contractors, in fact, are much more likely to agree that there is a right to one's natural talents than they are that there is a right to shareable bodily parts. As none of them knows what talents he will have, he will want to ensure that he gets maximum benefit out of whatever ones he has. As he cannot gain talents by depriving others of them, the only objection he can have to their talents is that these put his maximising of his own values at risk. It is not their possession of these talents, however, that constitutes this risk but the use to which these talents are put. He will, therefore, want limitations on how talents are used. If A has an IQ of 150 then the contractors will not want him to use it to cheat B but they will want him to use it in ways that benefit both himself and B. If they turn out to be A they will not want to lose it just because they might use it to cheat someone. They would, therefore, agree that everyone had a right to own their natural talents, but not a right to use these talents as they wished. The same would be true of personal proclivities, such as homosexuality. The contractors would not give society the right to deprive people of these inclinations by force.

Flew, then, had no need to appeal to the argument from constitutive body parts to show that Rawls was committed to a right to our natural talents. This right is derivable from Rawls' own contract theory. It seems, then, that Rawls must admit at least one case of being entitled to something not deserved, namely, one's natural talents. In fact, Flew's attempts to show that Rawls was committed to unearned entitlement was quite unnecessary, as it is a consequence of Rawls' theory that people are entitled to whatever the contractors behind the veil of ignorance agree is their entitlement, regardless of whether it is deserved or not.

We have already seen, for example, that the contractors would unanimously confer on each person an entitlement to his own unshareable bodily parts. They would even confer an entitlement to one's own shareable bodily parts except under certain very special conditions. These entitlements hold even though none of us deserve to possess these bodily parts, as this entitlement is conferred prior to any particular acts of ours that might have earned merit.

In a way, however, the fact that Rawls and Flew agree on this issue will be small comfort to Flew, because Rawls still derives these entitlements from the consensus of contractors behind the veil of ignorance. There is still a sense, then, in which Rawls' theory of justice 'prescribes universal equalisation' even though he and Flew are in agreement on undeserved entitlement.

Does Flew fare any better on what he believes is the other assumption of Rawls' theory, namely, that nothing could be earned or deserved unless everything which made the earning or the desert possible was itself deserved or earned?

6. The Presuppositions of Desert

Once again there is no need for Flew to show that there are cases of desert in which everything that made the desert possible was not itself deserved. Rawls would not disagree with him. As we have seen, Rawls believes that no one is entitled to a better than equal share unless this benefits everyone else as well. Imagine, then, two people, person A with an IQ of 150 and person B with an IQ of 100. Neither deserves the IQ they have, any more than people born with two eyes deserve their good fortune. Nonetheless, it may be the case that people with an IQ of 150 are able to

improve everyone's standard of living, provided they are given an incentive to do so. For example, if they are paid \$500 a week, they work so much more productively that everyone else's real income increases from \$200 a week to \$300.

Given that this financial advantage has no nasty disadvantages, all Rawls' contractors will agree to this inequality. It is now possible for A to deserve \$500 a week whereas it is not possible, given the story so far, for B to deserve \$500 a week. Even so, if A has an IQ of 150 and does what contributes to the common good, whereas C has an IQ of 150 but does no more than B, then C's mere possession of an IQ of 150 does not entitle him to \$500 a week. It seems then, that if A does what is necessary he deserves his \$500 a week even though his IQ of 150, which made this earning or deserving possible, was not itself earned or deserved.

7. Universal Equalisation

Rawls, then, is not committed to the view that no one can be entitled to anything they have not deserved nor that nothing can be earned or deserved unless everything which makes the earning or deserving possible was itself earned. Insofar as Flew rejects Rawls' theory because he thought it rested on these assumptions, this rejection is unwarranted.

In the light of these criticisms, I shall now look at the four theses to which Flew believes that Rawls is committed.

In summary, these theses are that our individual and differentiating characteristics are irrelevant to our deserts and entitlements. Rather, our deserts and entitlements are normatively equal because the only morally relevant characteristics for determining our deserts and entitlements are

those in which we are all factually equal, namely, our defining characteristics as human beings.

Flew opposes Rawls' theory because he sees it as committed to these theses. Once again, however, Flew is wrong. He has misunderstood the kind of enterprise Rawls' theory is. Rawls does not believe that we are all equal in our deserts and entitlements, but he does believe that they are to be assumed to be equal until it can be demonstrated otherwise. He does not deny that such demonstrations can be made. In fact, the whole point of his theory is to set up the criteria by which the success of any such attempt is to be judged. Rawls, then, is not dismissing all our individuating and differentiating characteristics as morally irrelevant. He is merely assuming that a claim to any special merit because of some individuating characteristic has to be substantiated.

Nor does Rawls think that the grounds of all just entitlement and desert are common to all, so that these grounds, presumably have to be found in the universal and essential nature of humanity, never in any individual particularities. Flew here is confusing what decides whether this particular individual deserves something or not, with the quite different question of whether the kind of consideration which is being used to decide his desert is a relevant consideration for desert at all. For example, is it grounds for deserving better than equal treatment than others that you studied when you could have been earning, or that your responsibilities are greater than those of your employees, or that your work is more intellectually or emotionally demanding or that it involves a high level of skill, or that you work longer hours or your work is more dangerous, etc. These questions, Rawls says, are decided by the contractors behind the veil of ignorance. These contractors certainly possess no individuating and differentiating characteristics. This

shouldn't worry Flew, however, because they are not deciding the individual deserts or dues of person A or B.

Nonetheless, they would have to decide whether these were grounds for merit on some criteria, which, Rawls argues, justice requires be that everyone benefits by inequalities being allowed on these grounds. Suppose the contractors accepted all the above proposals as grounds for desert, then, to settle whether or not A deserves better than equal treatment, we would have to look at the particular facts about A, namely, does he work longer hours, did he forego a wage while studying, and so forth.

Don't we normally distinguish these two questions in our debates about merit? To decide whether A deserves extra money for working longer hours we have first to decide whether working longer hours is meritorious. That is, whether contractors behind the veil would agree that those who work longer hours should get a greater share of either liberty or goods other than liberty. As I argued earlier, in considering unearned or undeserved distributions, each contractor would insist on an equal distribution. Most benefits, however, are not unearned or undeserved. One's bodily parts and natural talents are one of the few examples. Others, perhaps, are land and natural resources such as water, or oil. It will be useful in understanding how we should deal with the unearned or undeserved cases if we look at cases of earned or deserved distribution. Consider the situation described earlier where everyone either receives \$200 or, if some receive \$500, the rest receive \$300. The Difference Principle requires non-envious rational people to opt for the \$500-\$300 option. This, however, is a \$500-\$300 option for those who do the work that produces this amount of money, just as the \$200 was for those who did the work. Behind the veil the contractors know that they could be A who is a worker or B who is a malingeringer. If they are A they will want to be able to get either the

\$500 or \$300, depending on their talents and efforts. They will not want their income reduced by \$50 to support a malingerer whose lack of effort has caused a drop in the aggregate income of the society. They will therefore want penalties attached to malingering severe enough to prevent malingering without generating costs of their own equal to or greater than the costs of malingering. If the contractors turn out to be malingerers, they will want to benefit from the increased wealth of the community while avoiding the work. They know, however, that if some people are seen to be parasites on the others, this will reduce incentive and reduce the aggregate wealth. They are faced, then, with the choice of malingering without penalty when this causes a drop in income or malingering with a penalty if they are caught or not malingering. How this would be resolved would depend on the particular weights they placed on monetary rewards versus not working too hard. In Rawls' theory, however, the principles adopted by the parties behind the veil must be general, universal and public. They must be general in that they can be formulated 'without the use of descriptions'. They must be universal in two senses. First, 'they must hold for everyone in virtue of their being moral persons', second, they must be chosen 'in view of the consequences of everyone's complying with them'. They must be public in that they must be openly acknowledged as one of the fundamental rules of the society. (1971, 131-132)

The contractors recognise that, even if they turn out to be malingerers in real-life, it would not be in their interests to have malingering built into the principles governing the society. The effect of this would be to make all those who could earn \$500 or \$300 and would do so if there was no malingering refuse to do so unless there were penalties to discourage malingering. Their responses would be that, if a malingerer can earn the same as me without penalty, they why should I not become a malingerer, causing the income of cash to drop either to zero or some figure well below

\$500 or \$300. Their thinking here, is not envious. They are not envious of the malingerer. They are merely concluding that, if you get \$300 whether or not you put effort into it, then it is more rational not to put effort into it and to put your effort in other things that you want to do. The contractors, therefore, will conclude that it be a principle of the society to reward effort and to penalise lack of effort with respect to the earning of the \$300. In addition, they will have agreed to reward talents, if it is combined with effort, by giving as much extra to those with talent as is necessary to maximise the income of those without these talents. Nonetheless, as differential rewards lead to differential power, which leads to differential freedoms, they will only allow differential monetary rewards to the extent that this is compatible with equal freedom or with unequal freedom if this maximises the freedoms of the least free. This could lead to rather low ceilings being placed on differences in income. There could be strong disincentives to initiative where initiative leads to differences in wealth that entrenched differences in power into the system. However this turns out, the important things to note with respect to Flew's case against Rawls, is that the personal characteristics of individuals are relevant to their entitlement, in that the malingerer will be entitled to less than the worker, and will receive less if caught. Admittedly, in some sense, a person may not be able to help being a malingerer in that that is the kind of person he is and he lacks the capacity to change. On certain analyses of free-will, this means that he doesn't deserve to be penalised for his actions. Unless this is pathological, however, it will not be an excuse in the Rawlsian model. Even if it is pathological, it will only exempt the person from the penalty rather than still entitle him to an ordinary wage. Even so the contractors realise any of them could become pathological cases, therefore there will be provisions to care for these cases. Malignerers, as a consequence of the adoption of a policy of penalising malingering, may become the worst-off

members of society. The contractors would, however, not operate maximin to choose that outcome with the least bad worst-off position if that position is one brought about by the person's own fault, however this is determined behind the veil. They will, however, use maximin to ensure that the families of malingerers are, as far as possible, not also penalized for the action of the malingerers themselves.

There seems nothing, then, in Rawls' theory that commits him to the view that in real life we are all equal in our deserts and entitlements or that we are all equal in all morally relevant respects regardless of what we have done or are capable of doing. His theory does require, however, that we are all considered equal in that, prior to determining which of our personal characteristics or kinds of effort earn merit, no particular characteristic is favoured merely because it is ours.

Rawls has claimed, correctly as I hope I have shown, that the mere possession of particular talents or temperaments is not desert-making.

As he says,

'It seems to be one of the fixed points of our considered judgements that no-one deserves his place in the distribution of native endowments, anymore than one deserves one's initial starting place in society.' (1980, 139)

Flew accuses Rawls of setting up a straw man here. He says,

'This constitutes the erection and demolition of a straw man ... For those of us who still employed this challenged concept of desert refer, as bases of good and ill desert, not to people's native talents and temperaments, but rather to what they themselves have actually done or abstained from doing.' (1980, 139)

If Rawls is constructing straw men here, this presumably means that Flew believes that no one denies Rawls' claim. Flew therefore seems to be admitting that it is quite possible for people not to deserve their native talents and temperaments, yet still to gain desert from how they exercise these talents and temperaments. I have tried to show that this is precisely the consequence of Rawls' theory.

Flew summarises his discussion of desert and entitlement in two consecutive claims. The first is:

'The conceptual truth seems to be ... that desert presupposes entitlement - entitlement, that is, to whatever of themselves people may exercise or fail to exercise in the acquisition of good or ill desert.'

Rawls would agree with this claim. This entitlement, however, derives from the contract. No one has the right to deprive us of our unshareable bodily parts or our natural talents and temperament because the contractors behind the veil of ignorance would unanimously agree that it was in everyone's impartial interests that this not be allowed. As we have seen, even shareable bodily parts are sacrosanct except under very special conditions. However, this right is merely the right to continued possession of these things, not the right to possess the benefits gained from their use unless we have earned these benefits, and it is in the interests of the least well-off that we be allowed to earn such benefits.

Flew's second claim is that

'Contract, too, (and similarly,) presupposes rights which were not themselves contractually gained. For the contracting parties have likewise to be entitled to whatever it is they propose to trade.' (1980, 142)

It is, however, a non-sequitur to derive the first sentence from the second. It is quite true that contracting parties in the real world have to be entitled to whatever it is they propose to trade, but that entitlement is conferred upon them by the decisions of the parties behind the veil of ignorance.

Prior to the contract, there will be natural talents, bodily parts, natural resources, land, etcetera, to which no one has any entitlement, including those in actual possession. It is not to be construed as if those in possession of these things come to the contracting situation behind the veil of ignorance proposing to swap a slightly arthritic leg for a short-sighted eye. How can they? Justice requires that they don't know in the real world whether they have any legs at all, let alone arthritic ones. Even if the contractors behind the veil of ignorance know their own bodily parts and natural talents behind the veil, there is no guarantee they will have the same endowments once the veil is lifted. There is no point then in contractor X agreeing to swap a leg for an eye with contractor Y behind the veil of ignorance to gain some benefit once the veil is lifted because he has no reason to believe he will keep the benefit. In this case, then, the contractors do not have to be entitled to what they propose to trade because they are not going to trade anything. Instead, they are establishing the principles which entitle people to the things they propose to trade, as well as the principles by which a trade will be a fair or just one.

8. Impartialists and Unearned Entitlements

The impartialist utilitarians J and K will reason much as Rawls' contractors about the characteristics that are meritorious about the relationship between talents and effort, about the treatment of malingers, and about people not deserving their natural allocation of talents and bodily parts.

They will, however, reason differently about what rights people have to their bodily parts. This is because J and K are not constrained by the Difference Principle. Instead they are to apply the impartialist axioms to the conflict of desires where A wants some of B's bodily parts and B doesn't want A to have them. This means that A will have to coerce B to get those parts even though B did not acquire these bodily parts from A by coercion. Consequently, as J and K are noncoercionists they will agree that A is not to coerce B in order to obtain A's bodily parts for himself. The same will apply to natural talents, to the extent that they can be transferred from one person to another by force. The velorians, then, would agree with Flew that we have a right to our bodily parts, whether shareable or unshareable.

Chapter 14

'Oughts', 'Wants' and Egoism1. Introduction

So far I have cast conflicts of desires in terms of the wants of A and B then used the velorians J and K to arrive at the principles which tell us what A and B ought to do. In this chapter I wish to describe what I believe to be the logic of the word 'ought'. I have used it as a 'hypothetical' imperative, that is, the 'ought' conclusions the velorians reach are what the conflictants 'ought' to do from the viewpoint of impartialism. I will argue that all uses of 'ought' are hypothetical. Supposed categorical 'oughts' are really disguised want claims. I will show that even egoists can use universalizable 'oughts' without contradiction in spite of arguments from various philosophers to the contrary. It is my contention that egoism is an unacceptable moral theory not because it is irrational but because it would not receive unforced agreement. I will take up why impartialism is to be preferred to egoism in Chapter 16.

2. The 'Ambitious Egoist'

Egoists have been charged with irrationality because it has been claimed the logic of the word 'ought' is such that an egoist who says that he ought to act egoistically is guilty of a contradiction.

One way in which this charge has been sustained is by the claim that categorical 'ought' statements either uttered by, or endorsed by, the egoist are not universalizable. My strategy will be to see whether these supposed logical difficulties disappear if egoism is recast in terms of hypothetical 'oughts'. To begin with I'd like to look at an argument used by both Frankena and Kalin against a certain kind of egoist. For

reasons that will become apparent later I will call this kind of egoist an 'ambitious' egoist.

The ambitious egoist, according to Frankena (1973, 17-20) believes -

- (a) If A is judging about himself, then A is to use this criterion:
A ought to do Y if and only if Y is in A's overall self-interest;
- (b) If A is a spectator judging about anyone else, B, then A is to use this criterion: B ought to do Y if and only if Y is in A's overall self-interest.

Kalin (1968, 68) sets out to show that anyone who believes (a) and (b) believes that he ought both do and not do a particular action, therefore the principles fail to be a guide to action. His demonstration of this is as follows -

- (i) Suppose A is the evaluator, then

What ought A to do? A ought to do what's in A's interests (by (a))

What ought B to do? B ought to do what's in A's interests (by (b))

What ought C to do? C ought to do what's in A's interests (by (b))

And so on.

Therefore, everyone ought to do what's in A's interests (by (a) and (b)).

- (ii) Suppose B is the evaluator, then

What ought A to do? A ought to do what's in B's interests (by (b))

What ought B to do? B ought to do what's in B's interests (by (a))

What ought C to do? C ought to do what's in B's interests (by (b))

and so on.

Therefore, everyone ought to do what's in B's interests (by (a) and (b)).

Conclusion: Everyone ought to do what's in A's interests, and everyone ought to do what's in B's interest, ... and so on.

(iii) A's interest is at least sometimes incompatible with B's. When this occurs, it will be the case that A ought to do what is both in A's interest and against it (that is what is in B's interest at A's expense). Under these circumstances, ethical egoism gives self-defeating advice. It therefore fails as a moral theory.

This argument clearly depends on the use of categorical 'oughts'. What I wish to do now is to investigate how the Kalin-Frankena argument against the ambitious egoist works if we turn the categoricals into hypotheticals. In other words, can an egoist consistently refuse to countenance the existence of categorical imperatives? If he can, does he thereby rescue the rationality of his egoism?

3. Deducing Action

The egoist could maintain that he will interpret all categorical 'oughts' as the conclusions of arguments in which the categorical 'ought' is the consequence of a hypothetical imperative whose antecedent is also the minor premiss. For example,

Argument 1

If A's self-interest is to be maximised, then A ought to do what he believes is in his self interest.

A's self-interest is to be maximised,

therefore

A ought to do what he believes is in his self-interest.

This has the standard form of modus ponens, namely,

If p then q

p

therefore

q

When 'p' and 'q' are propositions, the fact that 'p' appears as the minor premiss indicates that its truth is no longer to be considered as hypothetical but as actual. Argument 1 could be given the following 'propositional' interpretation.

Argument 2

If 'A's self-interest is to be maximised' is true, then

'A ought to do what he believes is in his self-interest' is true.

'A's self-interest is to be maximised' is true,

therefore

'A ought to do what he believes is in his self-interest' is true.

Here 'p' appears to be a prediction that it will be the case that A's self-interest is maximised and 'q' is a statement to the effect that A is likely to have done what he believed was in his self-interest.

Both 'p' and 'q', however, can be interpreted as having a prescriptive component. When this occurs, I will use upper case letters such as 'P' and 'Q' to represent sentences with a prescriptive component

It is, however, logically improper to substitute 'is prescribed' for 'is true' in Argument 2. This would give us Argument 3, the major premiss of which will in many cases be false, thereby making the argument unsound.

Argument 3

If 'A's self-interest is to be maximised' is prescribed, then
 'A ought to do what A believes is in his self-interest' is
 prescribed.

'A's self-interest is to be maximised' is prescribed,
 therefore

'A ought to do what he believes is in his self-interest' is
 prescribed.

For it to be the case that 'A's self-interest is to be maximised' is prescribed, there must be someone who prescribes it, that is, who wants A's self-interest maximised. Just because someone wants A's self-interest maximised, it doesn't follow that he wants A to do what A believes is in A's self-interest. Consequently, the major premiss may well be false, thus making Argument 3 unsound. There is a parallel here with propositional arguments. Consider Argument 4, assuming it to be true that all haemophiliacs are male.

Argument 4

If 'A has haemophilia' is true, then 'A is a male' is true.

'A has haemophilia' is true,

therefore

'A is a male' is true.

If we replace each 'is true' in Argument 4 with 'is believed', then we have the same kind of problem I discussed with 'is prescribed' in Argument 3, that is, some one has to do the believing and just because he believes that A is a haemophiliac it doesn't follow that he believes that A is a male. So, while the major premiss of Argument 4 is true when each proposition in it is treated as 'true', it may no longer be true when each proposition is treated as 'believed'. See Argument 5.

Argument 5

If 'A has haemophilia' is believed, then 'A is a male' is believed.

'A has haemophilia' is believed,

therefore

'A is a male' is believed.

Nonetheless, in both the propositional case and the imperatival case, the premiss of Arguments 1 and 4 'entail' their conclusions, in Hare's sense of 'entail'. Hare defines this sense of 'entailment' as follows

'A sentence P entails a sentence Q if and only if the fact that a person assents to P but dissents from Q is sufficient reason for saying he has misunderstood one or the other of the sentences.' (1952, 25)

In Argument 1, the two premisses can be treated as a single sentence consisting of two clauses. Together with the minor premiss they entail

the conclusion in the sense of 'entail' defined above. The relationship between the premisses and the conclusion is such that a person's refusal to assent to the conclusion after assenting to the premisses is normally sufficient for us to say that he has misunderstood one or the others of the sentences. The only exception here is if there is some state of affairs more strongly desired than, and incompatible with, that indicated in the antecedent of the first premiss. This exception can be dealt with by treating the minor premiss as infelicitous when a state of affairs is more strongly desired, and incompatible with, that indicated by the minor premiss. When this occurs, a person will be regarded as not having really assented to the minor premiss, which means that he has not assented to the conjunction of the major and minor premisses as a whole. He does not really want the state of affairs indicated by the minor premiss because he prefers some other state of affairs when the two conflict. With this modification, anyone who really assents to the major and minor premisses yet refused to assent to the conclusion, behaves in a way that is incomprehensible except on the hypothesis that he doesn't understand one or the other of these sentences. As Hare says,

'Speakers may on different occasions use words with different meanings, and this means that what is entailed by what they say will also differ. We elicit their meaning by asking them what they regard their remarks as entailing.' (1952, 25)

Using this test, if people don't see the premisses of Argument 1 as entailing its conclusion, then they and I have different understandings of the terms in these sentences, just as we would do if they assented to the proposition 'If A has haemophilia then A is a male and A has haemophilia' yet denied that A was a male. The propositional and imperatival cases seem quite parallel here. For the purposes of this paper, I will treat them as parallel, regarding counter-examples to the imperatival case as unsuccessful if they also apply to the propositional case. If the

relationship between imperatival premisses and conclusion is no tighter than that between propositional premisses and conclusion then it is tight enough.

5. Defending the 'Ambitious' Egoist

The ambitious egoist, it will be remembered, believes:

- (a) If A is judging about himself, then A is to use this criterion:
A ought to do Y if and only if Y is in A's overall self-interest.
- (b) If A is a spectator judging about anyone else, B, then A is to use this criterion:
B ought to do Y if and only if Y is in A's overall self-interest.

It is, however, odd to operate on a criterion which requires a person to do what is, in fact, in A's overall self-interest. The best we mere mortals can manage is for A to do what he believes is in A's self-interest, and I will modify all future cases to take this into account. I will take A as believing that B ought to do Y if and only if A believes Y is in A's overall self-interest because this is the formulation that would strike any actual egoist as most likely to be true. Admittedly it could be formulated as, 'B ought to do Y if, and only if B believes that Y is in A's overall self-interest' but A would be rather worried that B's concept of A's self-interest could be a long way away from A's own.

The ambitious egoist, as we have seen, believes the following:

'If A's self-interest is to be maximised, then everyone ought to do what A believes is in A's self-interest and if B's self-interest is to be maximised, then everyone ought to do what B believes is in B's self-interest ... AND if n's self-interest is to maximise then everyone ought to do what n believes is in n's self-interest.'

This, however, is merely a technical belief, that is, a belief about the means by which different ends will be achieved. As a technical belief it is probably true. A's self-interest is much more likely to be achieved if everyone does what A believes is in A's self-interest than if only A does.

Nonetheless, it is logically possible that A's self-interest is better achieved by other means, as we mentioned earlier.

The ambitious egoist, then, is someone who has the technical belief, which he thinks is as true of others as of himself, that each person's self-interest will be maximised if everyone pursued it. In addition, any particular ambitious egoist, A, wants his own self-interest maximised, being either indifferent or hostile to the self-interest of other people. Given our previous analysis, this means that he is not entitled to, but required to, acknowledge that everyone ought to do what they believe is in his self-interest, or else he is irrational. See Argument 6.

Argument 6

If A's self-interest is to be maximised, then everyone ought to do what A believes is in A's self-interest.

A's self-interest is to be maximised,

therefore

Everyone ought to do what A believes is in A's self-interest.

A, however, does not want B's self-interest to be maximised, therefore he is not committed by the hypothetical to the conclusion that everyone ought to do what is in B's (...n's) self-interest. Counter to Kalin and Frankena, then, he is not committed to self-defeating courses of action by his egoistic principles.

6. Vicious Universalisation

The anti-egoist will probably accept that, if A's self-interest is to be maximised, then everyone ought to do what A believes will maximise A's self-interest, but he will also want to know why A's self-interest ought to be maximised. In other words, where a categorical imperative is detached from a hypothetical imperative by means of prescribing the antecedent of the hypothetical, the anti-egoist is always entitled to ask why the action prescribed in the antecedent 'ought' to be done, hoping to produce a situation where the 'ought' used in the answer can't be universalised but produces instead the kind of self-contradiction Kalin and Frankena thought they had found. I will call this 'vicious' universalisation,⁽¹⁾ in that the anti-egoist pursues a regress of universalisations until he gets a 'vicious' consequence. Let us see if it works on the ambitious egoist A. 'Why', says the anti-egoist, 'ought A's self-interest be maximised?' Our ambitious egoist only believes in hypothetical imperatives, so he sees this question as a request for a goal, end or purpose for which the maximisation of his self-interest is a means. He might produce the hypothetical, 'If A is to be happy, then A's self-interest ought to be maximised.' This is a contingent causal claim. Perhaps A will be happier if his self-interest isn't maximised. Perhaps it's part of human nature only to be happy in self-sacrifice.

(1) I use Kalin's definition of 'universalisation' here, namely, 'If it is reasonable for A to do S in C, it is also reasonable for any similar person to do similar things in similar circumstances.' (1970, 66) Hudson (1970, 184) interprets Hare's notion of universalisability in much the same way. He gives an example of how 'ought' judgements must be universalisable. He says 'To say, for example, that we ought to encourage immigrants to return to their countries of origin is to say that no-one, placed as we are, ought to fail to encourage them to do so.' The viciousness of the universalisation only takes effect if the "ought" is used categorically.

Nonetheless, A happens to believe he will be happy if his self-interest is maximised and his answer is at least logically respectable. He would arrive at it via Argument 7.

If A is to be happy, then A's self-interest ought to be maximised.

A is to be happy,

therefore

A's self-interest ought to be maximised.

A is not committed to wanting other people to be happy, especially where this prevents his happiness, nor is he committed to wanting other people's self-interest maximised, especially where this prevents the maximisation of his own self-interest.

The anti-egoist could still ask, 'Why ought A to be happy?' At this point A will probably not want to claim that he ought to be happy. After all, he believes that 'ought' statements properly derive from hypotheticals, which means he would need to find some goal or end which his being happy serves. This he will be reluctant to do because he is likely to regard his own happiness as the most basic goal, the final justification for behaviour. If the anti-egoist can commit A to believing that A 'ought' to be happy, then he may be able to show that this universalises to 'Everyone ought to be happy'. A does not believe that everyone ought to be happy, nor does he want everyone to be happy, so he will not want to take any step that commits him in this way. A, therefore, will refuse to say that he ought to be happy. He doesn't believe he ought to be or, for that matter, that anyone else ought to be either. He just wants to be. Logically, though, this puts him on a par with the anti-egoist. Suppose the egoist goes on the attack, asking the anti-egoist why he believes something ought to be the case, e.g., that people ought to keep their promises. The anti-egoist will have to show either that people keeping

their promises is a means to some end, for which further justification could be asked, or else he has to say that there is no such further end but that keeping promises is an end in itself. It ought to be done just because it ought to be. As an argument to the unconvinced this is singularly unpersuasive. What the anti-egoist has done is use the 'ought' claim as a substitute for a want claim. All he is entitled to do is assert that he wants everyone to be happy but he substitute a categorical 'ought' statement instead to give the appearance that he is appealing to some objective realm of moral facts. Unlike Zimmerman (1969), I am not proposing that we can reject "ought" statements and replace them with "want" statements. Instead my position is the same as that expressed by MacIntyre (1981, 11-22). MacIntyre claims that emotivism should be understood as a theory of the use of moral language rather than a theory of its meaning. MacIntyre argues with the critics of emotivism that judgements 'such as "this is right" or "this is good" ' do NOT 'mean the same as "I approve of this; do so as well" or "Hurrah for this!" or any of the other attempts at equivalence suggested by emotive theorists; but even if the meaning of such sentences were quite other than emotive theorists supposed, it might plausibly be claimed, if the evidence were adequate, that in using such sentences to say whatever they mean, the agent was in fact doing nothing other than expressing his feelings or attitudes and attempting to influence the feelings and attitudes of others ... We could NOT safely infer what someone who uttered a moral judgement was doing merely by listening to what he said. Moreover the agent himself might well be among those for whom use was concealed by meaning. He might well, precisely because he was self-conscious about the meaning of the words that he used, be assured that he was appealing to independent impersonal criteria, when all he was in fact doing was expressing his feelings to others in a manipulative way.' (1981, 13)

The egoist might well regard the anti-egoist as doing precisely this when the anti-egoist says that the reason why you ought to keep your promises is just that you ought. If MacIntyre is right, then the egoist is no worse off than the anti-egoist. The egoist has traced through a chain of "oughts" until he has arrived at a position where he no longer says that something ought to be the case. He merely says that he wants it to be. The anti-egoist ends his chain with the claim that something ought to be done because it ought. The anti-egoist will be hard-pressed to explain what the difference is.

This does not mean that there are no reasons for rejecting egoism as a moral theory but merely that these reasons are not to be found in either the self-defeating nature of egoistic principles or in the egoist's supposed inability to describe his theory without contradiction. Egoism is not to be rejected because it is irrational or incoherent but because it would not be accepted by unforced agreement. The egoist, of course, is free to ask why ought we reject what is NOT accepted by unforced agreement. If this is answered by reference to some other goal, the same question can be asked again, and so on. I will consider how we might stop this infinite regress in Chapter 16.

7. The Humble Egoist

Although Kalin believes, wrongly as we have seen, that the ambitious egoist holds a self-contradictory principle, he does believe that there is a defensible version of egoism, which I will call 'humble egoism'.

According to Kalin, the humble egoist believes the following two principles, the first (i.e. (a)) being identical with the first principle of ambitious egoism.

- (a) If A is judging about himself, then A is to use this criterion:
A ought to do Y if and only if Y is in A's overall self-interest.
- (c) If A is a spectator judging about someone else, B, then A is to use this criterion:
B ought to do Y if and only if Y is in B's overall self-interest.

In future statements of these principles I will modify them, as indicated earlier, by talking of someone doing Y if and only if he believes that Y is in his overall self-interest.

Medlin (1957) claims that even humble egoism is self-contradictory. His argument, briefly, is as follows. If A believes that he ought to do Y if and only if he believes that Y is in his overall self-interest, and his belief that he ought to do so means that he approves of his doing so, or wants himself to do so, then the principles (a) and (c) will lead to contradictory advice because his believing that B ought to do what is in B's interest also commits him to wanting B to do what is in B's interests even when this conflicts with his (A's) interests.

Kalin replies to this criticism by denying that A's belief that B ought to do Y implies that A wants B to do Y.

Kalin argues for this claim by showing that there are perfectly respectable precedents for this kind of position. He uses the example of games. A chess-player can believe that his opponent ought to move his bishop yet not want him to move the bishop.

Kalin's argument has been criticised by Carlson (1973).

'There is a similar confusion in Kalin's attempt to show that I can consistently both believe that my opponent in a chess game ought to check my king, and yet not want him to. Note that he passes imperceptibly from asserting (1) "This is how he ought to move," to (2) "But believing that he ought to move his bishop and check my king does not commit me to wanting him to do that ..." without realizing that I can consistently assert (1) without asserting (2), insofar as the most plausible interpretation of (1) is as a hypothetical (the antecedent being suppressed), which does not entail the categorical at (2). I say this, because in Kalin's example what I really believe is not that this is how my chess opponent ought to move, period. (i.e., for the sake of simplicity, in such a way as to put my king in inextricable check), but that this is how he ought to move if the game is to be won by him (according to the rules); that he ought therefore to try to move that way. Since however, (ex hypothesi) I do not want him to win the game, it is plausible to assert that I do not really believe that he ought to succeed in moving in the prescribed way, and despite the fact that (with Kalin) I can meaningfully assert (1) (so long as its suppressed hypothetical force is understood). I conclude that this particular putative counter-example does not in fact support Kalin's key (more general) claim, namely, that one's belief in what anyone else ought to do or ought to make the case does not commit one to wanting any one else to acting so. (1)

Carlson, I believe, is correct. The chess-player holds the causal hypothetical, 'If my opponent is to win, then he ought to move his bishop'. If the chess-player proceeds to the categorical 'My opponent ought to move his bishop' he can only do so via the minor premiss, 'My opponent is to win', which, as we have seen, is not to be interpreted as a prediction but as an imperative indicating the speaker's desire to have that state of affairs realised.

(1) Kalin (1975) , in his reply to Carlson tries to avoid this criticism by distinguishing a teleological and a deontological sense of "ought". If my arguments are correct this is unnecessary. The word "ought" can be treated as unambiguous in meaning, although the different kinds of hypotheticals that can occur it may make it appear as if there is a different sense of "ought" at work in each case.

Even though Carlson's point is correct, it doesn't follow that Medlin's criticism of the humble egoist works. The humble egoist can be regarded as believing the following:

If A's self-interest is to be maximised, then A ought to do what he believes is in his self-interest AND If B's self-interest is to be maximised then B ought to do what he believes is in his self-interest ... AND If n's self-interest is to be maximised then n ought to do what n believes is in n's self-interest. As A prescribes the imperative 'A's self-interest is to be maximised' but does not prescribe that B's ... n's self-interest is to be maximised, he is only committed to the conclusion, 'A ought to do what he believes is in his self-interest'.

He accepts all the other 'ought' statements hypothetically. The anti-egoist could try to perform the same vicious universalisation on him as on the ambitious egoist, with the same effect, namely, that ultimately the humble egoist would assert that, if he is to be happy, then his self-interest ought to be maximised, but denying that he ought to be happy. Perhaps he, and even the ambitious egoist, might agree that everyone ought to pursue their own happiness, but only because he regards this as definitive of 'rationality'. Here the imperative isn't causal but, in Harsanyi's sense (Harsanyi, 1976) 'formal'. Again, the egoist need NOT believe that everyone ought to be rational.

The humble egoist, therefore, is not vulnerable to Medlin's criticism. Just because A believes that he ought to do Y if and only if he believes Y is in his self-interest, and this means that he wants himself to do what is in his own interests, he is not committed to wanting other people doing what they believe to be in their interests, especially when this frustrates his own self-interest. He is merely considering hypothetically

how others ought to behave if they prescribe to themselves the same kinds of imperatives about their interests that he prescribes to himself about his own.

The humble egoist, then, only appears to differ from the ambitious egoist in his opinion about the best means to his own self-interest, and neither is guilty of any logical fallacy. The ambitious egoist, ambitious because he entertains the hope that everyone wants to do what's best for him, believes that their doing so would maximise his self-interest. The humble egoist, humble because he doesn't really entertain the hope that everyone wants to do what's best for him, believes that his self-interest is most likely to be maximised if at least he pursues it. It is odd, therefore, that Kalin ever thought that one could be logically defensible when the other was not, as logically they are counterparts of each other. In fact, the ambitious egoist is probably closer to the truth than the humble egoist, at least as concerns what would be the desirable situation for an egoist to find himself in. He would, however, have to have the power of a Caligula, so the humble egoist is probably more realistic.

8. Motivating by Practical Arguments

In conclusion I'd like to clear up some points raised in the discussion of 'vicious universalisation'. When the anti-egoist asks either kind of egoist "Why ought your self-interest be maximised?" it is undoubtedly the case that the anti-egoist will not regard as satisfactory the answer which the egoist gives him, namely, 'Because it maximises my happiness'.

On the analysis given so far, the anti-egoist's dissatisfaction should not be surprising. The anti-egoist is looking for an answer that will enable him, too, to prescribe the categorical imperative. What he needs is both

a hypothetical major premiss and a prescriptive minor premiss from which he can deduce the conclusion "A's (the egoist's) self-interest ought to be maximised". The trouble is that the prescription that A uses as a minor premiss, that is, "A's happiness is to be maximised" is not one that the anti-egoist, unlike A himself, has any obvious reason to prescribe.

Nor, of course, does it provide other egoists with a reason why they should want A's happiness to be maximised. This draws our attention to an important feature of the way in which I have handled sentences containing 'ought'. I will now spend some time elucidating this feature and its implications.

Consider the belief of the ambitious egoist A to the effect that everyone (B, C,n) ought to do what is in his (A's) interest.

A believes the following hypothetical.

Argument 8

If 'A's self-interest is to be maximised, then B ought to do what A believes is in A's self-interest'.

'A's self-interest is to be maximised',

therefore

'B ought to do what A believes is in A's self-interest'.

B will not find this argument convincing unless he prescribes the minor premiss, that is, unless he wants A's self-interest to be maximised. Even though he may well accept the truth of the major premiss and the validity of the argument form, it is unlikely that he will accept the conclusion. This fact, however, does not render A's argument invalid. It merely renders it unsound in B's eyes because B regards the minor premiss as

false. Were B to become a slave to A's charisma, he might well find the argument perfectly sound, prescribing both the minor premiss and the conclusion. As slave the only reason he might not regard the argument as sound would be if he doubted the truth of the major premiss, thinking perhaps that A's self-interest would be better served if B did what B thought was in A's self-interest. Be this as it may, the point to be noted is that the soundness of the argument depends on the psychological state of the person responding to the minor premiss. The truth of the minor premiss is always person-relative, making the argument sound if the person who ought to do the action indicated in the consequent also prescribes the end indicated in the antecedent, provided that the means proposed truly is the best one for attaining that goal.

As we have seen, the person who ought to do the action indicated in the consequent need not be the person whose self-interest will thereby be maximised. A person may, for example, wish to maximise the interests of some sectional group. He might argue as follows,

Argument 9

If Group G's interest is to be maximised, then everyone (or some particular person) ought to do X.

G's interests are to be maximised,

therefore

Everyone (or some particular person) ought to do X.

We will only agree that his argument is sound, however, if we also prescribe the maximisation of G's interests, that is, if we share his patriotism, or religious or political convictions or whatever. We can meaningfully argue with him about the truth of the major premiss and if we can produce factual

evidence to show that G's interests are more likely to be maximised if everyone does Y rather than X, he will change his argument. He is, therefore, rational in the appropriate sense.

Rather than maximise the interests of some group, however, he may wish to maximise the achievement of some ideal. A suitable substitution of 'maximising the achievement of ideal I' for 'maximising the interests of Group G' could be made in Argument 9 to produce Argument 10, with similar comments as made on Argument 9.

We can even extend the model to cover what are claimed to be 'moral' arguments, for example,

Argument 11

If people are to act as would be determined by rational contractors behind a veil of ignorance (or a benevolent ideal observer) then everyone (or some particular person) ought to do X

People are to act as would be determined, etc.

therefore

Everyone (or some particular person) ought to do X.

If utilitarianism and rational contractor theories are to be counted as typical moral systems, then even morality has no categorical imperatives. As Mrs. Foot (1972) maintains, if morality is a system of imperatives, it must be a system of hypothetical imperatives. Nonetheless, the analysis I have given of the use of 'ought' explains why philosophers have thought of morality as a system of categorical imperatives. Moral systems like utilitarianism or Rawls' contractarianism are systems that provide their believers with ideals. In utilitarianism the ideal is the state of

affairs in which happiness has been maximised. This provides a standard against which to measure any act a person performs. There is the act he actually did and the act he ought to have done if the state of affairs in which happiness had been maximised was to obtain. Likewise, with Rawls' model, the acts we actually do can be compared with the acts we would have done if we had acted in accord with the contractors' principle.

Once one is within such a system, there is no need to preface every 'ought' statement with the antecedent of the conditional. It can just be assumed. As a result, 'ought' statements look categorical even though they are really hypothetical. It is only when party A from one system argues with party B from another that it is essential that both parties realise the hypothetical nature of their 'ought' statements. Within any moral system, once its believers have adopted categorical 'oughts', all the moral judgements within the system take on the appearance of descriptive judgements. There may, for example, be certain categories of killing that transgress the ideal. These may be similar to that category of killings we call 'murder'. They all fall under the concept of 'murder', however, because of their relationship to the ideal, not because of any common empirical features they happen to possess. It may be that they all transgress the ideal because they are similar in their empirical properties but it is not this that makes them 'murder'. Anyone who tried to find what made an act a 'murder' by looking for empirical properties common to all murders would not have grasped the concept of murder.

It may be, for example, that someone thinks up a new kind of killing which transgresses the ideal in the way the others did. It still falls under the concept 'murder' even though its empirical properties differ from those of other murders. Anyone who refused to call it a murder because of its lack of the usual empirical properties associated with

murder would have mistaken what Kovesi (1967, 4) calls the material element of the concept of murder for its formal element. Within the one moral system, then, there would develop an inter-related set of moral notions that appear to describe the world. When making moral judgements, we do not appear to ourselves to be applying an ideal. We appear to be more concerned with whether an act falls under a notion such as 'murder' or 'manslaughter'. These notions, however, derive from some standard that we think actions ought to meet. As a consequence, we cannot both classify an act as a case of 'murder' or whatever and remain free to value it or detest it as we wish. Given our ideal, we must detest it if it is properly classified as 'murder'. When we say that murder is 'wrong' or 'bad' the words 'wrong' and 'bad' are used as reminders, to use Kovesi's term, (1967, 26) rather than as providers of additional information, or as a discriminator (1967, 109). Someone who doesn't share our ideal can legitimately wonder whether to value or detest an act we classify as 'murder' but he cannot use the word 'murder' in the same sense as we use it without also committing himself to detesting acts of 'murder'. He will have to find a new vocabulary to express his attitudes or else give our vocabulary a new meaning. Foot (1978, 137), for example, claims that ordinary English words like 'father' or 'daughter' determine criteria of goodness in the sense that, if a foreign language said of a person that she was a good X when she denounced her parents to the police, then X could not be translated by our word 'daughter'. Montefiore, (1961) in his reply to Foot suggests that her examples indicate a need for the prescriptivist to invent a new language that does NOT have these value-judgements built into it.

It is likely, then, that if we all talk to each other from within a particular moral system that we will treat its hypothetical imperatives as categorical imperatives and regard its moral notions as describing the

world. The problem, however, is to identify what the ideal is that directs the development of these apparently descriptive notions. Kovesi calls it 'the moral point of view'. He distinguishes it from other points of view in that the notions we form from this point of view 'are not only formed by ourselves but they are about ourselves' and they 'are about ourselves in so far as we are rule-following rational beings'. As Kovesi notes, however, 'there are also other notions that are about ourselves in so far as we are rule-following rational beings, like 'clever', 'consistent', 'learned', etc., which are not necessarily moral notions'. (1967,-147)

Theories like utilitarianism and Rawls' rational contractor theory can be seen as attempts to try and specify what it is that makes moral notions different from these other notions. Utilitarianism postulates that moral notions derive from our concern with maximising happiness. Rawls postulates that they derive from our concern with our status as free and equal beings. The moral notions we would develop to describe the world from a utilitarian point of view would not necessarily be the same as those we would develop from a contractarian point of view, yet they will share what Kovesi believes to be a crucial requirement of any moral notion, namely, 'they not only have to be formed from the point of view of anyone, but they also have to be about those features of our lives that can be the features of anyone's life'. (1967, 148)

Whenever we have concepts which have these two features they could derive from either the utilitarian or the rational contractor point of view, or there may even be other points of view that satisfy these conditions. Kovesi himself appears to consider the moral point of view to be sui generis. Anyone who asks what is wrong with murder or stealing thereby shows that they haven't understood the rules that govern these concepts and which enable us to agree on which other concepts are moral and which are not. This, however, merely shows that there is a concept called 'wrongness' that is internal to what Kovesi calls the 'point of view of right and wrong, (1967, 26), that is, the moral point

of view. It doesn't show that this concept of 'wrongness' is one that rationally requires us not to do acts that are 'wrong'. In fact, this kind of claim about the unquestionability of the wrongness of murder is precisely what we would expect with the apparently descriptive terms that have been generated from within a single moral system. Anyone who says, 'What's wrong with murder?' either doesn't understand the meaning of the word 'murder' or else is outside the moral system in which acts that the system calls 'murder' are automatically wrong. There is, then, a danger that those within a particular moral system will refuse to debate the justification for their moral judgements on the grounds that anyone who challenges these judgements thereby demonstrates that they have misunderstood them. This is a moral version of Tertullian's paradox, with the egoist playing the same role with respect to the moral point of view as the atheist does with respect to the religious point of view. As a consequence, they run the risk of regarding themselves as possessing a faculty which enables them to follow rules that those who lack this faculty cannot follow, whereupon they begin to look very like non-naturalists. Unfortunately, however, the egoist, like the atheist, cannot accept it as a reason against his position that he just doesn't understand the view he thinks he is opposing unless he can see that believers possess a rule-following ability that he lacks that is more than just a case of verbal rule-following. In fact, there is no reason to believe the egoist or, for that matter, the atheist, cannot follow the rules of the system they oppose. They can use words in the same way as believers. They can even say how believers will act under certain circumstances if the believers are sincere. What the egoist needs to be shown is that there is something akin to the practical consequence of being colour-blind that follow from his not agreeing with the judgements of those who are committed to the moral point of view. Unless this can be shown, then there is a conflict of desires between egoists and holders of the moral point of view that can only be resolved by force. There is

a similar problem within the moral point of view. There appear to be irreconcilable differences between non-egoists over issues such as abortion, euthanasia, the scope of individual liberty and so on. Different moral theories produce different answers to these questions. Which set of non-egoists has got the rules that link concepts to particular judgements right? It is a strategy open to either side of any contentious issue to claim that the other just fails to understand the rules that govern the moral point of view. What is needed is some more objective decision-procedure by which to arbitrate between these various theories or between egoism and the 'moral point of view'.

Rawls (1951) proposed a set of conditions he believed such a decision-procedure needed to meet. His own candidate for a decision-procedure meeting these requirements he calls 'reflective equilibrium'. This decision-procedure requires that we already have consensus on some matters. As Rawls says 'Mere proof is not justification. A proof simply displays logical relations between propositions. But proofs become justification once the starting points are mutually recognised, or the conclusions so comprehensive and compelling as he persuades us of the soundness of the conception expressed by their premises. It is perfectly proper, then, that the argument for these principles of justice should proceed from some consensus.' (1972, 581) In the next chapter I will examine whether or not reflective equilibrium is an appropriate procedure by which to justify either our particular moral judgements or the theories by which we hope to render these judgements consistent.

Hare is unconvinced that Rawls' justification of principles should begin from a consensus. He says, 'It is true that any justification which consists of a "linear inference" must so proceed; but Rawls' justification is not of this type. Why should it not end in consensus as a result of

argument? There may have to be a prior consensus on matters of fact, including facts about the interests of the parties (though these themselves may conflict); and on matters of logic, established by analysis. But not on substantial moral questions, as Rawls seems to require.' He then proceeds to say that he thinks that 'moral argument can succeed in reaching normative conclusions with only facts, singular prescriptions and logic to go on'. (1975, 84-85) In the next chapter I will also look at whether Hare's decision-procedure is any better able to produce consensus than Rawls'.

Chapter 15

Reflective Equilibrium1. Introduction

A state of 'reflective equilibrium' is one where there is an harmonious fit between our considered judgements about what ought to be done and the principles from which these judgements can be derived. Our judgements are arrived at intuitively, then given due thought. Rawls says that such an harmonious fit is an 'equilibrium' 'because at last our principles and judgements coincide; and it is reflective since we know to what principles our judgements conform and the premises of their derivation.' (1972, 20)

A distinction needs to be drawn between 'personal' and 'communal' reflective equilibrium. Personal reflective equilibrium obtains when the harmonious fit is between the judgements and principles of just one individual. Communal reflective equilibrium obtains when the harmonious fit is between the judgements and principles of more than one individual. Both the advocates of reflective equilibrium discussed in this paper, namely Rawls and Haksar, are concerned with communal reflective equilibrium. They each engage in argument with other people in order to arrive at a consensus on which judgements and principles are right. With personal reflective equilibrium, an individual is not concerned to harmonise his judgements and principles with those of other people so he has no need of arguments by which to persuade them to adopt his personal balance. He merely needs to satisfy himself that the relationship he has achieved between judgements and principles is the best from his own point of view. Arriving at a personal equilibrium, then, would appear to be a much easier proposition than arriving at communal reflective equilibrium. In this chapter I want to investigate under what conditions communal reflective equilibrium is

possible and what the status is of those judgements and principles that actually achieve a communal reflective equilibrium. Does the fact that there is consensus within a particular community show anything about the rightness, or if appropriate, the truth, of these judgements and principles?

2. The Scientific Model

There are a number of models of how consensus can be achieved between individuals on some disputed matter. One of them is the method of coercion, to which impartialism is the alternative. There are other major models, namely, mathematics, sciences such as physics, and sciences such as linguistics. Hare dismisses the mathematics and linguistics models as inappropriate models by which to arrive at moral consensus. He says,

'The analogy with these sciences is vitiated by the fact that they do NOT yield substantial conclusions, as moral philosophy is supposed, on Rawls' view, to do, and in some sense clearly should.' (1975, 86)

This leaves the model of sciences such as physics. I now wish to explore what parallels there are, if any, between reflective equilibrium and physics as methods of obtaining a consensus by rational means, where 'rational means' are those means that would be adopted by people concerned to arrive at truth rather than concerned with some other purpose such as advantaging themselves at the expense of other people.

There seems, at least at first sight, to be considerable parallels. Intuitive judgements might play a similar role in moral theories as perceptions do in physics.

In physics, there is a fixed point of comparison between theories, namely, what is observed. Admittedly, each theory may have its own terms to describe what is observed, these terms being comprehensible only in the context of the theory as a whole, but this rarely, if ever, is the cause of a disagreement between theories about the ordinary language features of what was observed. No matter how a theory describes red flashes on a screen, for example, as electron traces or whatever, if the holders of the theory refused to acknowledge that the flashes on the screen looked red because their theory required that the flashes looked yellow, then this would be regarded by scientists in general, regardless of their theoretical persuasion, as a case of intellectual dishonesty. The holders of this theory may wish to argue that the flashes would look yellow if initial conditions were properly controlled, but what would definitely make their theory scientifically disreputable would be their insistence that the colour the flashes appear to be to the observer depends on the theory he holds.⁽¹⁾ Ultimately, scientific theory has to fit with how people perceive things in controlled conditions. If an advocate of a theory wishes to claim that things are not as they appear, then he does not do this by denying that this is how they actually do appear but by explaining why they appear like this.

Similar considerations apply to reflective equilibrium theories. A moral theory does not deny that people make certain intuitive judgements. Instead, it suggests that, when these intuitions cannot be fitted into the best set of principles, then these intuitions can't really be intuitions of moral

(1) This point, and the others made in this chapter, holds against writers like Werner (1983) and Flanagan (1982) who try to use the naturalist and holistic epistemology developed by Quine (1953, 1960, 1978) to defend the claim that there is as much justification to believe in a moral reality as there is to believe in a scientific one.

reality. While they truly are the intuitions that some people have they are not intuitions of the moral truth, consequently they are to be abandoned as descriptions of what morally ought to be the case, just as the description of the colour an object looks to have under abnormal conditions is to be abandoned as a description of the colour it really has.

There is, however, a radical difference between how scientific theories relate to perceptions and how moral theories relate to intuitive judgements. With a true scientific theory we could correctly predict how things would appear to people when all variables are controlled, including the internal states of the observers and the conditions under which the observations take place. The more accurately a scientific theory enables us to predict how things will appear to people under controlled conditions, the more the theory behaves as we would expect of a true theory, therefore the more rational it becomes for us to treat the essential components of the theory as true, particularly if no other theory is as successful over the same, or a similar, range of kinds of phenomena. A completely true scientific theory would enable us to correctly predict that, under certain circumstances, there would be a total consensus among people about how things appeared. The presence of the predicted consensus under specified conditions is the test of the truth of the theory. If the theory says that things have properties different from those they appear to have, then the occurrence of the predicted consensus under specified conditions is confirmation that things really have properties other than those they appear to have.

If a theory enables us to predict more accurately than any other theory the consensus that will be reached on how things appear to be under specified conditions, then there will be a consensus among rational people

concerned to believe a true scientific theory that this particular theory is closer to the truth than any of its rivals. The consensus among rational people as to which theory is closer to the truth depends on which theory enables us to predict most accurately the consensus that will obtain on how things appear to be under specified conditions in the greatest range of kinds of cases.

As we saw earlier, reflective equilibrium allows us to distinguish between how things really are morally, and how they appear to be. Reflective equilibrium, however, is not like science in providing us with a method for discriminating between theories with different claims as to which properties are the ones things really have and which are the ones they merely appear to have. It is not part of a reflective equilibrium approach to moral theories to use a theory to predict that there will be a consensus on how things morally appear to be under given conditions, then determine from the success or failure of this prediction whether it is rational to accept the theory's claims as to what properties things morally really have and what properties they merely appear to have. Reflective equilibrium, then, cannot use the presence of a consensus on how things morally appear to be to produce a consensus on how things morally are.

The significance of the difference between scientific methodology and reflective equilibrium becomes apparent when we see how each handles the case where our predictions fail, either because a consensus doesn't occur that should if the theory is true, or a consensus does occur that shouldn't. When a scientific theory fails in test conditions where there is no instrument or observer error, etc., then the theory has to be modified so that the properties it claims things actually have are made consistent with the properties that things appear to have. If a scientific theory fails to explain those cases of how things appear to be that were

inconsistent with the predictions based on it, this is a weakness in the theory. We may still prefer this theory to its competitors because they are even less successful but, nonetheless, we acknowledge that this theory has a mistake in it somewhere. When a moral theory, however, says that things morally really are a particular way yet there is no consensus that they morally appear to be as the theory requires, then reflective equilibrium does not demand that the theory be modified so that its claims about how things morally are be made consistent with how things morally appear. It does not even demand that the variables be more rigorously controlled so that we find out under what conditions we do get consensus on how things morally appear. It does not direct us to obtain enough information on the various relationships between initial conditions and consensus on how things morally appeared to be so that we can use this data base to hypothesise about how things morally really are. If we did do this however, I suspect that all we would get would be theories about the psychological causes that lead to consensus on how things morally appear to be. The use of a truly scientific methodology would lead us to theories of what empirical reality is behind people's beliefs about morality rather than theories about what moral reality is behind these beliefs. This is of little help in our current enterprise. We are after theories about what makes people's moral beliefs true or false NOT theories about what makes them have these beliefs.

Reflective equilibrium, then, does not provide us with the same kind of fixed criterion of theory preference as scientific methodology does. The consequences of this are fatal to its pretensions as a method for obtaining consensus on particular cases of moral judgement. With scientific methodology, advocates of a theory that cannot be used to predict correctly how things will appear under specified conditions have the onus of explaining away this predictive failure. They can appeal to instrument

error, observer error, unknown variables, etc. but they must make some appeal. If they can offer nothing, their theory has been shown to be seriously flawed. Reflective equilibrium places no such onus on advocates of a moral theory whose claims about how things morally are is at odds with how things morally appear to some people. The advocate of the theory can, if he wishes, take into account only how things morally appear to those who agree with him. It is as if those who believed in ESP only took into account the experiments whose results agreed with what they had observed, rather than taking experiments with contrary results as possible evidence that controls hadn't been rigorous enough, or observer error had been a factor, or some such.

Suppose Person P1 holds particular moral judgements J1 to J10 whereas P2 holds J4 to J14. Both P1 and P2 will want principles that generate J4 to J10 but it may be that some of these principles generate some or all of J1 to J3 while others generate some or all of J11 to J14. If P1 adopts the principle (maxim) M1 that generates J1 to J10 whereas P2 adopts M2 that generates J4 to J14 there is no onus on P2 to show that P1 only had intuitive judgements J1 to J10 because some variable relevant to moral perception hadn't been controlled such that, had it been, P1 would have had intuitive preference J4 to J14, thereby showing that M2 was closer to moral reality than M1. In fact, the only way that reflective equilibrium can contribute to consensus between two or more parties on moral matters where they initially disagree is if the parties themselves are more committed to arriving at consensus with each other than they are to keeping their intuitive judgements intact. If they value certain intuitive judgements more than they value consensus they will need to have these in common. This means they are only prepared to reach consensus on those intuitive judgements that matter less to them than consensus itself does. The more intuitive judgements there are that are valued more than consensus the

more theories there will be whose advocates cannot reach consensus with each other. Some of these intuitive judgements may be very far-reaching, such as Rawls' intuitive preference for liberty over utility. Anyone who prefers utility to liberty will be unable to reach consensus with Rawls, nor does Rawls have the kind of method available to science that would enable him to show that his intuitive judgements correspond with moral truth whereas the utilitarian's doesn't.

In physics, theory A may enable us to predict that flashes on a screen will appear red and theory B that they will appear yellow. It may turn out that either some people see red and others see yellow or that the same people sometimes see red and sometimes see yellow. The theories lead us to predict different results in the properties things appear to have to people because they postulate different properties that things actually have. If the advocates of A wish to accommodate the fact that some flashes appear yellow, which contradicts predictions based on A, they need to show that there is some variable present when the flashes appear yellow that, if controlled, results in the flashes appearing red, and that the real properties attributed to things by Theory A would lead one to predict that things would appear yellow to people if this variable wasn't controlled. If Theory B failed to do this, then rational people concerned with truth will prefer A to B, provided A doesn't have any other failings that B lacks. This ability to eliminate counter examples to the theory by identifying an uncontrolled variable which produces the predicted result when controlled in a way to be expected if things really are as the theory says they are is an ability not present with moral theories on a reflective equilibrium model. This means that reflective equilibrium doesn't provide moral theories with a methodology which achieves consensus by showing that one theory accommodates better than other theories a set of facts that must be accommodated by any true theory.

Reflective equilibrium allows a moral theory to ignore the fact that things do not appear morally to people as the theory says they should. Instead of attempting to find a theory that fits a set of facts that must be accommodated, a group of individuals seeking consensus can pick and choose which facts their principles will accommodate. This, however, creates the possibility of endlessly diverging groups, each of which has made a different choice at each branch in the decision-tree. While there is consensus within each group, there is nothing that pulls all the groups back towards a common theory, as the facts about how things appear to people do with a scientific theory. There is nothing that leads all groups to a consensus about what the properties of moral reality are through the avenue of a consensus on how they appear to be. In fact, there is nothing that leads to such a consensus by any other means. Even if the strongest intuition of every member of every group is the desirability of consensus, reflective equilibrium gives no clues as to which intuitive judgements are to be preferred as the basis of theorising when these intuitions clash.

The problems this generates can be seen by looking at how Haksar tries to arrive at consensus on the status of the human foetus as a rights-bearer.

3. Haksar's Strategy

Haksar believes that any human has a right to equal respect and consideration. Whenever he considers a particular case where there is a dispute about how some human should be treated, his intuitive judgement will be that this human should get the same treatment as any other. These intuitive judgements can be derived from the principle that all humans have a right to equal respect and consideration. I will call this the principle of 'human equality'. Haksar now has two problems if he is to obtain consensus. He must get people to accept his intuitive judgements in this area, and

he must get them to derive these judgements from the principle of human equality. There are, however, people who do NOT intuitively judge in accord with the principle of human equality, forexample, those who do not believe that human foetuses have the same rights as other humans.

Unlike the scientific case, this difference about how things morally appear to be does not provide Haksar with an unavoidable datum that his theory must explain predictively. There is no attempt to uncover a variable that, were it controlled, would empower the theory to predict a consensus on how things morally appear whenever the variable is controlled in this way. What we have instead is a branch in the decision-tree where the havers of each kind of intuition can go their own way unrestrained by any requirement to accommodate the differences in intuition. The search for consensus need not be abandoned here, however. Suppose Haksar puts his principle of human equality to 2000 people, and 1000 accept it because they have his kind of intuitive judgements in particular cases. Haksar has to try to obtain consensus from the remaining 1000. He cannot just postulate the principle of human equality as an axiom. Those who disagree with it will be unpersuaded by this kind of question - begging. He therefore needs to derive the principle of human equality from some higher principle but not any old higher principle will do. Whatever it is, it must prove attractive to the 1000 unbelievers. He will need a principle, then, that states that all members of some class X have a right to equal respect and consideration such that the class X includes all humans but need not be identical with the class of all humans, where "human" = 'members of "homo sapiens".'

Haksar, then, will need to find some property Q possessed by all members of X that is not logically limited to humans.

As I said earlier, not any old property of this kind will do, however. Whatever it is, it must be of a kind likely to eliminate or, at least, drastically reduce the categories of opposition to the principle of human equality. In other words, most of those who were unhappy with the principle that all members of 'homo sapiens' have a right to equal respect, etc. will need to be happy with the principle that members of the class X, all of whom possess property Q, have a right to equal respect, etc. This amounts to getting people to agree that the reason for giving beings a right to equal respect is their possession of property Q. In the first instance, Haksar will try to find a property whose presence is empirically demonstrable because this increases the chances of consensus about which beings possess it and which don't.

Reflective equilibrium, then, does find itself tested by the facts after a fashion. If there is absolutely no empirical property Q that can be embodied in a principle from which the principle of human equality can be derived, then Haksar will be forced by the facts to abandon the principle of human equality. Nonetheless, the way the facts test his moral theory is quite different to the way that facts test scientific theories.

Having decided which intuitive judgements his moral theory is to accommodate, Haksar must now find some property Q possessed by all humans that enables him to postulate a principle that all possessors of Q have a right to equal respect and consideration that will be accepted by the 1000 remaining unbelievers. If we used this method in science, we would decide which perceptions our scientific theory was to accommodate, for example, we will accommodate the cases of people who see red flashes on the screen rather than the case of people who see yellow flashes. Our choosing to accommodate those who see red flashes is a consequence of our theoretical commitment to the view that the flashes really are red, just as Haksar's commitment

to intuitive judgements giving foetuses a right to equal respect and consideration is a consequence of his theoretical commitment to the view that it morally really is the case that foetuses have such a right. Having chosen to treat the flashes as really being red, we will then look for some property Q, as Haksar does, that is common to all cases of people seeing red flashes. When we find Q, we will try to show that Q's being common to all cases of people seeing red flashes is exactly what you would expect if the flashes really were red. This still involves getting a consensus, for it involves getting the people who believe the flashes really are yellow, to see that they experience the flashes as red under conditions which would be predicted by a theory that treated the flashes as really being red but not by a theory that treated the flashes as really being yellow. It is this last step that is missing from Haksar's methodology. The 1000 unbelievers will not experience an intuitive judgement that foetuses have a right to equal respect and consideration under conditions which would be predicted by a moral theory that treated foetuses as having this right but would not be predicted by a moral theory that didn't treat them as having this right. The role, then, of the property Q, will be quite different in reflective equilibrium methodology to what it is in scientific methodology.

When Haksar finds his property Q, he will not be able to show that its being common to all cases of people experiencing the intuitive judgement that foetuses have a right to equal respect and consideration is what you would expect if it morally really was the case that foetuses had this right. He would not be able to show this because it lacks the appropriate kind of prediction. It is not predicting an experience someone would undergo who was ignorant of the theory. In fact, it usually doesn't even predict the experiences of those who are familiar with the theory because it has no interest in accommodating all the original moral intuitions

people have. It is interested, rather, in getting people to convert to new intuitions in those cases which don't fit the theory. This is as if the proponents of red flashes appealed to those who saw yellow flashes to try their hardest to see the flashes as red because this is a lot easier than changing the theory to accommodate the case of those who see yellow. If the proponents of red flashes did try this, it would show that their theory wasn't really a theory about empirical reality. Likewise, when reflective equilibrium theorists do this, it shows that their theories aren't really about moral reality. If there is a reality, and we can perceive it, then we should be able to use a consensus about how it appears to us as a means for arriving at a consensus about how it really is. Reflective equilibrium pretends to be a method for arriving at such a consensus but provides no unavoidable datum which a true theory must accommodate. As a consequence, an indefinite, possibly even an infinite, number of equally consistent moral theories can be generated with no criterion for choosing between them.

Reflective equilibrium, then, is too weak to achieve consensus about what principles should resolve conflicts of desires. It is unable to restrict the options available to disparate groups. Instead, it proliferates options. If there were a moral reality, it would NOT be able to lead us to it. Given the more likely alternative that there is no moral reality, it does not provide us with a method that, if successful, eliminates all but one uniform system of principles for deciding which desire should be satisfied in any conflict of desires.

If reflective equilibrium, then, is to help us arrive at consensus, we need to correspond in our intuitions at each of many branches of the decision-tree. Hare thinks that he can avoid this dependence on intuition by a reliance on 'facts, singular prescriptions and logic'. (1975, 85)

I shall now consider the extent to which Rawls' methodology is any less 'intuitionist' than Rawls'.

4. Hare's Methodology

In a footnote to the above quote, Hare refers us to 'Freedom and Reason' (1963), especially Chapters, 6, 7, 10 and 11 for an exposition of his views on the logic of moral argument. In these chapters, he claims that a 'moral' judgement is one in which a person universally prescribes some action. To say, for example, 'I ought to put A into prison' is, according to Hare to be understood in ordinary English as entailing the universal prescription, 'Let anyone who has done what A has done be put into prison', where the fact that I utter this prescription means that I want it to be the case that the state of affairs so prescribed comes about.

So far this thesis is a purely linguistic thesis. As such, its truth or falsity is a matter for the lexicographer to determine. Does the ordinary speaker of English think that 'I ought to put A into jail' entails 'Let anyone who has done what A has done be put into jail' (with suitable riders about the conditions being 'relevantly similar')? If he doesn't, then 'ought' can't have universal prescription as part of its ordinary language meaning and Hare's analysis of the "ordinary language" meaning of 'ought' would fail. Hare himself recognises this. (1963, 96)

However, even if Hare's analysis of the meaning of 'ought' failed to capture ordinary usage, it could still be the case that there are good reasons why people should engage in universal prescription, and this is the crucial thing that Hare needs to establish if his position is not to be just as intuitionist as Rawls.

There are three alternatives to universal prescription, namely, prescribing without universalising, universalising without prescribing and, finally, neither universalising nor prescribing. Hare provides reasons for rejecting the first two alternatives.

Prescribing Without Universalising

If a man uses the word 'ought' to prescribe but not to universalise then, in Hare's view, he is not using the word 'ought' in its moral sense. For the sake of the argument, I will accept that the word 'ought' is used morally if, and only if, it is used to make a universalisable prescription.

Consider the case where a person B says, 'B ought (non-universalisably prescriptively) put his debtors A into prison but his creditors ought (non-universalisably prescriptively) NOT put him in prison'. It may appear that he is in substantial moral disagreement with people who say 'B ought (morally) put his debtors A into prison and his creditors ought (morally) put him in prison'. In Hare's view, however, there is only a verbal disagreement between them, even though their utterances involve quite different prescriptions. He says, 'the moral, evaluative (i.e. the universal prescriptive) disagreement is only verbal, because, when the expression of B's view is understood as he means it, the view turns out NOT to be a view about the morality of the action at all. So B, by this manoeuvre, can go on prescribing to himself to put A into prison, but has to abandon the claim that he is justifying the action morally, as we understand the word "morally" '. (1963, 99)

The non-universalising prescriber, then will presumably endorse the prescription, 'Let me jail A if he owes me money but can't pay but let C not jail me if I owe him money and can't pay', whereas the universal

prescriber will presumably endorse either 'Let anyone who owes another person money and can't pay be jailed' or 'Let anyone who owes another person money and can't pay not be jailed'. Depending on which universal prescription is endorsed, the universal prescriber will either endorse 'Let me jail A if he owes me money and can't pay but also let C jail me if I owe him money and can't pay' or 'Let me not jail A if he owes me money and can't pay but let C not jail me if I owe him money and can't pay'.

Now, if Hare's position is not ultimately to rest on pure intuition, he needs to supply us with a good reason why we should endorse either of the courses of action recommended by the universal prescriber rather than that recommended by the non-universalising prescriber.

Universalising Without Prescribing

According to Hare, the moral judgement 'C ought to put me into prison' entails the singular prescription, 'Let me be put into prison'. It presumably does so because 'C ought to put me into prison' entails 'Let anyone who is in position X (C's position) Y (a verb, in this case - jail) anyone who is in position Z (my position). With the factual minor premise 'I am in position Z and C is in position X' it is possible to deduce the singular prescription 'Let me put into jail (by C)'.

This conclusion does not follow, however, if I use 'ought' universalisably but not prescriptively. This is because 'C ought to put me into prison' does not entail 'Let anyone who is in position X (C's position) Y (jail) anyone in position Z (my position)'. Consequently, even if I'm in position Z in relation to C, it doesn't follow that I am committed logically to endorsing the injunction 'Let me be jailed'.

Hare dismisses this difference between the universalising prescriber and the non-prescribing universaliser as, again, purely verbal. In one sense as with the non-universalising prescriber this is true. In neither case is the disagreement a 'moral' one, in Hare's sense. If the disagreement were kept within the 'moral' framework that Hare imposes, the non-prescribing universaliser's conclusion that he was not committed to 'Let me put in jail' would no longer hold. By definition, in merely uttering 'C ought to jail me' he has prescribed 'Let me be jailed by C'.

Hare, however, has merely shown what would follow for the universal prescriber. He has not shown us why we should be universal prescribers. What's wrong with being a non-prescribing universaliser?

It is time to turn to his answer to this question, and to what's wrong with being a non-universalising prescriber.

Why Be A Universal Prescriber?

Hare poses the question as follows.

'But if a man wants to flee from my concepts' that is, universalisable, prescriptive ones, 'where is he going to flee to? To singular prescriptions, expressing selfish desires? Or to universal but non-prescriptive judgments? He is at liberty to take either of these courses, but if he does so, he will not disturb us. For then, though we shall be in dispute with him about what to do, or about what the facts are, we shall no longer be in dispute with him about what we ought to do. We are in a position to say to him, "If you do not consent to talk on our terms, the remaining points of dispute between us will be such as can be expressed without using any terms that anybody could call moral or even evaluative. We are ready to have disputes with you of all kinds, but let us keep the kinds distinct".' (1963, 201)

This hardly solves our problem, that is, why be a universalising prescriber. True, it points out to those who are not universalising prescribers that they are not using a particular kind of 'ought', that is, a universalisable and prescriptive 'ought' but it hardly provides them with a reason to use such an 'ought'.

To provide such a reason, Hare provides the following argument.

'We can get the better of our present attacker' that is, the person who isn't a universalising prescriber 'because our language is general enough to express any dispute that he may say that he is having with us. If A has a language in which he can express everything that B wants to say, and more, then A is bound to be the winner in this philosophical game ... And our language contains means of expressing all that our present attacker could wish to say, but also means of expressing universal prescriptions, such as his language forbids him to utter. And so, when both these factions have had their say, we shall be left saying something else which they cannot express, but which we all know perfectly well how to express - namely moral and other evaluative judgements'. (1963, 202)

In one obvious way, this argument fails. Let us imagine three languages. Language 1 contains a universalisable prescriptive 'ought', language 2 contains a non-universalisable prescriptive 'ought' and language 3 contains a non-prescriptive universalisable 'ought'. Each language contains only its kind of 'ought'. If this is so, it is not at all clear how language 1 can say all that can be said in language 2 and 3. After all, in language 1 a person can't say 'You ought to do X' in a non-prescriptive universalisable way nor can he say 'I ought to do X' in a non-universalisable prescriptive way. If we have language 4, which has all three kinds of 'ought', then the universal prescriber is no better off than his attackers.

Each has three kinds of 'oughts' to appeal to if they wish.

Furthermore, even if Hare's point was true, all that has been established is that the universal prescriber can make universal prescriptions in his language, as well as non-universalisable prescriptions and non-prescriptive universalisations. Whatever the others can do he can do better, but why should the non-universalising prescriber or the non-prescribing universaliser want to make universal prescriptions? Fair enough, if Hare is correct, he will need to choose a language that has the apparatus to make universal prescriptions if he wants to make them but surely he isn't going to want to make them just because there's a language around that allows him to. There's a set of concepts around called 'theology' that allows me to make statements about God, if I want to, but why should I want to? The mere fact that the language makes it possible for me to is hardly a reason.

Hare, therefore, appears to have supplied no reason why anyone should be a universal prescriber. His preference for universal prescribers, therefore, appears to have no more basis than Rawls' particular set of intuitions.

This is not a happy conclusion to reach at the end of our examination of the justifications underlying the major two influential moral views in ethics today.

I will now turn to the question of whether there is a methodology available that can rescue impartialism from a similar charge of intuitionism.

Chapter 16

Why Be An Impartialist?1. Introduction

We have seen that neither reflective equilibrium nor universal prescription provides us with a decision-procedure that rational people have good reasons to adopt as the means of achieving consensus. Reflective equilibrium of its very nature lets rational people with conflicting desires dissent from each other in their choice of the principles to govern the conflict, provided each person can harmonise his principles with his other principles and his intuitions about particular cases. Rawls gives no reason to believe that there could not be an indefinite number of such 'reflective equilibria', each equally coherent. Universal prescription fails to give any good reasons why a person should be a universal prescriber rather than a non-prescribing universaliser or a non-universalising prescriber. Even if it did give such reasons, it still allows universal prescribers to adopt any of an indefinite number of ideals or, as Hare calls them, 'fanaticisms'.⁽¹⁾

It seems, then, that if we are to arrive at a consensus we must find a decision-procedure that is neither purely formal, as Hare claims his to be nor that already presupposes consensus on substantive moral matters, as Rawls says that his does. A purely formal procedure, as Foot (1958; 1958-9) and others have argued, admits principles that seem to have nothing

(1) Hare now argues (1976; 1981) that universal prescription leads only to utilitarianism. His arguments are criticised by McCloskey (1979) and McDermott (1983)

to do with morals, or, as I said above, lets in conflicting moral principles. A method like Rawls' that presupposes a consensus on substantive moral matters just begs the question against those who hold positions not encompassed by this consensus. In Rawls' case, even with an initial consensus, Rawls' reflective equilibrium procedure will soon have people departing from consensus down endlessly branching, but consistent, inter-connections of principles and intuitions about particular cases.

R.S. Peters (1966) provides us with an example of a different model. I do not think that Peters' own application of this model is successful but I believe it is an example of the kind of strategy we need to adopt to show why impartialism should be preferred to any of the alternative moral systems.

2. Peters' Methodology

Peters ⁽¹⁾ suggests that, 'If it could be shown that certain principles are necessary for a form of discourse to have meaning, to be applied or to have point, then this would be a very strong argument for the justification of the principles in question. They would show what anyone must be committed to who uses it seriously.' (1966, 115) Peters goes on to argue that the principles of equality, freedom and the considerations of interests are presupposed in this way by anyone who sincerely asks the question, 'What ought I to do?' or 'Why do this rather than that?'

Peters thinks that these principles are presupposed in the sense that we can only avoid being committed to them by refusing 'to talk or think about

(1) His argument here is a development of the argument in Benn and Peters (1959, 31)

what ought to be done, which would constitute an abdication from a form of thought into which all in our society are initiated in varying degrees. No adducing of reasons for the guidance of conduct would be permissible thereafter'. (1966, 116)

I will concentrate on his arguments for how engaging in discourse about how to act or what to do presupposes two principles, namely, the principle of consideration of the interest of others and the principles of freedom. I will show that this kind of discourse does not presuppose a commitment to these principles, in the sense that the person must either endorse these principles or abandon discourse about how to act. Nonetheless, the reasons why Peters' arguments fail do provide us with an understanding of the nature of moral argument and the extent to which its conclusions can escape a reliance on intuition.

3. The Consideration of Interests

Peters begins his discussion of the consideration of the interests of other people with the question,

'Why, in other words, should not a man who asks the question, "Why do this rather than that?" limit consideration of possible ends of actions to those that affect only himself? Why should he not ask, "What ought I to do considering only myself?" ' (1966, 171)

Peters' answer is worth analysing in detail. He begins,

'The answer is surely that consideration of the interests of others is a presupposition of asking the question "Why do this rather than that?" seriously.'

Prima facie, this claim is false. If I ask myself, 'What is there in it for me if I do A rather than B?' I am asking a specific example of the question, 'Why do this rather than that?' yet I am quite clearly excluding consideration of the interests of others. How does Peters propose to enforce such consideration on me? How do I presuppose such consideration?

Peters continues,

'This question, as has already been pointed out, is a question in public discourse. It presupposes a situation in which men are concerned with finding answers to questions of practical policy, in which they need the help of other men.'

In one sense, this claim is true. Such a form of words would not have entered a language unless there were issues of shared interest that warranted such a form of speech. In another sense it is false, if it is meant to show that the question is only legitimately asked when we need the help of others. All Peters can establish here is that a particular linguistic practice will not get off the ground unless it is in the common interest of the group of language speakers to have such a practice. This, however, makes language no different from any other institution, e.g. money. Unless it were in the common interest of a group to have a standard means of exchange the group would not have instituted the custom of 'money', but it does not therefore follow that a man who uses money must want the interests of others to be achieved. The public nature of the institution of money does not make the egoistic use of money irrational.

Peters continues,

'In entering into such a discussion any rational man must assume not only that there are worthwhile things to do but also that he might want to engage in such worthwhile things. If he thought that, having discussed such matters with his fellows, his stake in such a worthwhile life was going to be completely

ignored, it is difficult to conceive how he would ever take the step of engaging in such a public discussion.'

This makes clear that Peters is talking about the presupposition of public debating of why do this rather than that. It can be accepted that public debating of why do this rather than that must historically precede private debating, that is, debating with oneself or thinking through for oneself the options open to one. It isn't clear, however, why this commits one to the consideration of the interests of others.

The explanation, I believe, lies in the fact that Peters has conceived the public debate in a particular way, namely, as a debate among people who are more or less equal in power. Also, I believe, he is using the word 'consider' in an ambiguous manner.

It is true that if person A is going to feel that a debate with persons B to N about whether to do this or that is going to be worthwhile to A, then it must be the case that the debaters consider, i.e. discuss, or give attention to, the interests of all the debating members. None of them need care tuppence for each other, but in order to get help from the others on their own problems they have to be prepared to exchange advice. This will be true, anyway, if A - N are all egoists. If A - N contains A and his non-autonomous slaves B - N, then in this debate B - N will not necessarily want advice on their own affairs, although they may well wish to exchange views on how best to help A.

If A - N are egoists in a position of equal power, then Peters is right when he says that consideration of the interests of others is a presupposition of asking the question, 'Why do this rather than that?' seriously but this doesn't show, as he intends it to, that an egoist

cannot properly ask himself 'What ought I to do considering only myself?' At least, it doesn't show that the egoist cannot ask himself. 'What ought I to do considering only myself or, if others, only as a means to my ends?' He may have to give lip service to the interests of others but this hardly limits his egoism in any significant manner.

What has not been established is that the egoist is logically committed to giving his own interests no more status than anyone else's.

Peters goes on to say,

'It is not, of course, being argued that the interests of some cannot, on any particular occasion, be shown to be more important than the interests of others. The argument is only that this has to be shown.'

But to whom is it to be shown? If the question is asked privately then does A have to show to himself that his interests are more important than those of others? Why does he have to? If he is an egoist, he will already believe they are. The only sense in which he will need to show himself this, perhaps, is to convince himself that he is logically consistent, as we have already seen that he can be.

Peters seems to suggest this interpretation. He says,

'Private ploys are parasitic on public practices in the same sort of way as, more generally, a private language presupposes a public language, with all the principles (e.g. non-contradiction) that are necessary for it to have meaning and point and to be applied.'

But we have already seen how the egoist escapes self-contradiction if "oughts" are hypothetical.

5. The Principle of Freedom

Peters claims,

'If a person is asking seriously what he ought to do or what there are reasons for doing, he must obviously demand absence of interference in doing whatever there are reasons for doing.' (1966, 180)

Consider the egoist.

Is it the case that such a person, if he sincerely asks this question, must (as Peters claims he must) "obviously demand absence from interference in doing whatever there are reasons for doing?"

The problem here is - to whom is he going to make this demand? It makes little sense for him to demand absence of interference from himself, so he must be demanding it of other people. If this is so, then we need to distinguish between categories of other people, namely, egoists, coercionist idealists, or impartialists.

Suppose an egoist A demands of other egoists that they do not interfere with his pursuit of his own interests. They will be unimpressed because as far as they are concerned their interests, not his, are the only reasons for action. It seems that A would be foolish to make such a demand of other egoists unless he can back it up with force. He would be better advised to get his way by secrecy and deceit therefore it is not obviously true that he must demand absence of interference.

Suppose, on the other hand, that he makes this demand of impartialists. They would only accept his demand for freedom from interference in pursuing his interests if his interests are compatible with impartialism. However, A wants no interference at all, even in his partialist actions. Yet if he demands freedom to do what he has no impartialist right to do,

then impartialists will not accept his demand. Therefore it is again the case that such a demand would be foolish. There is no reason, certainly no obvious reason, why he must demand absence of interference, even if he wants it.

The final possibility is that he make this demand of coercionist idealists. Two types of coercionist idealists are relevant here. Those who share his own ideal and those who don't. Those who don't will treat him as the impartialists do. As the others regard his ideal as their most important goal there is no need for him to demand that they do not interfere. They would not want to. So once again, Peters' claim has no force.

Perhaps Peters means that, if a person is seriously asking what he ought, in his own overall self-interest, to do, then he must obviously want "absence of interference in doing whatever there are reasons for doing".

Again, this is not true, for a person may ask what he ought to do if he is to maximise his overall self-interest but discover that this is not what he ought to do if he is to meet the requirements of impartialism. It is not obvious that he want absence of interference in doing what he ought (in his overall self-interest) to do because he may want to be an impartialist more than an egoist. He would, then, be grateful to anyone who stopped him doing what he ought (in his overall self-interest) to do if this was incompatible with what he ought (as an impartialist) to do. He may welcome efforts to bolster his resistance to temptation in situations where his present desire overcomes what he knows in the long-run is what he really wants.

From this discussion it would appear that, if a person is asking seriously what he ought to do (in his overall self-interest), then it will only be true that he will want to be free to do what he ought to do (in his overall self-interest) if what he most wants to do is what he ought to do (in his overall self-interest).

This would appear to be true enough, but does not show he must also want freedom for others.

A similar detailed examination could be made if the person seriously asked himself what he ought to do as an impartialist. If the word 'seriously' is understood to mean that he asks the question because he most wants to do what he ought to do as an impartialist, then it will be the case that he does not want to be stopped from doing it. Here, however, there is more point to the use of the word 'demand' because within impartialism, a person has the right to do what he ought to do and the right to be allowed to do it by others.

The problem for Peters here, however, is that he has tried to show that freedom to do what there are reasons for doing is justifiable to all men. This, however, is not so, as our argument has shown. The impartialist will only accept that a person is justified in doing what there are impartialist reasons to do, whereas each egoist will accept that he in particular only has reasons for doing what is in his interests to do. Consequently, the egoist will not accept that other people should be free to do anything that must interfere with his freedom. Of course, he will not be able to give them any reason why their freedom should be limited, but they will not be able to give him any reason he finds convincing why they should not have their freedom limited (other than force). The egoist will see the issue entirely in terms of how much freedom he can force or trick others into

surrendering so that he will be better off. As a justification, then, this argument of Peters is a failure.

Let us return to the case of the person who most wants to do what he ought to do in his overall self-interest.

Peters says of him, 'Purely on ground of prudence, too, if a person is genuinely concerned about what he ought to do he would be very foolish to shut himself off from other rational beings who also have views about what there are reasons for doing. It would be even more foolish to impose constraints on others so as to prevent them giving him advice.'

(1966, 181)

We have already seen that there are at least three different kinds of rational beings who might have views about what there are reasons for doing, that is, the egoist, the coercionist idealist and the impartialist.

Would the egoist be foolish to impose constraints on members of each of these classes so that they cannot give him advice. Should he listen to them all? If Peters' argument is to go through, he must.

Does the egoist have any reason to listen to the advice of other egoists and therefore reason to allow them freedom? What kind of advice could he ask of them? As an egoist, he only wants to know what is in his own best interests. As egoists, they will only give him such advice if they believe it to be in their best interests. Each bit of advice they give him is likely to be planned to give them some advantage. He could not trust them to give him honest answers therefore there is no reason why he should want their advice and therefore no reason why he should want them to be free to give it.

Does the egoist have any reason to listen to the advice of impartialists? Provided he acts on behalf of his own interests in an impartialist way, they will give him honest and often helpful advice. The problem with impartialists, however, from his point of view is that they will also give him impartialist advice, which he doesn't want. Not only that, they will give each other advice on how to stop him doing what he has no right to do, therefore they are a mixed blessing, particularly as they are likely to act on that advice. The only advice the egoist really wants is not advice on what personal goals are desirable but how to achieve the goals he already has or how to organise them to maximise their chances of being achieved or else what are the consequences of various planned courses of action so that he can choose the one that he prefers. As a result, he may wish to live in a society neither of other egoists, nor of impartialists, but of coercionist idealists whose ideal is to service his every wish. Unfortunately for him, there are few such people, but this is no reason why he should not seek great political power, repress and render unfree the vast majority of the population while keeping relatively free a body of fawning technical advisers who are too intimidated by his power to do anything other than provide him with technical information about the means to his goal. In other words, the true egoist might want to be a Nero, or a Hitler, an absolute despot. He may even want to be a Svengali who controls their desires but leaves their minds free to calculate what will maximise his self-interest. All Peters' arguments about the need for freedom would fall to have any effect on him, not because he is irrational, but because Peters' arguments do NOT give him a reason to have reasons of a particular sort.

This distinction between having a reason and giving a reason, however, will help us see the extent to which consensus over the resolution of conflicts of desires is possible.

6. Impartialist Consensus

The arguments throughout this thesis have shown that we can divide people into two, mutually exclusive, kinds, namely, impartialists and partialists. Partialists are the kind of people who would resolve a conflict of desires in their own favour if they had superior power even though this resolution would not be adopted in a position of equal power. Impartialists are the kind of people who would not resolve a conflict of desires in their own favour just because they had the superior power if this resolution had NOT been adopted in a position of equal power. As we saw in Chapter 2, this means that impartialists are noncoercionists who want the resolution of conflicts of desires to be the ones agreed to in a position of equal power behind a veil of ignorance. Partialists are either egoists or one of a variety of coercionist idealists. There is, then, already a consensus within impartialism that is missing within partialism. Whereas all impartialists will agree on what should be done in any particular conflict of desires, unless some of them have made factual or logical errors, each egoist and each type of coercionist idealist will not agree with each other on what should be done in all conflicts of desires, or, if they do, this will be an outlandish coincidence. There will, of course, be a consensus among the adherents of each coercionist ideal. It may even be the case that, at a particular point in human history, there are more people involved in this consensus than in any impartialist consensus, although this is unlikely unless people have made logical or factual errors. Even so, a consensus that exists among a minority could still be the one that should exist between everyone.

What reasons, then, can be adduced as to why the consensus between impartialists should be extended to other groups? In order to answer this question, we need to return to a distinction I made earlier between

speculative and serious questions. A speculative question is one that the questioner asks in a particular context such as a philosophical debate where his audience does not assume that he has a genuine need to know. In ordinary discourse, when we ask a question the presupposition is that we do not know the answer. If we ask, for example, why we should perform some act it is because we are genuinely puzzled about why we should. We do not regard ourselves as having a reason for doing it and want to be given one. In a philosophical discussion, however, a speaker is not understood to regard himself as having no reason to be an egoist if he asks why shouldn't he be an egoist. No-one assumes that he is an egoist who is challenging non-egoists to give him reasons to change. No-one assumes that he is the kind of person who will sacrifice their most important desires for the satisfaction of his most trivial desire if he had the power to get away with it. This is because his question is seen as speculative. It poses a technical problem within philosophy about the nature of the kinds of proofs available for different moral positions. In a serious discussion, however, all the assumptions that are suspended in a philosophical context about the character of the questioner, are allowed full play. This thesis, of course, is an example of speculative questioning, but what I want to do now is think speculatively about people who seriously ask about the reasons for being a partialist or an impartialist.

Suppose, for example, that an egoist A wishes to obtain a consensus in which everyone agrees that all conflicts of desires be resolved as he wants them to be. If he asks another egoist B to agree with his resolutions of conflicts of desires, he will find that B may agree with A's resolution when A's conflict is with C, but not when it is with B. If he wants B to agree to all his resolutions, then he will have to force B. If he lacks force, he will not be able to give B any sincere reason why B should agree with him. The only reason he can sincerely offer B is that the resolution

of the conflict between A and B that he proposes is the resolution that favours A. As a rational egoist B will find that this fails to count as a reason from his point of view, although he will understand why it counts as a reason from A's. A, then, will not be able to get a consensus that the reason why a conflict of desires should be resolved one way rather than another is that this resolution favours his side of the conflict. At least, this is so if the people to whom he puts this proposal do not already have it as their fundamental motivation that A's desire-satisfaction be maximised. Anyone who has this attitude to A I will call his 'worshipper'. A worshipper, as I indicated earlier, is a coercionist idealist whose ideal is to maximise the desire-satisfaction of his idol, in this case, A. If the only audience A has, then, is an audience of worshippers, there will be no conflicts of desires. There may still be what I will call 'technical disputes; in that worshipper B may think that A's desire-satisfaction is best maximised in some way other than A himself thinks. If this occurs, however, it is only because either A or B has made some logical or factual error. There are no incompatible desires preventing agreement.

Unfortunately for A, it is only rarely, if ever, that his audience will contain any worshippers at all. If his audience consisted only of his worshippers, then A could openly admit to being an egoist without this producing any backlash from his audience. If his audience, however, is not entirely composed of worshippers then he will come into a conflict of desires with its non-worshipping members. With his worshippers, he had to give reasons why his desire-satisfaction would be maximised by his strategy rather than theirs, but he did not need to give them reasons why they should act so as to maximise his desire-satisfaction. With the non-worshippers, however, he will have to give them reasons why they should resolve conflicts of desires in ways that maximise his desire-satisfaction. One of these reasons could be force. If he had sufficient

power, he could offer them arguments which demonstrated his ability to invoke sanctions against them if they disobeyed him and his determination to use these sanctions. If these threats failed to provide them with reasons to act, he could then physically force them to do as he ordered. If he has this degree of power, he can let them know the real reason why he resolves conflicts of desires as he does. He can be honest. He can be sincere in what he gives as his reason for his resolution of conflicts of desires. They cannot retaliate or refuse to deal with him. On the other hand, if he doesn't have this power, it will be irrational of him to give as the reason for the resolution of conflicts of desires what is his real reason, namely, that this resolution favours his desires. In fact, he will have to disguise the fact that he is the kind of person who would use his superior power if he had it to resolve conflicts of desires always in his own favour. He will have to pretend to be motivated by reasons other than his real reasons. In public discourse, then, although a person may have egoistic reasons for what he does, he will not give these as his reasons unless he has sufficient power to be able to force people to do what is necessary to maximise his desire-satisfaction. Even if he had this power, it would only be rational of him to use it if none of his important desires would only be satisfied by people who treated him as an equal. In situations, then, where everyone has a comparable power, particularly at the level of everyday interpersonal exchanges, there will appear to be a consensus to the effect that one should not be an egoist.

The explanation of this consensus is that it is the interests of any particular person that other people act for impartialist reasons. If a person A is in a position of uncertain power, that is, he may have the advantage in certain situations but others will have it at other times, then he will not want them to act for egoistic reasons or for the reasons required by coercionist ideals other than his. If they do, they will

nearly always give his desires less consideration in a conflict of desires than an impartialist would. Any impartialist will always give even an egoist's desires equal consideration with those of other people. Another egoist B, however, will give A no consideration whatsoever other than what A has the power to force from him. This is just as true if B holds a different coercionist ideal to A. A has to ensure that his interests have their maximum chance of satisfaction even if he always has the inferior power, as he nearly always will whether he is an egoist or a coercionist idealist. This means he will have to use impartialist reasoning on other people and try to ensure that they themselves sincerely act for impartialist reasons. At least, this is true, as long as he has inferior or more or less equal power. Even if A is a member of a majority coercionist ideal, he will have to use impartialist reasoning to his fellow idealists on all matters not resolved by appeal to the ideal itself. Unless a majority coercionist ideal has a monopoly of power, it will find itself engaged in constant violent exchanges with minorities that neither it nor they can afford unless it conducts its relationships with them by offering impartialist principles to govern the behaviour of members of both groups. Where coercionist ideals have more or less equal power, the pressure to resolve disputes in accord with impartialist reason-giving becomes stronger. It will be in the interests of any particular coercionist ideal to ensure that non-members at least give it equal consideration to their own ideals.

There will, then, be a tendency to move towards an impartialist consensus as groups within society, or as societies within the world community, are forced to give each other reasons for what they do or want done because they lack the superior power to compel. Once reason-giving starts between people who are more or less equal in power, or between states where coercion is not a realistic option, then the kinds of reasons they give

will be ineffective if they are seen to be egoistic or coercionist individualist.

I set out in this section to see whether there were ways of arriving at a consensus in favour of impartialism other than by the methods of reflective equilibrium; the logic of moral language, or intuitionism. I have argued that the contingencies of human life place each individual and nearly all ideals in a position of inferior power. There will nearly always be more people not wanting what a particular individual wants than there are who do. As a consequence, it will always be in the interests of any particular individual that the kind of reason-giving that other people sincerely accept be impartialist reason-giving as it guarantees at least equal consideration of his interests, even when they have a monopoly of power. If any other kind of reason-giving is the dominant form of public reason-giving, then, whenever his interests conflict with the interests of those who sincerely accept the public form of reason-giving, his desires will be given only as much consideration as he has the power to compel.

In a position of comparable power, then, when we engage in serious rather than speculative discussion about what should be done in a particular situation, we will want the people we are talking to to be sincerely committed to giving impartialist reasons for their proposals. We may not ourselves have impartialist reasons but it will still be rational of us to talk to others as if we did, and to be seen to act as if we sincerely held them.

This does not show that we ought to be impartialists, however. It is more a case of showing how the features of our actual situation in the world lead us to adopt certain kinds of reason-giving, much as Ullmann-Margalit (1977) has shown that we are likely to adopt certain kinds of

norms such as discipline or honour in certain game-theoretic situations. As Ewin (1981, 1) says, 'the acceptance of certain qualities of character as virtues and the rejection of others as vices is forced on us by the co-operative basis of human life', that is, by the fact that most of the time each of us is in a situation where we have to give other people reasons other than our possession of superior force. We bring children up to engage in the kind of reason-giving that we want them to apply to their own cases, and to ours, even when they happen to have superior power to us, or even though they can actually get away with acting egoistically or as a coercionist idealist. As Warnock says, we train them in this kind of reason-giving 'to countervail "limited sympathies" and their potentially most damaging effects'. (1971, 26)

Nonetheless, even if all this is true, it does not show that we ought to be impartialists. In my view the question, 'Why should I be an impartialist?' is open to the same analysis as the question 'Why should I be moral?' Mackie describes the analysis of this question as follows, 'If the "should" is a moral one, "You should be moral" is tautological, and if it is anything else, say a prudential one, this statement is sometimes false, so that our question either answers itself or, having a false presupposition, admits of no answer'. (1977, 189-90) 'Should' questions, like 'ought' questions, are the question-form of hypothetical imperatives, so either the question has to be answered by a further goal or end-state other than the satisfaction of impartialism itself or else the answer, if MacIntyre's claim is correct (1981, 11-22), amounts to no more than a declaration that one wants people to follow impartialism.

It appears, then, that we either want to be impartialists or we don't, that is, that we either want conflicts of desires to be resolved as they would be in a position of equal power or we don't. There is no rational

basis for being one rather than the other. To this extent, then, which we are is a matter of intuition or natural inclination. The intuitions have been narrowed down, however, to just two.

While there may be no argument that proves that it is rational to be an impartialist and irrational to be a partialist, it is the case that argument as a way of resolving conflicts of desires presupposes that the parties are impartialists. When there is a conflict of desires between A and B where A is an egoist, then A begins from the premiss that the conflict is to be resolved in his favour whenever he has the power to do so. Argument by itself cannot get him to agree to any other resolution of the conflict. The only way to achieve an alternative resolution is to constrain him by force. The same is true of the coercionist idealist over conflicts of desires. As rational agents, both will use argument to settle technical disputes about how best to maximise their desire-satisfaction, but neither will take any account of arguments to resolve conflicts between their desires and someone else's unless required to do so by superior power. This is because a conflict of desires can be resolved by argument only if the parties to the conflict are people of a certain kind. In entering an argument about how to resolve a conflict of desires between A and B, the parties A and B have to be genuinely open-minded as to whose desires shall be favoured. In addition, they must have a particular attitude to the use of force to implement the conclusions reached by the argument. They must enter the argument sincerely accepting that, once agreement has been reached, then, if they succumb to the temptation to break it, force may be used to prevent them. A conflict of desires can only be resolved by argument if the parties to the conflict are prepared to have force used against them if they dishonour the contract. It is not just they say they are prepared to, but they must sincerely believe that this is what should be done. Now, partialists are people who do

not sincerely believe that force should be used against them to make them honour contracts to which they have agreed in a situation where no force was used to make them agree to the contract in the first place.

Impartialists, however, do sincerely believe that this is what should happen. It is because partialists of their very nature do not enter an argument about a conflict of desires which an open mind about who is to be favoured at the outcome, and are not sincerely prepared to agree to the use of force to hold them to the outcome, that they cannot reach agreement in a position of equal power on what act should be done in a particular situation. It is only if people reason in a position of equal power as if they were noncoercionists behind a veil of ignorance that they reason as if they sincerely entered an argument over a conflict of desires with an open mind as to the outcome and a commitment to honour the resolution brought about by argument. If we are sincerely concerned to resolve conflicts of desires by argument then we will reason as impartialists. In fact, we will be impartialists.

Impartialism, then, is the theory that the right act is the one that follows from principles agreed to by a certain kind of person in a position of equal power, even when real people refuse to agree to these principles. It is a consequence of impartialism, as we have seen, that real people can be forced to act in accord with principles they disagree with in real life but which, it is argued, they would agree to if they were impartialists in a position of equal power. What if a person wants to argue that he should NOT be forced to act as impartialism requires? Can argument resolve a conflict of desires between impartialists as such and partialists as such? So far, we have looked at conflicts of desires where neither of the desires in the conflict was a desire to follow impartialism. Impartialism was a method for resolving conflicts of desires of other kinds. These included conflicts of desires where these desires were NOT for a particular ideal

as well as conflicts of desires between ideals. But what of the case where one of the ideals is impartialism itself. Admittedly, on the definitions given, impartialism is a non-coercionist ideal but can it fairly be used as the arbitor of conflicts between itself and other ideals about how to resolve conflicts of desires?

Any such conflict as I have shown, could be a conflict between those who sincerely wanted to resolve conflicts of desires by argument and those who didn't. A partialist, then, who challenges impartialism is really challenging whether or not conflicts of desires are to be resolved by argument. He is saying that he wants to resolve by argument whether or not conflicts of desire should be resolved by argument. As this argument is itself about a conflict of desires, he has to at least agree that one conflict of desires should be resolved by argument, namely, this one. This conflict of desires, however, cannot be resolved by argument because no-one can enter it with an honestly open mind as to what the outcome will be. One either enters it with the view that conflicts of desires are to be resolved by argument, or one doesn't. In so far as conflicts of desires are resolvable by argument then they must be conflicts of desires other than the conflict between those who want to resolve conflicts of desire by argument and those who don't. The conflict between those who want to resolve conflicts of desires by argument and those who don't can only be resolved by force. As we have seen, however, unless one group has a monopoly of power then argument will be used as a means of resolving conflicts of desires and it can only be sincerely used if people are impartialists. Argument can, of course, be used insincerely but it will usually lose its point if it is seen to be insincere. Argument that fails to meet impartialist criteria, then, runs the risk of appearing insincere unless its partialist nature can be attributed to logical or factual errors. This is why racist argument is usually seen as insincere. It would only be

sincerely offered if people had made factual or logical errors, and it is often difficult to believe that they can have made such errors.

Impartialism, then, cannot avoid the appeal to intuition although, as I said earlier, it at least reduces our choices to one of two intuitions. Once the choice is made in favour of impartialism, moral conclusions will be derivable from the impartialist axioms and the empirical facts. Nonetheless, we will have to hypothesise moral principles which we test against the axioms. Our principles will be provisional because new cases may arise that reveal to us that our principle gives results that would NOT be acceptable to the utilitarians so we will need to devise new principles. Our procedure, though, is not like that of science. Our standard against which everything is measured is the axioms. We hope the principles we produce properly articulate what is logically implicit in the axioms as applied to particular cases but must recognise that new cases may show that we did NOT think the principle out thoroughly enough. This is where there is always scope for moral argument. Unlike reflective equilibrium, our method does not allow a multiplicity of equally coherent alternatives. In a sense, a claim that an act is right can be falsified by demonstration that it is inconsistent with the axioms.

While impartialism is only one of two intuitive jumps we can make, it is the one that must be made by people who sincerely want to resolve conflicts of desires by argument. Once people engage in argument about how to resolve a conflict of desires then they have either committed themselves to impartialism or they have entered the argument insincerely, that is, they hold a position that allows no resolution of the conflict by argument. This reinforces the likelihood that consensus can be reached on particular cases in that, once people are convinced they are committed to impartialist reasoning if they are sincere about resolving conflicts of desires by

argument, then the scope of dissensus about particular cases is reduced to that caused by factual or logical errors. With someone who admits that he is insincere in arguing for a resolution of conflicts of desires, then there will usually be little point in continuing with the argument in so far as it is an argument about principles for resolving the conflict. The impartialist conditions provide us with the conditions of fair argument about conflicts of desires. Once we know that those we are arguing with are not committed to fair argument, then the only point of continued argument is to use it as a substitute for force, which those who were deliberately arguing unfairly were doing all along.

Conclusion

The aims of this thesis have now been met. The conditions under which unforced agreement on the resolution of conflicts of desires is possible have been identified. They are that people must argue as if they are non-coercionists in a position of equal power behind a veil of ignorance. Only if they argue like this do they argue as if they have an open mind about whose desires are to be satisfied. Only if they argue like this do they show themselves the kind of people who sincerely believe that they will abide by the decision reached in a position of equal power or, if they are tempted to act contrary to these decisions, that they sincerely believe that force should be used on them to make them comply. The conditions under which unforced agreement is possible are essentially those that Rawls has elaborated in his theory of justice with the assumptions on which he relied removed. The first part of this thesis have shown that a contractarian-type theory of right does not need to make contentious assumptions about the metaphysics of personal identity, nor does it need to presuppose a particular epistemology or operate on unacknowledged perfectionist premisses about the value of autonomy or potentiality or being human. Nonetheless, the contract situation itself does impose considerable restraints on the kinds of positions people will adopt on these issues. I then proceeded to see what substantive normative principles would appear to be derivable from the impartialist axioms. The result coincided in many respects with Rawls' conclusions. It differed from Rawls mainly in those areas where Rawls' conclusions have been most strongly criticised. The thesis finished with a discussion of the rationality of impartialism, in which I argued that the conditions under which people can reach unforced agreement are the only conditions under which resolution of conflicts of desires by argument was possible. If the decisions that people make under these conditions are the ones they make as free and equal rational agents, then impartialism is the kind of theory of right that Kant and Rawls were after.

Selected Bibliography

1. Works referred to in the text

- Barrow R, Happiness, Martin Robertson, 1980
- Barry B, On Social Justice, Oxford Review, No. 5, 1967
 -- The Liberal Theory of Justice, Oxford, Clarendon Press, 1973
- Bayles M, A Concept of Coercion, in Pennock J.R. and Chapman J.W., 1972
- Benn S.I. and Peters R.S., Social Principles and the Democratic State,
 London, Allen and Unwin, 1959
- Benn S.I., Freedom, Autonomy and the Concept of A Person, Proceedings
 of the Aristotelian Society, 1975/76
- Blocker H.G. and Smith E.H. (eds), John Rawls' Theory of Social Justice,
 Athens, Ohio University Press, 1980
- Carlson G.R., Ethical Egoism Reconsidered, American Philosophical
 Quarterly Volume 10, No. 1, January, 1973
- Corrado G., Rawls, Games and Economic Theory, in Blocker H.G. and
 Smith E.H., 1980
- Daniels N. (ed), Reading Rawls, Oxford, Blackwell, 1975
- Dworkin G, Non-Neutral Principles, in Daniels N. (ed), 1975
- Ewin R., Co-operation and Human Values, The Harvester Press, 1981
- Farrell D.M., Dealing with Justice in A Reasonably Just Society,
 in Blocker H.G. and Smith C.H., 1980
- Feinberg J., Rights of Animals and Future Generations, Philosophy and
 Environmental Crisis, W. Blackstone, 1974
- Fisk M., History and Reason in Rawls' Moral Theory, Daniels N. (ed)
 1975

- Flanagan O., Quinean Ethics, *Ethics*, Vol. 93, No. 1., October, 1982
- Flew A., Who are the Equals?, *Philosophia*, Vol. 9, No. 2., July, 1980
- Foot P., Moral Arguments, *Mind*, Volume 67, 1958. Also in
Foot P., 1978
- Moral Beliefs, *Proceedings of the Aristotelian Society*,
Vol. 59 1958-1959, Also in Foot P., 1978
- Goodness and Choice, *The Aristotelian Society Supplementary
Volume*, 1961. Also in Foot P., (1978)
- Morality as a System of Hypothetical Imperatives, *Phil Review*,
Vol. 81, July, 1972. Also in Foot P., 1978
- Virtues and Vices, Oxford, Blackwell, 1978
- Francis L.P., Responses to Rawls from the Left, in Blocker H.G. and
Smith E.H. (eds), 1980
- Frankena W., *Ethics*, Englewood Cliffs, Prentice Hall, 1973
(Second Edition)
- Gauthier D. (ed), *Morality and Rational Self-interest*, Englewood Cliffs,
NJ, Prentice Hall, 1970
- Gert B., Coercion and Freedom, in Pennock J.R. and Chapman J.W., 1972
- Glover J., *Causing Death and Saving Lives*, Harmondsworth, Penguin,
1977
- Haksar V., *Equality, Liberty and Perfectionism*, Oxford, Clarendon
Press, 1979
- Hare R.M., *Freedom and Reason*, Oxford, Clarendon Press, 1963
- Rawls' Theory of Justice, in Daniels N. (ed), 1975
- Ethical Theory and Utilitarianism, in Lewis H.D. (ed),
Contemporary British Philosophy 4, London, Allen and Unwin,
1976
- Moral Thinking, Oxford, Clarendon Press, 1981
- Harsanyi J.G., *Ethics in Terms of Hypothetical Imperatives*, in
Harsanyi J.C., *Essays on Ethics, Social Behaviour and
Scientific Explanation*, Utrecht, Reidel, 1976

- Hart H.L.A., Rawls on Liberty and its Priority, in Daniels N. (ed),
1975
- Held V., Coercion and Coercive Offers, in Pennock J.R. and Chapman J.W.,
1972
- Hubin D.C., The Scope of Justice, Philosophy and Public Affairs, Fall,
1979, Vol. 9, No. 1.
- Hudson W.D. (ed), The Is-Ought Question, London, MacMillan, 1969
-- Modern Moral Philosophy, London, MacMillan, 1970
- Kalin J., In Defense of Egoism, in Gauthier D. (ed), 1970
-- Two Kinds of Moral Reasoning : Ethical Egoism as a Moral Theory,
Canadian Journal of Philosophy, Vol V., No. 3, November,
1975
- Kant I., Groundwork of the Metaphysics of Morals, translated and
analysed by Paton H.J., as "The Moral Law", London,
Hutchinson, 1948
- Katzner L.I., The Original Position and the Veil of Ignorance, in
Blocker H.G. and Smith E.H. (eds), 1980
- Kovesi J., Moral Notions, RKP., 1967
- Lewis C.I., An Analysis of Knowledge and Valuation, Open Court, 1946
- Lewis H.D. (ed), Contemporary British Philosophy 4, London, Allen
and Unwin, 1976
- Lucas J.R., The Principles of Politics, Oxford, Clarendon Press, 1966
- MacIntyre A., After Virtue, London, Duckworth, 1981
-- Moral Arguments and Social Contexts, The Journal of
Philosophy, Vol. LXXX No. 10, October, 1983
- Mackie J.L., Ethics, Penguin, 1977
- McCloskey H.J., Universalised Prescriptivism and Utilitarianism: Hare's
Attempted Forced Marriage, The Journal of Value Inquiry,
Vol. XIII, No. 1, Spring, 1979

- McDermott M., Hare's Argument for Utilitarianism, *Philosophical Quarterly*, Vol. 33, No. 133, 1983
- Medlin B., Ultimate Principles and Ethical Egoism, *Australasian Journal of Philosophy*, XXXV 1957. Also in Gauthier D. (1970)
- Montefiore A., Goodness and Choice, *Aristotelian Society*, Supplementary Volume, 1961
- Nagel T., Rawls on Justice, in Daniels N. (ed), 1975
- Nozick R., Coercion, in Morgenbesser S. (ed), *Philosophy, Science and Method*, St. Martins Press, 1969
- Parfit D., Later Selves and Moral Principles, *Philosophy and Personal Relations*, ed. A. Montefiore, London, 1973
- Personal Identity, *Philosophy of Mind*, ed. J. Glover, O.U.P., 1976
- Pennock J.R., Coercion : An Overview, in Pennock J.R. and Chapman J.W., 1972
- Pennock J.R. and Chapman J.W. (eds), *Coercion*, Chicago, Aldine, Nomos XIV, 1972
- Peters R.S., *Ethics and Education*, London, Allen and Unwin, 1966
- Quine W.V.O., Two Dogmas of Empiricism, in Quine W.V.O., *From a Logical Point of View*, Harper Torch Books, 1953
- *Word and Object*, Massachusetts, M.I.T. Press, 1960
- Quine W.V.O. and Ullian J.S., *The Web of Belief*, Ind. Ed., New York, Random House, 1978
- Rawls J., Outline of a Decision-Procedure for Ethics, *The Philosophical Review*, 60, (1951)
- Justice as Fairness, *Philosophical Review*, Ixvii (1958)
Also in P. Laslett and W.G. Runciman (eds), *Philosophy, Politics and Society*, Second Series, Blackwell, 1962
- *A Theory of Justice*, Oxford, Oxford University Press, 1972
- Kantian Constructivism in Moral Theory, *Journal of Philosophy*, Vol. LXXVII, No. 9, September, 1980

- Richards D.A.J., A Theory of Reason for Action, Oxford, Clarendon Press, 1971
- Singer P., Practical Ethics, Cambridge, Cambridge University Press, 1979
- Taylor C.C.W., Critical Notice of R.M. Hare's Freedom and Reason, Mind, 1965. Also in Wallace G. and Walker A.D.M., 1970
- Ullmann-Margalit E., The Emergence of Norms, Oxford, Clarendon Press, 1977
- Wallace G. and Walker A.D.M., The Definition of Morality, Methuen, 1970
- Warnock G.J., The Object of Morality, Methuen, 1971
- Werner R., Ethical Realism, Ethics, Vol. 93, No. 4, July, 1983
- Wertheimer A.P., Political Coercion and Political Obligation, in Pennock J.R. and Chapman J.W., 1972
- Williams B., Ethical Consistency, in Williams B., 'Problems of the Self', Cambridge, Cambridge University Press, 1973
- Wolff R.P., Understanding Rawls, Princeton, Princeton University Press, 1977
- Zimmerman M., The 'Is-Ought' : An Unnecessary Dualism, in Hudson W.D. (ed), 1969

2. Works consulted

- Baier K., The Moral Point of View, Random House, 1965
- Barry B., Political Argument, RKP, 1963
- Beatty J., The Rationality of the Original Position; A Defense, Ethics 93, April, 1983
- Beauchamp T., Distributive Justice and the Difference Principle in H.G. Blocker and E.H. Smith, 1980
- Bourke U.J., Ethical Role of the Impartial Observer, Journal of Religious Ethics, Vol. 6, Fall, 1978
- Bowie N., Equal Basic Liberty for all in H.G. Blocker and E.H. Smith (eds), 1980
- Brandt R.B., A Theory of the Right and the Good, Clarendon Press, 1979
- Buchanan A., Marx and Justice : The Radical Critique of Liberalism, Methuen, 1982
- Cooper D., Equality and Envy, Journal of Philosophy of Education, Vol. 16, No. 1, 1982
- Daniels N., Equal Liberty and Unequal Worth of Liberty in N. Daniels, 1975
- Moral Theory and the Plasticity of Persons, Monist, Number 62, July, 1979
- Wide Reflective Equilibrium and Theory Acceptance in Ethics, Journal of Philosophy, 1979, pp.256-283
- Dworkin R., The Original Position, in N. Daniels, 1975
- Taking Rights Seriously, Duckworth, 1977
- Edgley R., Reason in Theory and Practice, Hutchinson, 1969
- Feinberg J., Rawls and Intuition, in N. Daniels, 1975

- Feyerabend, Against Method, Verso, 1975
- Fishkin J.S., Tyranny and Legitimacy, John Hopkins University Press, 1979
- Galston W.A., Justice and the Human Good, University of Chicago Press, 1980
- Gardenfors P., Fairness Without Interpersonal Comparisons, Theoria, 2, 1978
- Gauthier D.P., Practical Reasoning, Clarendon Press, 1963
- Gert B., The Moral Rules, NY., Harper & Row, 1966
- Grice G.R., The Grounds of Moral Judgement, C.U.P., 1967
- Hampton J., Contracts and Choices : Does Rawls Have A Social Contract Theory, Journal of Philosophy, Vol. LXXVII, No. 6, June, 1980
- Hare R.M., Language of Morals, O.U.P., 1952
- Henberg M.C., Impartiality, Canadian Journal of Philosophy, Vol. 8, December, 1978
- Howard N., Paradoxes of Rationality, MIT Press, 1971
- Kekes J., Morality and Impartiality, American Philosophical Quarterly, Vol. 18, October, 1981
- Lakatos I. and Musgrave A., Criticism and the Growth of Knowledge, Cambridge University Press, 1970
- Locke D., The Principal of Equal Interests, Phil Review, Vol. 90, 1981
- Lyons D., Nature and Soundness of the Contract and Coherence Arguments in N. Daniels, 1975
- Nagel T., The Possibility of Altruism, Clarendon Press, 1970

- Nozick R., *Anarchy, Utopia and the State*, Blackwell, 1974
- Papineau D., *Theory and Meaning*, Clarendon Press, 1979
- Paul J. (ed), *Reading Nozick*, Rowman and Littlefield, NJ., 1981
- Quine W.V.O., *Theories and Things*, Belknap Press, 1981
- Rawls J., *Two Concepts of Rules*, in P. Foot (ed), *Theories of Ethics*, OUP, 1967
- *Fairness to Goodness*, *The Philosophical Review*, 84, 1975
- *The Basic Structure as Subject*, *American Philosophical Quarterly*, 14, 1977
- *Social Unity and Primary Goods*, Sen and Williams, 1982
- Robinson H.M., *Is Hare a Naturalist*, *Phil Review*, XCI, No. 1, January, 1982
- Ryan C.C., *The Normative Concept of Coercion*, *Mind*, 1980, Vol, LXXXIX
- Sandel M.J., *Liberalism and the Limits of Justice*, Cambridge University Press, 1982
- Scanlon T.M., *Rawls' Theory of Justice* in N. Daniels, 1975
- Schaar J.H., *Equality of Opportunity and the Just Society*, in H.G. Blocker and E.H. Smith, 1980
- Schaeffer D.L., *Justice or Tyranny*, Kennikat Press, 1979
- Sen A. and Williams B., *Utilitarianism and Beyond*, C.U.P., 1982
- Shaw W.H., *How to do Ethics : A Question of Method*, *Metaphilosophy*, Vol. 13, No. 2., April, 1982
- Sterba J.P., *The Demands of Justice*, University of Notre Dame Press, 1980
- Van De Veer, *Of Beasts, Persons and the Original Position*, *Monist*, Number 62, July, 1979

- Wagner R.H., Impartiality and Equity, Theory and Decision, 12, 1982
-- Fishkin on Tyranny and Structural Principles of Justice,
Ethics 93, October, 1982
- Whitlock F.A., Criminal Responsibility and Mental Illness, Butterworths,
1963
- Williams B., The Self and the Future, Phil Review, 1980
- Zimmerman D., The Force of Hypothetical Commitment, Ethics 93,
April, 1983