FORM AND SUBSTANCE IN R.M. HARE'S UTILITARIANISM

THESIS

Submitted in fulfilment of the

requirements for the Degree of

MASTER OF ARTS (PHILOSOPHY)

of Rhodes University

by

PIETER HENDRIK COETZEE

December 1992

ABSTRACT

Throughout his career as moral philosopher Hare has insisted that there is a rational way of arriving at substantive moral judgements. Hare develops this view - first presented in 'The language of morals' (1952) and 'Universalizability' (1955) - into the claim that rational agents are required to adopt utilitarian solutions to moral disputes. In 'Freedom and reason' (1963) this claim is defended with reference to the view that the formal features of moral language (universalizability and prescriptivity) commit moral agents to a certain method of reasoning, and that this method of reasoning, when conjoined with facts about people's desires and preferences, leads us to accept substantive moral judgements consistent with those required by a form of utilitarianism. This view features throughout Hare's subsequent work, but the argument for it undergoes change. This means change in the defence of the claim that the meta-theory Universal Prescriptivism is consistent with a form of normative utilitarian theory, as this claim is argued for in 'Ethical theory and utilitarianism' (1976) and 'Moral Thinking' (1981).

I shall endeavour to trace the chronological development of Hare's thinking, and will concentrate on developments in the argument for a theory of act-utilitarianism. I shall argue that the argument for utilitarianism gives rise to two major problems which arise from a specific feature of the argument, namely, the attempt to run the resolution of bi-lateral and multi-lateral cases of conflict along lines analogous to the resolution of conflict in the single-person case. Hare's argument requires that a decision-maker must identify the person with whom he reverses roles as himself, and that he must be prepared to concede that the things his

recipient has good reasons for wanting are also reasons for him to want the same things. I argue that it is not possible to make coherent sense of the identity of the person in the reversed-role situation and that the motivational states a decision-maker is expected to deem 'his own' are not properly states of himself. If I am right, the 'identity'-question sits at the root of a motivational gap in Hare's theory.

TABLE OF CONTENTS

INTRODUCTION

Statement of the problem

Hare believes that a proper account of the nature of moral reasoning will lead one to accept substantive moral principles; and in particular it will lead one to accept a form of utilitarianism. The object of this thesis is to trace Hare's argument for this claim and to evaluate it.

I will argue that Hare's analysis of the nature of moral reasoning does not entitle him to this claim. I will do this by outlining the development of Hare's moral philosophy, and will concentrate on

(1) the article 'Universalizability' (1955)
(2) 'The language of morals' (1952 - revised 1961)
(3) 'Freedom and reason' (1963)
(4) the article 'Ethical theory and utilitarianism' (1976)
(5) 'Moral thinking' (1981)

Throughout Hare's writing there is an insistence that one must understand the logic of moral language if one is properly to understand the nature and limits of moral argument; and if one wants to adequately defend substantive moral judgements. Hare believes that an understanding of the logic of moral language commits one to a theory he calls 'universal prescriptivism', and that this theory is the most adequate theory of moral discourse. Hare also believes that the method of reasoning which universal prescriptivism generates, taken in conjunction with other things he sees as necessary features of moral argument (facts, imagination, inclinations, desires, preferences), leads to rational acceptance of a form of utilitarianism - preference act-utilitarianism in particular.

A skeleton outline of Hare's theory of moral reasoning includes the following kinds of claim:

(a) Singular moral judgements are universalizable i.e. they entail identical judgements about identical situations.
(b) Moral judgements are prescriptive in the sense that one's acceptance that x ought to be done entails the prescription 'Let x be done'.

These are what Hare calls the basic logical properties of moral language. However, the method of reasoning which this analysis generates will not - by itself - commit one to the acceptance of any substantive moral principles. We need, in addition, knowledge claims about

(c) one's own and other people's inclination/desires/preferences, and
(d) a certain power of imagination.

We begin our moral deliberation by reflecting on what we are committed to in virtue of our understanding of the logical properties of moral discourse. Any prescription we wish to offer must rest on a knowledge of the relevant facts of the situation we are judging. These facts include an awareness of the possible consequences of acting on the prescription, i.e. an awareness of how we and other people would be affected. The prescription must also be universal, i.e. it must be the one we are willing to act on were we the affected persons, i.e. were we in their positions with all and only the relevant universal features of their persons and circumstances. To arrive at a moral conclusion we must desire/prefer - on balance - that a certain thing be done or not be done in all relevantly similar situations, which include hypothetical situations in which we perform the role of each affected party in some random sequence.

If Hare is right, we are morally obliged to accept that line of action which we would accept if we were to be the persons in these situations. Hare claims that this line of action will maximally satisfy the desires/preferences of all concerned parties, i.e. it will be the course of action which a utilitarian would endorse.

There is a recognisable sense in which all of the ingredients mentioned above and the picture just sketched feature throughout the development of Hare's theory. But these things undergo changes of statement and elaboration, changing defences and changes in emphasis. This means that the attempt to marry universal prescriptivism to utilitarianism undergoes various changes. The most dramatic of these changes occurs in Moral Thinking (1981). Hare introduces a new thesis, which I call simply the Knowledge Requirement, and which Hare defends as a conceptual truth. According to this thesis our knowledge of other people's preferences entails our having the same preferences. As I judge, the introduction of this thesis marks a departure from a claim Hare advances in Freedom and reason and Ethical theory and utilitarianism to the effect that all the moves he needs to make to defend utilitarianism derives from a consideration of the logical features of only the moral concepts. In discussing Hare's work, I will attempt to show how his thinking develops and changes over time. I will also attempt to show how Hare defends his methodological assumptions and what necessitates changes in his thinking. My argument throughout will be that Hare fails to bring about the desired marriage between universal prescriptivism and utilitarianism.

CHAPTER 1

The initial theory of moral reasoning as presented in 'The language of morals' (LM) and 'Universalizability' (U)

Introduction

Hare insists that a crucial condition for developing an adequate theory of moral reasoning is a proper understanding of what he calls the logical features of moral language. These are universalizability and prescriptivity. They are the formal properties of moral language on which a theory of moral reasoning can be founded.

My object in this chapter is to explore Hare's initial understanding and defence of these features of moral language and to see how they work in his initial theory of moral reasoning. I shall concentrate on examining a central claim in this initial theory, which is that moral judgements are a class of universalizable prescriptive judgements. There are two premises which jointly support this claim:

(1) If some agent, A, accepts the moral judgement that x ought to be done in S, he is thereby committed to the view that anyone else in circumstances relevantly similar to S ought also to do x.[1]

(2) 'X ought to be done', if used evaluatively, entails 'Let x be done'. If an agent assents to the moral judgement 'X ought to be done', he logically cannot dissent from the prescription 'Let x be done'.[2]

The plan of this chapter is briefly as follows: I examine Hare's defence of universalizability and prescriptivity in sections 1 and 2 respectively.

My main contention will be that the so-called u-type requirements of universalizability demand more than mere applications of the principle of consistency, and that Hare's defence of universalizability, which proceeds via his account of supervenience, does not show what supports these requirements. In section 3 I examine the place of universalizability and prescriptivity in Hare's critical theory of moral reasoning. I concentrate on the notion of a decision of principle and argue that Hare's initial theory admits a form of subjectivism.

Section 1. Universalizability

It is, according to Hare's argument in U, a feature of the logic of evaluative judgements that they require justification. Their justification always involves, as major premiss, a moral principle to which appeal is made.[3] Moral principles, however, are of a peculiar nature. They must be formulable without using terms which function as individual constants and so their most distinctive feature is that they make no reference to particular individuals. Moral judgements, because they are moral, are 'u-type valuations', i.e. they are applications of

> '... a rule wholly devoid of personal references, a rule containing only predicates (descriptions) and logical terms'.[4]

Being supportable thus is what Hare means by saying that they are universalizable.[5] Expressed in the form of a principle, Hare makes the following claim:

(UP) If A accepts the moral judgement that x ought to be done in S, this implies that he accepts that x ought to be done in all circumstances

relevantly similar to S indifferently of who is actually who in these circumstances.

Suppose C is A's reason for holding that x ought to be done in S. Stated in terms of reasons for moral judgements, (UP) reads like this:

C is a reason why x ought to be done in S if and only if A accepts that it is also a reason for anyone else in relevantly similar circumstances to do x.

Hare's thesis is not that C <u>is</u> a reason for anyone else in relevantly similar circumstances, but rather that <u>A</u> is committed to this view.[6] In U, Hare's defence of (UP) rests on the claim - I call it Q - that moral judgements are u-type valuations. Another way of stating this is to say that whenever moral reasons are offered from some choice or action, they invoke or apply u-type principles. According to Hare, Q is necessarily true, i.e. true in virtue of the meaning of the word 'moral'.[7] This implies that moral reasons necessarily invoke or apply u-type principles. So, a moral reason just is a reason which makes no reference to particular individuals. It follows, Q being analytic, that to deny the truth of (UP), is to misunderstand the meaning of the word 'moral', and to contradict oneself.

Hare attempts to develop his thesis that moral judgements are u-type valuations into a distinction between the sorts of reasons that can be given for moral and other value judgements. This involves his distinction between e-type and u-type reasons. In U Hare is careful to point out that nothing in his thesis turns on the meaning of the word 'reason'. Hare allows that whenever a reason is given for some ought-judgement, the reason

may be of type-e, i.e. it may refer to a particular individual and still qualify as a reason.[8] It is analytically true that giving a reason involves reference to a rule or principle, but this rule or principle need not be of type-u. In giving a reason one says something about an action or a state of affairs, and this 'something' may involve a reference to a particular individual.

I use one of Hare's examples to illustrate this point.[9] 'It [the chancellor's squeeze] resulted in an improvement in Great Britain's balance of payments', is a reason of type-e, if offered by someone who thinks that it is a reason only if the squeeze improves Great Britain's balance of payments, and not if the same action in another country in qualitatively similar circumstances, improves that country's balance of payments. As Hare makes clear, to say that anyone in the chancellor's position ought to do x, if x improves Great Britain's balance of payments, is not universalizable in the sense required by (UP), since the reason refers to a particular individual, Great Britain. If the reason is a reason for anyone who is situated like Great Britain's chancellor in the relevant respects, i.e. if it is a reason for anyone who is Great Britain's chancellor, it cannot be of type-u.

Now, Hare argues that the reference to Great Britain is eliminable and that the reason may therefore be extended to relevantly similar situations (with the necessary adaptation). Hare allows that there is no logical barrier which prevents this. In this regard Hare argues, against Gellner, that e-type valuations are, in principle at least, translatable into u-type valuations.[10] To use Hare's 'balance of payments' example again, the claim is this: if the chancellor of some other country stands to Great Britain's chancellor in a similar relation vis-á-vis the balance of payments in their

respective countries, i.e. if his position in his country is like the position of Great Britain's chancellor in Great Britain in the relevant respects, then the principle that in circumstances of the kind in question, the squeeze is right if it improves the balance of payments, is properly universal, which means that any reference to particulars in the reason for the squeeze is, in principle at least, eliminable.

Hare recognizes that the suggested elimination is problematic.[11] One problem is to find or coin suitable universal terms as substitutes for the use of the expression 'like x', where 'x' is an individual constant. Another problem concerns designing precise descriptions of the similarity (where he means similarity only in the relevant respects) implied by the use of the word 'like'. In this regard Hare refers to a confusion in common uses of the words 'general' and 'universal'.[12] He holds that 'universal' is opposed to 'singular' or 'particular' (though the latter is ambiguous), and that 'general' is opposed to 'specific'. He suggests that generality and specificity are matters of degree (we can be as specific or as general as we like), but that universality and its opposites are not. He argues that moral principles may be highly specific and admit of exceptions or qualifications (in what respects and to what degree we specify this is a matter of choice), while remaining universal in the required sense.

In brief Hare is suggesting this: if anyone says that C is a reason why Great Britain's chancellor ought to do x in S, but he thinks that C is a reason only if it refers to some particular individual, Great Britain, then his use of the word 'ought' is not a moral use. There cannot be a sense of universal in which it is legitimate for Great Britain's chancellor to claim that C is a reason why anyone who is situated like him in the relevant

respects, ought to apply the squeeze, if the reference to Great Britain is included in the specification of the relevant respects. The relevant sense of the word 'universal' is simply this: a reason is universal if and only if there is no reference to individual constants in the predicate part of the reason. In this sense a reason for a moral judgement is simply the universalized form of the judgement, or the universal principle which the reason invokes. This, however, does not mean that my reasons <u>are</u> reasons for you. If C is my reason why I ought to do x in S, I am committed to the view that C is also a reason for you, were you in relevantly similar circumstances. It does not follow, and Hare does not claim, that you <u>have</u> the same reason to do x in S.

Now, I think (UP) may be interpreted as demanding consistency in applying moral judgements.[13] According to (UP), if features a, b, c ground the judgement that x is good, or that x is obligatory, then any other y, which also has features a, b, c (and no other relevant features) is also good or obligatory. But, (UP) does not demand consistency and nothing more, i.e. mere consistency in applying moral judgements employing the words 'good' and 'ought' is not sufficient evidence that their uses are moral uses. It seems that though consistency in applying moral judgements is a requirement of (UP), the demand for consistency does not arise as a consequence of Hare's definition of moral judgements as u-type valuations. The demand arises rather from the notion that moral and other value judgements are the sorts of judgement for which reasons can be given. I shall now restate (UP) for both 'good' and 'ought' in an attempt to lift out the requirement of consistency.[14]

(1) If A judges x to be good, then he is committed to the view that anything like x is also good, where 'like x' is similarity in respect of the non-moral but morally relevant properties of x.

(2) If A judges that x ought to be done in S, then he is committed to the view that x ought to be done in all situations like S, where 'like S' is similarity in respect of the non-moral but morally relevant properties of S.

(1) and (2) invoke the requirement of consistency and make explicit Hare's view that there are always reasons for moral judgements, which reasons mention the relevant properties or features of the thing or situation about which the judgement is made.[15] But, by itself the claim that moral judgements are made for reasons does not imply that the relevant properties are only those which can be stated in universal terms. If (1) and (2) are interpreted as asserting only that there is something logically amiss or contradictory about making different moral judgements about relevantly similar situations, then they assert no more than the principle of consistency.

One sense in which the principle of universalizability may be understood might be expressed by saying that the principles which support moral 'ought'-judgements apply without exception to anyone who falls within their scope. Thus, if A says that another person B ought to keep his promises, but denies that he is committed to the view that he himself ought to keep his promises, if and when he promises, then, unless he can show grounds to distinguish his case from B's, his denial is evidence that he does not know the meaning of the word 'ought' in its <u>moral</u> uses. The point is that u-type principles, if they apply, make no exception of particular individuals.

But I see no reason to accept that all u-type principles which are offered as reasons for ought-judgements are moral principles. For example, the principle 'One ought always to plant Erica's in acid soil' is u-type, but it is not a moral principle. What really is the criterion which distinguishes moral uses of 'ought' from non-moral ones? One might ask: What supplies the sufficient condition for an 'ought'-judgement to be a moral judgement?

I have so far attempted to clarify the sense in which moral judgements are u-type valuations. If my exposition of Hare's views in U is correct, then his thesis that moral reasons invoke u-type principles (for both 'good' and 'ought') rests on two claims.. These are:

(a) If x is good, this must be because of something about x, which implies that it is appropriate to ask for a reason why x is good;

(b) The reason why x is good can refer to a particular individual, i.e. it can be of type-e. But if this is so, then the judgement is not a moral judgement. The judgement is a moral judgement if and only if it is a u-type valuation. This is because of the meaning of the word 'moral', i.e. the claim that moral judgements invoke u-type principles, is analytic.

Now, I think Hare's attempt to develop his distinction between the sorts of reasons that can be given for moral and other value-judgements, relies in part on his account of supervenience. The link between (UP) and supervenience touches on a claim alluded to above: if it is true that there are always reasons for moral and other value-judgements, this must be because of some feature of the thing or situation about which the judgement

is made. For Hare's account of supervenience I turn to LM. Hare's notion of supervenience is this:

(SP^1): Two things cannot differ only in their value.[16]

I shall use Hare's 'painting'-example in LM to illustrate what (SP) entails.[17] Suppose that we are invited to judge whether two paintings, P and P_1 are good works of art. Suppose also that either P is a replica of P_1, or that P_1 is a replica of P, but that we do not know which is which. Now, in judging whether or not P and P_1 are both good works of art, there is one thing we cannot maintain, viz, that P and P_1 are exactly similar in all respects except that P is a good work of art and P_1 not. There is nothing logically amiss with saying that P and P_1 differ in this respect that P is signed and P_1 not, though they are exactly similar in all other respects. So, as Hare observes, why can we not say that P is good and P_1 not, though they are exactly similar in all other respects? Hare's answer is that there is a difference in the logic of 'good' and 'signed'. In the case of 'good' and other 'value'-words we account for different judgements about things with reference to differences of another kind. This concerns the so-called supervenient character of 'good' and other 'value'-words.

Suppose P is a good painting because P is C, where C is the conjunction of all the descriptive properties of P. Let G be the supervening property 'goodness'. In LM Hare maintains the following view:

(SP^2) If P and P_1 are exactly similar with respect to C, it follows necessarily that they are exactly similar in all respects pertaining to their value.[18]

The characteristic of 'good' which Hare calls its supervenience comes to this: P and P_1 cannot differ in any respect pertaining to G unless they also differ in some respect pertaining to C, i.e. differences with respect to G cannot vary independently of some difference with respect to C.[19]

Before I examine Hare's defence of (SP), I wish to note two features of this notion. (SP) implies that there is a relation of dependence of the property 'G' on the set or conjunction of properties 'C'. To say that P and P_1 are both good <u>because</u> they have certain specified features is to say that, <u>if</u> they have those features, they must be good. One might say that a judgement to the effect that they are good paintings depends on their having the specified features. Put more generally, Hare's point is that for any singular evaluative or moral judgement it is never logically amiss to ask in virtue of what something is good (bad, obligatory, right, wrong), i.e. it is necessarily true that if something is good, it is good because it has some specified set of descriptive properties. How does Hare account for this dependence-relation? One might ask: What explains the supervenient character of 'good'? This is to ask, in effect, how the 'because'-clause above is to be understood.

(SP) invokes the requirement of consistency. By (SP), it is inconsistent for anyone to say that P is good but P_1 not while acknowledging that there is no other difference to account for this. But there is nothing logically amiss in maintaining that P and P_1 are both bad, while acknowledging that they are exactly similar in all other respects. This is to say, if (SP) is true, then P and P_1 are either both good or both bad, depending on what standard of goodness one accepts.[20] Now, in so far as (SP) requires consistency, there is a link with (UP) as stated above in (1) and (2). To see this, suppose that P and P_1 are both good paintings of some school and

that it is generally accepted that good paintings of this school have the 'good-making' characteristics a, b, c. Suppose also that P and P_1 differ in some respects, for instance, as regards size and shape. It is obvious that if P and P_1 differed only in the latter respects, they could not differ in those respects relevant to their being good, at least no less than what two other good paintings of this school, which share all their descriptive properties, could differ in those respects relevant to their being good. There is no reason to suppose that P and P_1 must share all their descriptive properties to be good paintings of this school. (SP) requires some notion of relevant similarity.[21] So (SP) can be reformulated thus:

(SP^3) If P is good, then anything else like P is also good, where 'like P' is similarity in respect of the relevant descriptive or non-moral properties of P.[22]

(SP^3) makes explicit Hare's view that whenever a moral or value judgement is offered, it is always logically appropriate to ask for a reason. To say that 'goodness' is a supervenient property is to say that moral and other value-judgements employing the word 'good' are made for reasons which mention the relevant non-moral or descriptive properties of whatever is being judged. But, by itself, (SP^3) does not say what kinds of properties the relevant properties must be for the reason to be a reason for a moral or value-judgement. (SP^3) excludes only the view that the property 'goodness' itself can be a reason for a moral or value-judgement.[23] But, one might ask: What kinds of properties are appropriate instantions of a, b, c?

In U Hare maintains that only properties stateable in universal terms can be morally relevant properties. Hare, of course, carries this point by definition of the word 'moral'. This suggests that he is defending the supervenient character of 'value'-words like 'good' and 'ought', as regards their moral uses, with reference to the meaning of the word 'moral'. But what about non-moral uses of 'good' and 'ought'? In U Hare seems to be in doubt about whether e-type judgements can properly be called 'valuations', though he suggests that there are no grounds in common language to support the claim that a reason qualifies as a reason only if it invokes a u-type principle.[24] Hare seems to be adhering to a position adopted in LM that any account of the supervenient character of 'value'-words like 'good' and 'ought' must involve an account of how moral and other value-judgements employing these words can have reasons. For in U Hare is suggesting that e-type judgements invoke principles and that these principles can be reasons for ordinary value-judgements, though not, of course, for moral judgements. I think Hare's attempt to distinguish between u-type and e-type judgements is unsuccessful because there is no reason to suppose that ordinary value-judgements cannot be universal in the sense Hare wishes to reserve for moral judgements. So, we may ask again: what reason do we have to suppose that the kinds of properties which can be morally relevant are only those stateable in universal terms? Does the notion of supervenience help to account for this?

In LM Hare defends his account of supervenience with reference to the view that evaluative and moral language are used to guide choices and decisions. If Hare is right, the supervenient character of 'value'-words like 'good' and 'ought' is readily accounted for by the commendatory or prescriptive function which they perform in their ordinary uses. Other attempts to

explain their supervenient character either misconstrue this function or do not account for it at all.[25]

How does Hare's talk of the use of moral and evaluative language help him to explain the supervenient character of 'good'? I shall first simply state Hare's position.[26] Suppose that it is generally accepted that P is a good painting of its school because P exhibits the conjunction of properties C. The position Hare argues for is this: There is a necessary connection between the statement that P is C and the judgement that P is good, if we grant the further premiss that anyone who says that P is good because P is C also accepts the universal principle that all paintings of this school which conform to the given description are good, where his acceptance implies that he treats the principle as a prescription which has the force of a universal imperative. This is to say, a necessary connection obtains if his acceptance is an instance of the evaluative use of 'good' in the universalized form of his judgement. It is the necessary condition of the supervenience of 'good' in this context that there be something about paintings of this school which are the subjects of the judgements that they are good. It is the sufficient condition that the principle be accepted and asserted in the above sense. This is what makes it possible for the statement that P is C to be someone's reason why P is good, and hence a reason for him to commend paintings which conform to the given description.

This position rests on Hare's distinction between evaluative and descriptive meaning, and on the reasons he offers for the primacy of the former.[27] I begin with the notion of descriptive meaning and the role it plays in Hare's account of supervenience. If one were to teach the criteria for applying the word 'good' to P, one would in effect be teaching

that some set of describable characteristics of P is what makes it a good painting of its school. The task of teaching these criteria could not be successfully done unless one were consistently to call good all paintings of this school which have the required set of 'good-making' characteristics. This is a matter of being consistent about what standard applies whenever one says that a particular painting is a good painting of this school.[28]

Now, this feature of 'good' that it can be used purely to convey information about what characteristics of P make it a good member of its school, is what Hare calls the descriptive meaning of 'good'.[29] It seems that to teach the criteria for applying the word 'good' to P is, in effect, to teach the descriptive meaning of the expression 'A good P', and that within the class of all good paintings of this school the descriptive meaning of this expression is determined by the standard which makes certain features of these paintings the 'good-making' characteristics of this school. Hence, to know what the descriptive meaning of this expression is, is to know what standard is applied in this context.[30] If the standard is well-known or generally accepted, the descriptive meaning of the expression 'A good P' will be fixed to the extent that the standard is known or accepted. One might in this context speak of a <u>standard</u> meaning of the expression 'A good P', but only in the sense that the word 'good' can be used purely to convey information about a generally accepted standard.[31]

How does this help Hare to explain the supervenient character of 'good'? The suggestion seems to be that it is a characteristic of 'good' that its use in some judgement to convey information about 'good-making' characteristics, or about a standard of goodness, makes it possible that a

reason can be given for the judgement. I think Hare uses the informative function of 'good' as evidence for the claim that relevant similarity is always similarity qua some set of descriptive properties, and hence that the property 'goodness' itself can never be a reason why P and another relevantly similar P_1 can both be good. One might say that the possibility that the judgement 'This is a good P' can have descriptive meaning establishes the possibility that it can have a reason. But this way of putting the matter needs to be qualified. The possibility that there can be a reason for the judgement is logically tied to this characteristic of 'good' that its use on a particular occasion has descriptive meaning, if the relevant principle is asserted or affirmed on that occasion in Hare's evaluative sense.[32]

It needs now to be explained how Hare defends this notion of 'evaluative' meaning and what role the primacy of this meaning plays in Hare's account of supervenience. Hare develops two lines of argument, first to show that value-judgements resemble ordinary imperatives and hence that value-judgements are not reducible to mere statements of fact, and secondly to show that logical rules apply to imperatives and hence that it is possible to construct argument forms which contain universal imperative sentences along the lines of the syllogism.[33]

On the first line of argument Hare establishes that imperative sentences and value-judgements have a common function. This is that anyone in using them intends someone else to do something. By contrast, indicative sentences are used merely for telling someone that some state of affairs obtains. The basic theme is that in their ordinary uses value-judgements have the force of imperatives and hence that it is possible to distinguish an evaluative or attitudinal element which is not reducible to mere

description. Hare formulates the resemblance between value-judgement and imperatives in terms of a common criterion of assent. Put briefly, to assent to either an imperative sentence or to a value-judgement is to be committed to accepting a prescription for some actual or conceivable choice or decision. It is, according to Hare, a necessary condition of assent to a value-judgement that if one assents to it, one is thereby committed to doing what the judgement prescribes. For instance, if one assents to the judgement that P is a good painting of its school to study, one is not merely affirming that a standard of goodness has been accepted; one is also primarily 'instructing' oneself to decide or choose on the basis of a universal principle which mentions those characteristics of P which make it good for the intended purpose. One might say that the point of applying a standard, besides one's intention of informing someone or making known to him the criteria for applying the word 'good' to P, is primarily to guide or direct his choice within the given class of comparison.[35] It will become apparent below that Hare treats the notion of subscribing to the universal principle under which the judgement is to be subsumed, as analogous to obeying a command.

On the second line of argument referred to above, Hare establishes that a singular imperative conclusion can be validly drawn from a set of premises which contain a universal imperative sentence as major premiss and an indicative sentence as minor premiss.[36] The possibility of deductive inference rests on the rule that no imperative conclusion can be validly drawn from a set of premises which do not contain at least one imperative. Hare maintains that if inferences based on this form of reasoning were not possible, the word 'all' would have no meaning in the logic of imperatives. The assumption underlying Hare's attempt to construct a logic of imperatives and hence of value-judgements is clearly his conviction that

purely 'factual' or 'descriptive' principles could not be 'action-guiding' and hence could not perform their function of regulating our conduct.

If I understand Hare correctly, he is suggesting that the supervenient character of 'good' is to be accounted for with reference to his notion of 'evaluative' meaning. Hare's reasons might be made clear by bringing out a contrast between the following three statements:[37]

(a) P is a good painting of its school to study, but P_1, though similar in all other respects, is not a good painting of this school to study.

(b) P is an example of the 'blue'-period in Picasso's development, but P_1, though similar in all other respects, is not an example of the 'blue'-period in Picasso's development.

(c) P is rectangular, but P_1, though similar in all other respects, is not rectangular.

It is self-contradictory to assert (a), but not self-contradictory to assert (b) or (c). It would be self-contradictory to assert (c), if 'bounded by four right-angles' is included in the specification of what counts as 'similar in all other respects'. But, there is an important difference between (a) and the reinterpreted (c) which concerns the reasons why it is self-contradictory to assert either. One could give a complete description of P and say that P is good, and, one could give the same description of P and say that P is not good, which is (1) to contradict oneself, though (2) each judgement by itself is internally consistent. But one could not give a complete description of P and say that P is rectangular, and, give the same description and say that P is not

rectangular, without allowing that at least one of these statements is internally inconsistent. That this must be so follows from the meaning of the word 'rectangular'.

Now, Hare thinks that 'good' logically behaves like the expression "'blue-period' in Picasso's development" as regards (2), but not as regards (1), and that 'good' logically behaves like 'rectangular' as regards (1), but not as regards (2). As I understand Hare's point, this is to be explained with reference to his view that the descriptive meaning of 'good' does not exhaust its meaning. One might make the same point by saying that the characteristic "'blue-period' in Picasso's development" can be a reason why P is a good painting of its school to study for anyone who accepts (in Hare's evaluative sense) this characteristic as a standard of goodness. But there is no necessary relation, in the logical sense of necessary, between the statement "P is an example of the 'blue'-period in Picasso's development" and the judgement 'P is a good painting of this school to study' such that the former by itself necessarily entails the latter, whereas there is as necessary relation between 'P is bounded by four right-angles' and 'P is rectangular' which holds solely in virtue of the meaning of the word 'rectangular'.

Hare makes it clear that if one assents to the value-judgement that P is a good painting of its school to study, it must be the case that one subscribes to a principle. The reason is that teaching or learning involves principles: one does not teach or learn an individual act but rather a particular kind of act in a particular kind of situation.[38] That one must subscribe to a principle of the sort just mentioned seems to be a matter of necessity. But what kind of necessity does Hare have in mind? On the second line of argument referred to above, the judgement that P is good for

the intended purpose follows necessarily, in the logical sense of necessary, from the joint assertion of the principle and a statement of the 'good-making' characteristics of P. The suggestion is that if one assents to the value-judgement, it must logically be the case that one also subscribes to a principle of the suggested form. This implies that Hare has logical necessity in mind.[39] Hare's explanation implicit in LM, is that supervenience is a logical property of only 'value'-words. Purely descriptive words are not supervenient. The claim that the word 'good' has a supervenient character is partly the claim that it is not a purely descriptive word. Hare's explanation suggests that he is justifying the necessity of appealing to a major universal imperative premiss whenever one calls something good with reference to the logic of value-language.[40] As I see it, the necessity of subscribing to a principle of this nature in the suggested form of reasoning rests on Hare's claim that evaluative meaning is logically primary, and that the supervenience of 'value'-words is to be accounted for with reference to this primacy. How does Hare account for this?

Hare justifies his view that the word 'good' is irreducibly evaluative by arguing that it has two kinds of meaning. It is possible to distinguish between the criteria for applying the word 'good' to various classes of things, and the commendatory function which is common to all value-ascriptions involving the word 'good'.[41] For instance, it is possible to know what the criteria are for applying 'good' to P, without one knowing that to call P a good painting to study is to guide or direct a choice within the given class of comparison. Conversely, it is possible to know this without one knowing what makes P good for the intended purpose.[42] Hare offers two reasons for the primacy of evaluative meaning.[43] The commendatory or 'action-guiding' force of 'good' remains constant whereas the criteria

for its application to various classes of things are all different. The commendatory force of 'good' can be used to change generally accepted standards of goodness. The change is effected by changing the descriptive meaning of 'good' as it is standardly employed within a given class of comparison. This is to change accepted notions of what makes a thing of its kind good and hence which members of its class are to be called good. In effect, this is to propose that future decisions about the value of things of this kind be made on the basis of a new principle which mentions the new set of 'good-making' characteristics. This implies that what descriptive meaning the word 'good' can carry in a given context is always subject to what principle is asserted in this context.

For Hare acceptance of the principle and its subsequent assertion on a particular occasion, are always instances of the evaluative or 'action-guiding' use of 'good' in the judgement which the principle subsumes. But, the principle itself cannot be interpreted as stating or expressing a necessary truth.[44] It has descriptive content some of which it shares with the minor factual premiss of the syllogism. There is, however, a necessary connection between the statement that P has the 'good-making' characteristic C, and the judgement that P is a good painting to study, which is explained with reference to the notion of subscribing to a principle. To say that P is good because P is C, is to say that it being C is necessarily related to it being good, where the 'because' indicates that C is a necessary condition of the goodness of P, and a necessary condition of C being a reason for the judgement that P is good. But, for C to be a reason for this judgement, a sufficient condition must also be met, and this is that the principle must be asserted in Hare's evaluative sense.

If this is correct, it follows that to ask for a reason why P is a good painting to study, is to ask for the principle under which the judgement that P is good, is to be subsumed. This, in effect, is to ask for the major imperative or evaluative premiss. Since the principle contains an imperative component, it is possible to derive from it, together with a statement of the 'good-making' characteristics of P, a reason for choosing P. One might say that the reason consists in the principle under which the judgement that P is good is to be subsumed, and that the judgement guides choices or decisions by drawing attention to the principle and hence to a reason for choosing P on all occasions on which the principle is asserted or affirmed. For Hare there can be no connection between saying that P has the 'good-making' characteristic C, and saying that C is a reason for choosing P, if the principle is not asserted in his evaluative sense. The peculiarity of Hare's position might be summed up by saying that a reason for a value-judgement is a reason for action.

It follows from what is involved in Hare's notion of subscribing to a principle that anyone who accepts that all P's which are C are good paintings to study, contradicts himself if he says that P is good but another relevantly similar P_1 is not good, since this is to affirm and deny that he asserts the principle in this context.[45] But no contradiction arises if different persons assert different principles in this same context. The descriptive meaning of the judgement 'This is a good P to study' will be fixed by whoever accepts and asserts this principle, but this same judgement cannot carry different descriptive content, i.e. it is not possible to exhibit its descriptive meaning as a disjunction of different sets of descriptive properties. The obvious point is that for Hare acceptance of a principle determines relevant similarity.

Hare's account of supervenience implies a particular view of what disagreement about the value of a thing consists in. Since the statement that P is C does not, by itself, entail the judgement that P is a good painting to study, it is possible for two people to agree that P is C, yet disagree without logical or linguistic error about the value of P. One might say they disagree about <u>which</u> paintings of this school are the good ones to study, or about what <u>makes</u> a painting of this school a good one for the intended purpose, and that the disagreement is to be explained with reference to different standards. One might say that the property 'goodness' supervenes on whatever set of 'good-making' characteristics is the reason for a judgement that some paintings of this school are good ones to study. Put differently, depending on what principle is asserted, P and another relevantly similar P_1 are either <u>both</u> good, or <u>both</u> bad. One does not offend against any logical rule governing the use of 'good' in this context, should one fail to commend them.[46]

The reasons for Hare's rejection of alternative accounts of supervenience should by now be obvious. The basic theme is that alternative accounts do not show what is peculiar about value-language.[47] Naturalists accept that the relation between the property 'goodness' and the descriptive properties of P on which it supervenes is a necessary relation which holds in virtue of the meaning of the word 'good'. To say that P is good because P is C is to say that it being C is a necessary and sufficient condition of the goodness of P. It follows that any P which has the defining characteristic C must of necessity be good, and that being a good P necessarily implies that it has this defining characteristic. If the supervenience relation is understood as a relation of entailment due to equivalence of meaning, it is contradictory to claim that P and another relevantly similar P_1 can differ only in being good. For the claim that they differ in the latter respect

implies that they differ in other respects too, viz. in being C. On this account it is a conceptual truth that value-judgements have reasons. But this account reduces the supervening property 'goodness' to a mere descriptive property, which is to say that evaluation or commendation simply is description. Hence naturalists fail to account for the commendatory or 'action-guiding' force of 'good'.[48] In LM Hare makes this point by noting that if 'A good P' means the same as 'A P which is C', then it would be impossible to use the sentence 'A P which is C is good' to commend P's which are C. For this sentence would be analytic and equivalent to be sentence 'A P which is C is C'.[49]

The relation of dependence of the property 'goodness' on other kinds of properties is brought out by saying, as Hare does, that whenever something is called good, there must be something about it which is good. Hare accepts that this dependence relation provides an explanation of why there can be reasons for the goodness of things. What is puzzling or peculiar about saying that the goodness of P depends on P having the property C is, as I have understood Hare's account, a matter of explaining that P is called good by appeal to the relevant standard or principle, and hence for the reason that P is C. In other words, it is a matter of explaining why value-judgements employing the word 'good' have reasons and how their having them is to be understood.

What is the relation between (SP) and (UP)? I have tried to show above that on Hare's account of supervenience

(a) the property 'goodness' itself can never be a reason why P is called good;

(b) some set of descriptive properties of P is necessarily relevant to the goodness of P;

(c) a principle always determines relevant similarity (for all cases which may be subsumed under the principle).

There is a logically necessary connection between 'P is C' and 'P is good' which holds if one subscribes to the principle that all P's which are C are good paintings to study. C may be a reason for the goodness of P for anyone who subscribes to this principle. If I understand Hare correctly, there are logical limits to what reasons may be advanced for the goodness of P which are set by the three points just outlined, but which principle one subscribes to within the given class of comparison is a matter of choice or decision. P being C is always a necessary condition, but never by itself also a sufficient condition for the goodness of P. The sufficient condition is supplied by Hare's notion of subscription to the relevant principle. It follows that it is logically possible for any set or conjunction of properties to ground the claim that P is good (for the intended purpose). This is just to say that by itself (SP) does not determine what kinds of properties of a thing or act are relevant. P may not be good for anyone who subscribes to a different principle.

How does (SP) help Hare to establish (UP)? If (UP) requires mere consistency, then the three points outlined above establish that (UP) and (SP) are equivalent, i.e. (SP) entails (UP) and (UP) entails (SP). But this does not explain the assumption of a universal principle which serves as major imperative premiss in Hare's account of the supervenience of 'ought' and 'good'. There is nothing in (SP) itself to suggest that the morally relevant properties of some thing or act are only those which can be stated

in universal terms. To what extent, then, does (SP) lend support to Hare's claim that moral judgements depend on proper universal principles?

In defending his account of the supervenient character of 'good' and 'ought', Hare relies on the notion of subscribing to a principle. The significant point which follows from his account of what is involved in subscribing to a principle is that the subscriber has a reason to act as his principle enjoins him to act. But, there is no reason to suppose that he contradicts himself if he allows that anyone else in relevantly similar circumstances ought not to act as he does. It seems that (SP) can be distinguished from (UP) in that a reason for doing something which satisfies the constraints of (SP) can refer to a particular individual. It seems that Hare needs to find some justification for limiting what kinds of properties of a thing or act can be morally relevant properties. I think naturalism sets this limit by making it a necessary and sufficient condition of the goodness of a thing or act that it must have a certain set of descriptive properties.[50] It is not difficult to see that this supplies one possible foundation for universalizability, though Hare rejects this on the grounds that it makes the universalizability of moral judgements trivially true, i.e. they are universalizable only in the sense in which descriptive statements are universalizable.[51]

Hare writes as if it is just a feature of the logic of 'ought' that its use in some moral judgement presupposes a reference to a universal principle, and as if this requirement of moral uses of 'ought' is supportable by (SP). I find this difficult to harmonize with the view that ordinary value-judgements need not be universalizable in the sense required by (UP), and with the view that assent to an ordinary value-judgement commits one to a principle of action which has the force of a universal imperative, but

which is not universal in scope, as a moral principle (on his view) clearly is. If (SP) is equivalent to (UP), all value-judgements, not just moral ones, must be universalizable. But, if this is so, Hare cannot consistently maintain that the difference between a moral and a value judgement just is a difference in the scope of a universal imperative and a universal 'ought'-judgement.[52]

I think I am right in claiming that Hare does not succeed in defending his distinction between the kinds of reasons which it is logically appropriate to offer for moral and value-judgements. The problem just is that (SP) is not sufficient to support the claim that moral judgements must be supported by u-type principles. Hare has not shown why it is logically inappropriate to offer an e-type reason for a moral judgement. I sum up what I take to be Hare's problem thus: the problem of relevant similarity in the moral case just is the problem of what kinds of properties are morally relevant. This requires some justification for limiting what kinds of propertiesare to count as morally relevant, but Hare's defence of (SP) supplies no direct support for a distinction which he operates within LM and U. I think Hare needs to make explicit a criterion of relevant similarity consistent with the u-type requirements of (UP).

In Freedom and reason Hare accepts the view that all 'ought'-judgements are universalizable in the sense that u-type reasons can be given for them. Hare retains the claim that moral judgements must be supported by u-type principles, but abandons the view that the claim is analytic i.e. true in virtue of the meaning of the word 'moral', particularly as he attempts to phrase it in terms of values and not in terms of the word 'moral'. In Chapter 2 I shall examine the alternative justification for this claim in

order to establish whether Hare succeeds in advancing an alternative basis for the distinction between e-type and u-type reasons.

Section 2. Prescriptivity

The prescriptivity thesis might be stated thus:

(P) 'X ought to be done', if used evaluatively, entails 'Let x be done'. If an agent assents to the moral judgement 'x ought to be done', he logically cannot dissent from acting on the prescription 'Let x be done' (if now is the time to act, and if it is within his power, physically and psychologically, to do what the prescription enjoins him to do)[53].

In LM Hare's defence of (P) rests on the claim that moral and other evaluative language, including imperatives, are species of the genus 'prescriptive language' and so distinguishable from statements of fact. Hare presents the difference between statements and commands in terms of different criteria of sincere assent,[54] and develops an analogy between value-judgements and commands which suggest the following parallel. Sincere assent to a command implies that the commanded is sincere in his assent if he does or resolves to do what the speaker commands him to do. Analogously, sincere assent to a value-judgement implies that the addressee is sincere in his assent if he does or resolves to do what the speaker tells him to do. Hare says:

> 'It is a tautology to say that we cannot sincerely assent to a second-person command addressed to ourselves, and <u>at the same time</u> not perform it, if now is the occasion for performing it and it is in our (physical and psychological) power to do so.'[55]

Hare makes it a necessary condition of sincere assent to a value-judgement that if one assents, one thereby commits oneself to the prescribed course of action. The analogy suggests that value-judgements resemble imperatives in as much as they require, as a necessary condition of their satisfaction, that if one sincerely assents, one must do what the judgement says one should do.[56]

But the analogy is to some extent misleading. Hare's thesis is not that value-judgements are reducible to imperatives, but rather that they 'entail' prescriptions for action, if they are used evaluatively. Hare carries this point by definition of the word 'evaluative'.[57] Hare says:

> 'I propose to say that the test, whether someone is using the judgement "I ought to do x" as a value-judgement or not is, "Does he or does he not recognize that if he assents to the judgement, he must also assent to the command 'Let me do x'"?'[58]

Hare recognizes that his definition does not, as he puts it, prove anything substantial about how we use language,[59] which is to say that he assumes rather than shows that there are criteria which distinguish the genuine value-judgement from statements of sociological and psychological fact, i.e. statements like

> (1) '"X is required in order to conform to the standard which people generally accept" (statement of sociological fact);
>
> (2) "I have a feeling that I ought to do x" (statement of psychological fact).'[60]

The crucial point is that statements like these may be construed as having 'action-guiding' force, though they are merely 'descriptive'. This, as Hare notes, is a consequence of the fact that standards may become so thoroughly accepted that they have the force of intuition, and hence are treated as matters of fact.[61] So, the question arises: How might we tell when we are to count a judgement as a genuine value-judgement? It is this difficulty which Hare proposes to overcome by means of definition. I do not wish to dispute Hare's definition of the word 'evaluative'. I wish merely to note that if this definition be granted, the prescriptivity thesis becomes analytic. It becomes analytically true to say that the moral judgement 'I ought to do x' entails the imperative 'Do x', and that we cannot assent to the former and dissent from the latter unless we have, as Hare puts it, 'misunderstood the meaning of one or the other.'[62]

Hare also recognizes that his definition, combined with the criteria of sincere assent to a moral judgement, produces an odd result. It becomes analytic to say

> '... that everyone always does what he thinks he ought to (in the evaluative sense).'[63]

This, as Hare recognizes, is not how we use the word 'think'. The criteria for saying 'He thinks he ought' are 'exceedingly elastic'.[64] If omitting to do something we think we ought to do is accompanied by feelings of guilt, we say that we have not acted as we think we ought to. Hare therefore proposes to qualify his criteria of sincere assent by allowing that there are 'degrees of sincere assent',[65] not all of which involve actually obeying the command 'Do x'.

I noted above that it is a necessary condition of sincere assent to a moral judgement that if one assents, one is thereby logically committed to a particular course of action. Hare recognizes that there is a psychological phenomenon called 'weakness of will'.[66] But, barring cases of this phenomenon, Hare's position implies that it is contradictory to say that on some particular occasion one did not act as one sincerely believes one ought to have acted. One way of explaining why one may on occasion act 'contrary' to a principle touches on the question of what shall count as a relevant description of what one did on the occasion where one is said to be acting contrary to principle.[67] This is a matter of recognizing whether or not a given description fits the act, and hence whether or not it is subsumable under the principle. One might then say post eventum that one acted contrary to principle, though at the time one sincerely believed that one were acting as one ought to.

Another possible explanation touches on the problem of deciding in cases of conflict, which of one's principles one is willing to override and which one is not. A situation might be such that a person is faced with a conflict between two principles, each of which is designed to cope with circumstances of a particular kind, but which come into conflict because of the peculiarities of a particular kind of case.[68] For instance, do we say that the doctor, who decides to let a terminally ill patient die rather than to keep him alive at great cost in pain to the patient, has acted 'contrary' to his principle to save lives? Do we say, if he takes the opposite line and tries to keep the patient alive for as long as possible, that he acts 'contrary' to his principle to prevent pain? I think this is the kind of case in which a person might discover that his principles are inadequately qualified, i.e. he has not yet settled the prior question of which principle he is willing to qualify or which one he is willing to

treat as overriding, and what the justification for this could be. One might then say, not that he is insincere in holding to either principle, but rather that he is in doubt about what he ought to do in this particular kind of case.

I do not think Hare is likely to quarrel with the idea that considerations of one's being in error about the facts of a case or of one's being in doubt about the rightness of a particular line of action in a particular kind of case, do not define cases of insincere assent. A question about one's sincerity in holding to a principle is raised, if one recognizes that in some cases one is acting contrary to principle and hence that there is something one needs to account for,[69] though this is not always simply a matter of actual insincere assent. There is considerable tension between Hare's definition of a value-judgement as entailing at least one imperative and the criterion of sincere assent. The worrying point about the definition is that it seems to be stipulative. It helps to tidy up things in an area where dispute is likely to arise about when a person can be said to be sincere or insincere in his assent.[70]

Section 3. Decisions of principle

One conclusion that can be drawn from what I said in section 1 about the necessity of subscribing to a universal principle in the form of reasoning Hare defends, is that he pushes the question of how it is possible for moral judgements to have reasons back to the question of how the adoption of a universal principle is to be justified.[71] The position Hare defends implies that the principle which supports a moral 'ought'-judgement cannot be justified by appeal to fact (as the naturalists thought), or by appeal to self-evidently true propositions. How, according to Hare, do we justify a decision of principle? What principle one accepts for a particular kind

of case is a matter of individual decision, but one cannot be said coherently to be making a decision about what one ought to do in a particular kind of situation without one knowing what it is one should be doing, if one decided to do one or other of the things one could be doing in that kind of situation. The suggestion is that one's decision, to be right, must ultimately rest on a consideration of the consequences of adopting a particular line of action and hence that the justification of the principle ultimately rests on a decision to accept one set of envisaged consequences rather than another set. Hare says:

> 'The truth is that, if asked to justify as completely as possible any decision, we have to bring in both effects - to give content to the decision - and principles, and the effects in general of observing those principles, and so on, until we have satisfied our inquirer. Thus a complete justification of a decision would consist of a complete account of its effects, together with a complete account of the principles which it observed, and the effects of observing those principles - for ... it is the effects (what obeying them in fact consists in) which give content to the principles too. Thus, if pressed to justify a decision completely, we have to give a complete specification of the way of life of which it is a part'.[72]

If I understand Hare correctly, he is suggesting that we adopt the following principle as a guideline for determining the rightness of an action:

(PC): An act is right if the (probable) consequences of its performance in situations of a given kind are as good or not worse than the

(probable) consequences of performing an alternative act in situations of the given kind.

Two features of (PC) are worth noting. The first is that (PC) leaves open the possibility that an act may not be right, since the determination of its rightness depends on two things: (1) the extent of one's knowledge of the available facts of a situation, which must include an account of the (likely) effects of the performance of the possible alternative acts on the situation; (2) the reliability of one's expectations that a preferred outcome is likely to obtain, which is relative to the knowledge one has of the consequences of adopting a particular line of action at the time one makes one's decision. I think it can be said that a decision-maker does wrong in a given situation if and only if he knows, at the time he makes his decision, that he ought to do otherwise. But it does not follow that he is blameworthy should he discover that he chose wrongly, for (PC) allows that there is a sense in which what is right at the time he decides, may turn out to be wrong retrospectively.

The second feature of (PC) worth noting is that it defines a consequentialist perspective. Hare is not merely saying that consequences are morally relevant, but rather that only consequences and nothing else are relevant to a consideration of the moral rightness of an act. It seems that the rightness of an act rests on the utility of its consequences, though Hare does not explicitly say so. I take this to be the major import of (PC), though Hare advances no explicit argument which shows how (PC) may be linked with the principle of utility. The second feature raises two questions. What commits one to a consideration of only consequences in determining the rightness of an act? On what basis does one decide that a particular set of consequences is preferable to another? Hare provides no

explicit answer to the first question. Regarding the second question, the suggestion seems to be that one chooses that set of consequences which is best overall for everyone who may be affected by one's decision. This, it seems, is a direct consequence of the thesis of universalizability when combined with the prescriptivity thesis. In virtue of universalizability, one cannot be said intelligibly to judge that one ought to do a certain thing on a particular occasion while refusing to commit oneself to the view that anyone else in relevantly similar circumstances ought to do the same thing. An individual's decision, if it is to have moral import, must be universalizable in this sense. In virtue of prescriptivity, it cannot be said that one sincerely subscribes to a principle unless one acts in accordance with it, if now is the time to act on it, and if it is within one's power to do so.

Now, what does this entail for moral reasoning? One might say that anyone who accepts a principle for a particular kind of situation is committed to the view that anyone else who decides and acts differently in a relevantly similar situation, must be wrong. For, in subscribing to a principle, one defines a commitment to what is right in all cases which may be subsumed under the principle. For instance, the doctor who decides to let a terminally ill patient die rather than to keep him alive a little longer at great cost in pain to the patient, must hold that his colleague is wrong in taking the opposite line. In universalizing the reason for his decision, he implies that the reasons for deciding one way or the other are not of equal weight, i.e. the alternative courses of action cannot all be right.

Hare allows for the play of an element of choice, which implies that a moral principle may have whatever content a prescriber chooses to give it, and that criteria of value are open to choice. There is, however, an

independent check on what one may choose to prescribe universally. This is, to put it bluntly, that one has to live with the consequences of one's decisions. Hare says that the decision-maker

> '... has to decide whether to accept that way of life or not; if he accepts it, then we can proceed to justify the decisions that are based upon it; if he does not accept it, then let him accept some other, and try to live by it. The sting is in the last clause'.[73]

It seems fair enough to say that in LM Hare is advancing the thesis that his notion of subscribing to a principle is the essential notion in morals.[74] But it is not clear where this is supposed to lead to. There is a suggestion that what is right in circumstances of a given kind is, all other things being equal, what particular agents decide to do, which implies that decisions about the moral rightness of acts of particular kinds are always relative to agents. I think there is a case for saying that in LM Hare is primarily concerned with individual welfare functions rather than with developing a method of reasoning which shows how a transition is to be made to social welfare functions.

But, if what I said in Section 1 is in order, it is not difficult to see that ultimate decisions of the kind suggested here admit a form of subjectivism.[75] If I am right, anyone who says that anyone else who decides and acts differently in a relevantly similar situation does wrong, cannot rest his case on an appeal to the requirement of consistency. (UP) requires some notion of relevant similarity consistent with the u-type requirements which Hare lays down for the justification of moral 'ought'-judgements. If I am right a question arises on what grounds one might justify one's ascription of wrongness to the alternative decision or act.

Conclusion

I conclude this chapter with a brief indication of the direction in which Hare develops his views in Freedom and reason. Hare attempts to place his defence of (UP) on a different footing, though he holds to the view that moral 'ought'-judgements require u-type principles for their justification. He develops the notion of a universal descriptive meaning-rule in an effort to provide an alternative foundation for universalizability. I shall examine this view in Chapter 2 in order to establish whether it is sufficient to sustain the claim that moral judgements are u-type valuations. Two important issues in Freedom and reason concern the development of a method of reasoning which Hare claims is dictated by the logic of moral language, universalizability and prescriptivity, i.e. the establishment of a theory of moral reasoning which he calls universal prescriptivism, and the claim that his method of reasoning commits us to accepting substantive moral judgements which are consistent with a form of utilitarianism. The argument for this centres on the reversibility test and a set of rationality constraints, and consists essentially in the claim that a form of utilitarianism may be derived from universalizability and prescriptivity taken together with these things.

Notes

1. In U Hare does not offer an explicit formulation of the principle of universalizability. My formulation follows Don Locke's interpretation of Hare's intention. As Locke interprets Hare, in accepting a moral 'ought'-judgement, we are 'legislating for everyone', and are committed to the view that 'anyone who decides differently has decided wrongly'. See Locke (1968): The trivializability of universalizability. The philosophical review, vol. LXXVII, p. 30.

2. The clause 'if used evaluatively' is crucial. Hare says: 'I do not wish to claim that all 'ought'-sentences entail imperatives, but only that they do so when they are being used evaluatively.' See Hare (1952): The language of morals (reprinted 1982) p. 164. Oxford University Press. All subsequent references to LM are to the 1982 reprinted edition.

3. Hare says that 'giving a reason for any action involves reference (explicit or implicit) to a rule, maxim or principle. This appears to me to analytically true'. See Hare (1955): Universalizability, reprinted in Essays on the moral concepts (1973) University of California press. p. 15. All subsequent references to U are to the reprinted version of 1973. Hare expresses a similar view in LM. He says: '... it would be logically illegitimate to give a type C prescription [i.e. one involving the use of the word 'ought'] while denying that there was any principle on which it depended.' See Hare (1982) p. 157.

4. Hare is quoting Gellner. See Hare (1973) p. 13. The terms 'E-type' and 'U-type' are Gellner's. The thesis that moral 'ought'-judgements are 'U-type' valuations is true 'in virtue of the meaning of the word "moral"'. See Hare (1973) p. 16.

5. This way of putting the matter is due to C.E. Caton. See Caton (1963): In what sense and why 'ought'-judgements are universalizable. The philosophical quarterly, vol. 13, p. 49.

6. Here again I follow Locke. See Locke (1968) p. 32. In U Hare says: '... a reason cannot be a reason on just this occasion, and not on other similar occasions ...' unless, of course, there are relevant differences between the cases. See Hare (1973) p. 15. I take it that Hare means that the reason 'cannot be a reason on just this occasion' for A.

7. Hare (1973) p. 16.

8. Hare (1973) p. 15.

9. Hare (1973) p. 15.

10. Hare (1973) p. 18. Hare says: 'Universality, in the sense in which U-type maxims are universal, is not a matter of degree; to suppose that it is, is to confuse the term "universal" (as opposed to "singular") with the term "general" (as opposed to "specific").'

11. Hare (1973) p. 23-24.

12. Hare (1973) p. 27.

13. In LM Hare maintains that to call some x good and another exactly or relevantly similar y not good, is inconsistent (i.e. contradictory), because '... to do so would be to attempt to teach or advocate two mutually inconsistent principles in the same breath.' See Hare (1982) p. 159. I think Hare understands (UP) as requiring consistency. Hare says: '... if I have said, "That is a good motor-car" and someone asks "Why? What is good about it?" and I reply "Its high speed combined with its stability on the road", I indicate that I call it good in virtue of its having these properties or virtues. Now to do this is eo ipso to say something about other motor-cars which have these properties. If any motor-car whatever had these properties, I should have, if I were not to be inconsistent, to agree that it was, pro facto, a good motor-car ...'

14. These formulations are suggested by Andrew Oldenquist. See Oldenquist (1968) Universalizability and the advantages of nondescriptivism. The journal of philosophy, vol. LXV, p. 58-59.

15. In LM Hare says: 'Whenever we commend, we have in mind something about the object commended which is the reason for our commendation. It therefore always makes sense, after someone has said "That is a good motor-car", to ask "What is good about it?" or "Why do you call it good?" or "What features of it are you commending?"' See Hare (1982) p. 130.

16. This formulation is suggested by Oldenquist. See Oldenquist (1968) p. 57.

17. Hare (1982) p. 80-85.

18. This formulation is suggested by Oldenquist. See Oldenquist (1968) p. 58.

19. Hare says: '... there is some one characteristic or group of characteristics of the two pictures on which the characteristic "good" is logically dependent, so that, of course, one cannot be good and the other not, unless these characteristics vary too.' See Hare (1982) p. 81.

20. Jaegwon Kim argues that Hare's concept of supervenience has the force of 'weak supervenience'. Kim says: '... weak supervenience ... only requires that any two things having the same natural properties must be either both good or both not good.' He says further: 'Under weak supervenience there would be an inconsistency in one's commending an object (saying that it is good) but failing to commend another that is, or believed to be, exactly like it in all descriptive details; however, there is nothing inconsistent, or incoherent, in failing to commend either while acknowledging the same descriptive properties of the two objects. Weak supervenience, therefore, gives us the much discussed Principle of Universalizability of ethical judgements understood as a consistency requirement.' See Kim (1984) Concepts of supervenience. Philosophy and phenomenological research, vol. XLV, p. 161.

21. Hare says: '... the implication of the judgement "That is a good motor-car" does not extend merely to motor-cars exactly like that one. If this were so, the implication would be for practical purposes useless; for nothing is exactly like anything else. It extends to every motor-car that is like that one in the relevant particulars; and the relevant particulars are its virtues - those of its characteristics for which I was commending it, or which I was calling good about it.' See Hare (1982) p. 129-130.

22. This formulation is suggested by Oldenquist. See Oldenquist (1968) p. 59.

23. Hare says: '... since, as we have already remarked, "good" is a "supervenient" or "consequential" epithet, one may always legitimately be asked when one has called something a good something, "What is good about it?". Now to answer this question is to give the properties in virtue of which we call it good. Thus, if I have said, "That is a good motor-car" and someone asks "Why? What is good about it?" and I reply "The high speed combined with its stability on the road", I indicate that I call it good in virtue of its having these properties or virtues'. See Hare (1982) p. 131. The upshot is that my answer can never be '... just its goodness and nothing else'. See Hare (1982) p. 130.

24. Hare (1973) p. 15-16.

25. Hare says: '... a natural response to the discovery that "good" behaves as it does, is to suspect that there is a set of characteristics which together entail a thing being good, and to set out to discover what these characteristics are. This is the genesis of that group of ethical theories ... called "naturalist" ... I shall argue ... that what is wrong with naturalist theories is that they leave out the prescriptive or commendatory element in value-judgements, by seeking to make them derivable from statements of fact. If I am right in this opinion, my own theory, which preserves this element, is not naturalist.' See Hare (1982) p. 81-82.

26. In formulating Hare's position I consulted his article on Supervenience. This article was written in 1950 and first published in Proceedings of Aristotelian Society, Supplementary volume 58, 1984. It is reprinted in Hare (1989): Essays in ethical theory. Claredon Press. Oxford. All subsequent references to this article are to the version printed in 1989. The position Hare defends is, according to him, implicit in the 1950 article and in LM. Briefly, Hare argues thus: 'A typical practical syllogism has one universal premiss, one singular premiss subsuming a particular case under the universal premiss ... and a singular conclusion ... Let us call the two premisses and the conclusion of one of these inferences "p", "q", and "r". In full the inference will be of the form

p: For all x, if Gx then Fx
q: Ga
So r: Fa

We have to distinguish two theses ... It is the second, not the first that supervenience requires. The first is:

(1) Necessarily if q then r.'

Hare then notes that '... if we took as the universal premiss not "p" but "necessarily p", then (1) would follow directly'. But Hare argues that '... the universal premiss, whose existence is necessarily posited when we ascribe a supervenient property, does not itself have to be a necessary truth - at least not in the sense that we are constrained logically or in any other way to subscribe to it.' He then goes on to say that the second thesis, which also states a necessary truth, may be formulated thus:

'(2) Necessarily, if r, then there is a valid inference of the "p, q, so r" form, the two premisses of which hold.'

Hare notes that '... in saying a premiss holds is to subscribe to it'. In this regard see also Hare (1982) p. 191. 'This', claims Hare, 'is what is involved in the claim that "r" ascribes a supervenient property; it is not, of course, universally true whatever "r" is, but is what distinguishes "r's" which ascribe supervenient properties from those which do not.' The upshot is that the '... universal premiss should not be analytically true, or true in virtue of the meanings of words.' It should be noted that the '... "necessarily" at the beginning of (2) means conceptual or logical necessity, simply because supervenience is ... a logical property'. See Hare (1989) p. 69-71. See also Oldenquist (1968) p. 78.

27. Hare (1982) p. 111-126.

28. Hare says: '... the instruction [cannot] be done successfully unless the instructor is consistent in his teaching.' He also says: '... a standard is by definition consistent.' See Hare (1982) p. 132. See also note 13 above.

29. Hare says: '... the relation of the expression "good motor-car" to the criteria for its application is very like the relation of a descriptive expression to its defining characteristics, and this likeness finds an echo in our language when we ask "What do you mean, good?", and get the answer "I mean it'll do 80 and never breaks down". In view of this undoubted fact of usage, I deem it best to adopt the term "descriptive meaning".' See Hare (1982) p. 118.

30. To teach the descriptive meaning of 'A good P' is tantamount to "... explaining or conveying or setting forth the standard of goodness in [P's]'. Hare also says: 'To know the descriptive meaning is to know by what standards the speaker is judging.' See Hare (1982) p. 114 and 146.

31. Hare (1982) p. 122.

32. See note 26 above. As I judge, if it is necessarily true that a reason can always be given for a value - or moral judgement, as Hare's account of supervenience implies, the reason must be simply the universalized form of the judgement, and must be mentioned in the universal principle which Hare posits as the major premiss of the practical syllogism. Being prescriptive, the universal principle has 'action-guiding force'. To subscribe or assent to the principle is to give it its 'action-guiding' force, and this determines that if C is my reason for commending P, I am committed, on pain of self-contradiction, to commending an exactly or relevantly similar P_1, assuming that I have not changed my mind.

33. Hare (1982) p. 17-31.

34. Hare (1982), p. 18-20.

35. Hare (1982) p. 144.

36. Hare (1982) p. 25-28.

37. The examples and the analysis which follow are adapted from Hare (1989) p. 68.

38. Hare (1982) p. 60.

39. See again note 26 above. Hare says that in LM he employed the concept of supervenience in '... an attempt to find clear logical criteria for distinguishing between evaluative and descriptive words. I thought at that time that this characteristic of value-words could be used for that purpose.' But he concedes that this was wrong. 'For supervenience is a feature, not just of evaluative words, properties, or judgements, but of the wider class of judgements which have to have, at least in some minimal sense, reasons or grounds or explanations.' See Hare (1989) p. 66.

40. See note 39 above.

41. Hare (1982) p. 97, 102 and 106.

42. Hare (1982) p. 108.

43. Hare (1982) p. 118-120.

44. See again note 26 above.

45. See again note 13 above.

46. See again note 20 above.

47. See again note 25 above, and Oldenquist (1968) p. 63-64.

48. See again note 39 above.

49. Hare (1982) p. 85.

50. Oldenquist (1968) p. 64.

51. Hare (1989) p. 71.

52. Hare (1982) p. 175-178.

53. Hare (1982) p. 19-20.

54. Hare (1982) p. 20.

55. Hare (1982) p. 20.

56. See Gardiner (1955): On assenting to a moral principle. Proceedings of the Aristotelian Society, vol. 55, p. 28-29.

57. Hare (1982) p. 164.

58. Hare (1982) p. 168-169.

59. Hare (1982) p. 169.

60. Hare (1982) p. 167.

61. Hare (1982) p. 164-165.

62. Hare (1982) p. 172.

63. Hare (1982) p. 169.

64. Hare (1982) p. 169.

65. Hare (1982) p. 169-170.

66. Hare (1982) p. 169.

67. See Gardiner (1955) p. 29 and 32.

68. I borrow this point and the example from Hare (1981) Moral thinking. Clarendon Press. Oxford. p. 175-178. See also Gardiner (1955) p. 39-40.

69. See Gardiner (1955) p. 43.

70. Gardiner puts the point thus: '..."assenting to a moral principle P" excludes "not doing as P requires" in the absence of a wide and variegated range of circumstances and factors', which are often the things that count. See Gardiner (1955) p. 44.

71. See Oldenquist (1968) p. 78.

72. Hare (1982) p. 68-69.

73. Hare (1982) p. 69.

74. I borrow this point from R.B. Braithwaite. (1954). Critical notice of Language of morals. Mind, vol. 63, p. 252.

75. See Braithwaite (1954) p. 250.

CHAPTER 2

An analysis and critique of 'Freedom and reason' (FR)

Introduction

Hare's argument for utilitarianism turns crucially on premisses which commit an agent to settling a conflict of interests between distinct individuals as if it were a conflict of interests in his own person. This strategy is of particular importance in Hare's attempt to commit an agent to 'balancing' and maximally satisfying desires as a form of utilitarianism requires. At bottom Hare's strategy depends on what he thinks the thesis of universalizability can do. In this a demand for role-reversal, on a particular interpretation of what this demand requires, and a demand for imaginative identification with the desires of concerned parties, are the critical moves Hare makes to constrain an agent to accept utilitarian conclusions.

In this chapter my object is to examine the claim that the moves Hare makes in the argument for utilitarianism are all justifiable with reference to the logical features of moral language. In Section 1 I offer an analysis of Hare's theory of moral reasoning in an attempt to show how universalizability and prescriptivity generate a method of moral reasoning, and how Hare attempts to defend the moves he needs to make in order to defend the argument for utilitarianism. I shall argue that there are various gaps in this defence. In Section 2 I offer an analysis of Hare's defence of the thesis of universalizability, and a critique of the claim that moral judgements are universalizable prescriptions. I attempt to arrive at a sense in which this claim should be understood. I try to show that the claims Hare makes for universalizability are of central importance

in establishing the strategy referred to above, and why I think Hare is unlikely to succeed in generating premisses from this logical feature of moral language which might commit an agent to accepting a form of utilitarianism. In Section 3 I isolate and discuss two problems in Hare's strategy which I think undermine the argument for utilitarianism.

Section 1: Hare's theory of moral reasoning

My object in this section is to present an outline of Hare's account of moral reasoning as it is presented in Freedom and reason. The account of moral reasoning which Hare presents in FR, contains a claim worth noting at the outset. This may be put thus:

1. The logical features of moral language (i.e. universalizability and prescriptivity), and
2. claims about people's inclinations and desires, when conjoined in
3. a method of reasoning required by 1,

compels a prescriber to give the inclinations and desires of all recipients the same consideration as he gives his own. I mention this claim at the outset because it performs a pivotal role in Hare's attempt to attain a specific objective. This is to discover a moral justification for a course of action a prescriber may want to adopt, but which will affect other persons in various ways. The universalizability and prescriptivity theses may be formulated as follows.

(UP) If some agent, A, accepts the moral judgement that X ought to be done in S, he is thereby, on pain of self-contradiction, committed to the view that X ought to be done in all situations like S, where 'like S' is similarity in respect of the non-moral but morally relevant features of S.[1]

(P) 'X ought to be done', if used evaluatively, entails 'Let x be done'. If an agent assents to the moral judgement 'X ought to be done', he cannot dissent from acting on the prescription 'Let x be done', if he is sincere in his assent, and if he is able (physically and psychologically) to act on it.[2]

I shall have occasion later (Section 2) to revise my formulation of (UP). For my present purposes, the stated version is quite adequate. To simplify matters, I shall separate the argument Hare presents in a case of bi-lateral conflict from the argument he presents in the more complex case of multi-lateral conflict. The former proceeds in two stages. Hare first presents a case in which the parties actually stand to each other in exactly the same relation, and in which their inclinations do not differ.[3] This serves as a preliminary to his argument for the case in which they are not, in fact, in the same position.[4] I shall refer to the former as the 'actual', and the latter the 'hypothetical' bi-lateral cases. Hare then presents his argument for the multi-lateral case of conflict as a generalization of the method he outlines in the 'hypothetical' bi-lateral case.[5]

1.1 The 'actual' bi-lateral argument

I begin with the 'actual' bi-lateral case in which Hare makes the simplifying assumption that the parties to the conflict share at least one inclination. This, in the creditor-debtor example below, is the desire to avoid imprisonment in case of non-payment of debt.[6] Hare's argument may be reconstructed along the following lines. Suppose A is both creditor to B and debtor to C.

1. A wants to imprison B, because B cannot pay. This being so, he accepts - initially - the prescription 'Let me imprison B, if B cannot pay.'[7]

2. Suppose now that A wonders whether imprisoning B would be the morally correct thing to do. If he accepts the judgement 'I ought to imprison my debtor, if he cannot pay', as a moral judgement, it must be the case that he also accepts the universalized form of this judgement - 'Anyone in my position ought to imprison his debtor, if he cannot pay'. (This is justified by appeal to universalizability.)

3. However, it follows from the demand for universalization that A must accept that his creditor ought to imprison him, if he cannot pay. This will commit him to accepting the prescription 'Let C imprison me, if I cannot pay.' (This is justified by appeal to prescriptivity.)

4. But, the prescription he has to accept in Step 3 is inconsistent with what he has to prescribe should happen to himself regarding his problem with C, for A is <u>in fact</u> in the same position as B, i.e. their situations are identical in all relevant respects - in respect of the relation between debtor and creditor, and particularly in respect of the desires which characterize this relation.[8]

5. Now, in order to be consistent, A must reject the prescription that C should imprison him, if he cannot pay. If he rejects it - as he must, given that he strongly desires to avoid imprisonment and that this desire is stronger than his desire to exact payment - he must also reject the universalized form of the judgement which entails this prescription.

6. Hence, if A is to be consistent, he cannot accept that the judgement he contemplates endorsing in 2, can be a moral judgement.

As I judge, the stipulation that A actually stands to C in exactly the same relation as B stands to A, plays a key role, in conjunction with the appeal to universalizability and prescriptivity, in yielding Hare's conclusion. The significant point about this stipulation is that A's desire not to be imprisoned in case of non-payment of debt must be read as the desire he actually has for the envisaged situation in which C wants to exact the penalty which the law entitles him (C) to. This is critical in Step 4 of the argument. The assumption that A in fact strongly desires to avoid imprisonment is intended to force him, via the application of universalizability and prescriptivity (Steps 2 and 3) to reflect on his position in relation to C, when assessing what he ought to do as regards his problem with B. The point is that if he wants to decide the issue morally, he must be willing to accept one universal judgement for the situation sketched above. This means that A cannot without logical offence derive a moral judgement from the mere fact that he desires to imprison B, unless, of course, he also accepts a universal desire to the effect that every creditor imprison every debtor, if the debtor cannot pay. However, to have such a universal desire, he must also desire, or at least accept that C should imprison him, if he cannot pay.[9] To accept the corresponding prescription, however, leads to inconsistency, which is exactly what Hare needs to bring A to a decision, via Step 5 of the argument.

Now, the fact that A is actually in the same position with respect to C as B is with respect to himself, is obviously of critical importance, if Hare is to be able to maintain that A can be forced into a position of inconsistency (Step 5). As I judge, this stipulation runs together with

Steps 2 and 3 to yield Hare's conclusion, for by itself it does not determine that any inconsistency arises between A desiring to imprison B, yet desiring at the same time not to be imprisoned by C. But, the conditions under which A may be trapped in inconsistency require careful qualification. As I have formulated Hare's principle of universalizability above, the requirement is that A's position must be like B's in the relevant respects, i.e. similar in respect of the non-moral but morally relevant features of their respective situations, if A's decision about what he morally ought to do as regards his problem with B, can be made to turn, as an essential ingredient of the argument, on what he desires should happen to himself regarding his problem with C.

The point I am stressing here concerns a question about the link between Steps 1 and 2 of the argument. I think the contention that, to be moral, a desire must be universal, supplies one possible answer to the question why Step 2 is compelling. As I judge, to get the argument going, A must be manoeuvred into a position in which he is compelled to ask the question 'How would I like it, if that were done to me?' The expected response, of course, is that he would not like it, i.e. he would not assent to the prescription 'Let me be sent to prison, if I cannot pay.' It may be suggested that our expectation of this response is justified by Hare's stipulation that A's position is in fact like B's in the relevant respects (i.e. he cannot pay C as B cannot pay him), and particularly with respect to the fact that he actually shares B's desire not to be imprisoned in case of non-payment of debt. Hare, of course, supposes that the shared desire is the same in <u>all</u> respects, and particularly in respect of its strength or intensity.[10] It may now be suggested that it follows, from Steps 3 and 4 of the argument, that A would not assent to the prescription 'Let me be sent to prison, if I cannot pay' <u>because</u> he, in his position as creditor to B,

shares a desire which is like B's in all relevant respects. I am saying that his own desire not to be imprisoned in case of non-payment of debt must be thought to play a role in deciding what he morally ought to do regarding his problem with B. I think in compelling A to ask the question 'How would I (the creditor) like it, if that were done to me?', Hare is presuming that it is A's desires, i.e. those he has in his position as creditor to B, which are crucial. He (A) considers those desires in deciding whether he is morally justified in sending B to prison.

1.2 The 'hypothetical' bi-lateral argument

However, would our expectation of the response just referred to be justified if we were to suppose that there are relevant differences? Suppose, for instance, that though A is in fact debtor to C, he can pay his debt. Consider the following reconstruction of Steps 2-5 of Hare's argument.

2. If A is seriously considering whether he ought to imprison B, i.e. whether the prescription he accepts in 1 above, is the one he is willing universally to endorse, he must imagine a hypothetical situation in which he stands to C in exactly the same relation as B stands to himself in the actual situation, and treat this situation as counting for as much as the actual situation.[11]

3. If the situations in question are identical in all relevant respects, i.e. with respect to the relations between the parties and with respect to the desires which characterize their relations,[12] it follows from the universalizability of 'ought', that A can accept the judgement that he ought to imprison B, as a moral judgement, if he also accepts its universalization, which means, in effect, also

accepting the judgement that C ought to imprison him, if he cannot pay.

4. But, if he accepts this judgement, it follows from the prescriptivity of 'ought', that he also accepts the prescription 'Let C imprison me, if I cannot pay.' This, however, leads to inconsistency, since in the hypothetical situation he has also to accept the prescription 'Let me not be imprisoned, if I cannot pay.'

5. In order to avoid inconsistency, A must reject the prescription that C should imprison him, if he cannot pay. If he rejects it - as he must, given that he strongly desires to avoid imprisonment in the hypothetical situation in which he is situated like B in the relevant respects - he must also reject the universalized form of the judgement which entails this prescription.

In this version of the argument, A is faced with inconsistency though he is not ex hypothesi actually in the same position as B. He is, however, hypothetically in a relevantly similar position, since Step 2 requires that he imagines himself to be the person, B, who cannot pay his creditor. How is role-reversal supposed to work?[13] As I judge, A must take B's desire not to be imprisoned into his imagined role as debtor who cannot pay. This is the ground of our expectation that he will respond in the way mentioned above, i.e. that he would not accept the prescription that he not be imprisoned, were he B who cannot pay. He has to treat this desire as his dominant desire for what should happen to himself in that case. I do not think it need be supposed that he has a desire like B's at the outset, for there is no requirement in the 'hypothetical' bi-lateral argument presented above, that this must be the case. Step 2 of this argument is intended to

show that, if he were now, through an act of the imagination, adequately to represent to himself what it would be like to be in B's position, he would acquire a desire which is like B's in all relevant respects. The adequacy of his representation obviously depends on this desire being of the same strength as B's, for otherwise it could be said that he misrepresents B's case in an important respect. I am saying that B's desire need not pertain to A in the actual situation, for he may in fact lack B's aversion to being imprisoned at the time that this is relevant to deciding what he ought to do regarding his problem with B.[14] This, of course, is not incompatible with being averse to himself suffering at the hands of his creditor, at the time he is considering imprisoning B.

Now, I am suggesting that our expectation of A's response in the hypothetical situation is justified on the strength of our supposing B's desire to be the appropriate desire A has in that situation. Hare's appeal to imagination as a necessary ingredient of moral thinking is a point worth noting here.[15] This appeal sits at the heart of the role-reversal procedure in the 'hypothetical' bi-lateral case. But it is not clear how Hare thinks this procedure is supposed to work. Putting A in B's position requires that he imagine himself to be the person, B, who cannot pay his debt. But, does this require that any other desires A may have for what should happen to himself in that situation, must be set aside at the time he considers whether he is willing to be imprisoned? Or, is the requirement at issue here that A must consider only B's desire in that situation? Step 2 of the argument can be read as requiring the latter. This, however, says something quite different from what I suggested above regarding how the argument works in the 'actual bi-lateral' case, for there A's own actual desire not to be imprisoned is the desire we suppose him to be willing to act on, whereas in the 'hypothetical' bi-lateral case B's desire appears to be this

desire. It is, however, worth noting that putting A in B's position does itself not say why B's desire not to be imprisoned should take precedence over any other desires A may have for what should happen to himself in the hypothetical situation nor why A must be thought to be willing to act on B's desire, as Step 5 of the argument suggests.

I have noted the importance of appeal to imagination in the argument. This appeal helps to establish a claim critical to the argument. This is that, in reversing roles with B, A will be compelled to concede that the actual and hypothetical situations are identical in the relevant respects, and hence that no morally relevant difference can arise between himself and B in these situations.[16] This is a highly contentious claim, particularly since it requires clear identity conditions both with respect to circumstances and persons.[17] For Hare the required identity is identity with respect to the relations between A and B in the actual and imagined situations and with respect to the desires which characterize their relations in these situations. I shall examine the identity requirements in some detail in Section 2. For the moment I shall accept that Hare's claim is in order. It is worth noting that this claim supplies the link between Steps 2 and 3 of the argument. Now, if there can be no disagreements about the morally relevant facts, logic in the shape of universalizability and prescriptivity prohibits A from accepting different moral judgements, or the correlated prescriptions, for the situations in question. In this the argument appears to be that, whatever judgement A feels himself constrained to universalize, he must be willing, if he is to make a moral judgement, to accept that the same judgement applies to himself, irrespective of which position he in fact occupies in the circumstances.[18] This, I shall grant for the moment, follows from the (alleged) identity of the relations between A and B in the actual and hypothetical situations, and the desires which

characterize their relations in these situations, taken in conjunction with universalizability and prescriptivity.

Now, there is no doubt that for Hare A is compelled to accept the prescription 'Let me be imprisoned, if I cannot pay' (Step 4) in virtue of the identity relations just referred to, together with universalizability and prescriptivity. Merely appealing to A's imagination, and to his inclinations, including inclinations we might reasonably suppose him to have, were his position to change and become like B's in the relevant respects, are not sufficient to compel him to accept this prescription. Acceptance, however, leads to inconsistency, since A is ex hypothesi not willing to prescribe that he should be imprisoned. As I judge, Hare understands the inconsistency to which A falls prey as one of not being able, without self-contradiction, to prescribe that incompatible desires be satisfied. How he chooses, however, is a matter of inclination, for there is nothing self-contradictory about him prescribing the frustration of his desire to avoid imprisonment.[19] If his inclination is such that he cannot accept the prescription 'Let me be imprisoned, if I cannot pay', he must (in accordance with Step 5 of the argument) also reject the moral judgement that he ought to imprison B. Hare thinks that this is what he will do. But, as I judge, this assumes that A must be thought to care less about exacting the penalty in the actual situation in which he stands to B in the relation of creditor to debtor, than he cares about himself avoiding imprisonment in the hypothetical situation in which his position is like B's in the relevant respects. I think this points to a suppressed premiss in Step 5 of the argument. If the relevant desires are (1) that A desires to imprison B, and (2) that A desires not to be imprisoned by C, it has to be supposed that A is prepared to balance (1) against (2), and that (2) is stronger than (1), if Hare's argument is to yield the suggested conclusion.[20] I am

suggesting that trapping A in self-contradiction, in the logical sense in which identity conditions and the demand for universalization make it impossible for him to accept different prescriptions, is itself not sufficient to yield Hare's conclusion.

Step 5 of the argument raises a few issues which need brief discussion at this point. It is worth noting that (a) the requirement that A balance (1) against (2), and (b) the requirement that (2) be treated as the stronger desire, and hence the one he should satisfy, are genuinely different requirements of the argument. The 'hypothetical' bi-lateral argument requires that A compare and balance his own and B's desires as if they were, for him, incompatible desires for what should be done in a situation of a particular kind. But, it may be asked: What conceptual moves commit A to this strategy for settling his problem with B? A, of course, is expected to exchange roles with B and then to accept a prescription which might be called 'his own' on the basis of the desires he imagines himself to have in that role. In this the argument appears to be that, in reversing roles with B - on the interpretation of this procedure which I think Hare endorses - A is expected to identify with the person in B's position, and in particular with the desires B has in his role as recipient of the prescription which A's initial judgement entails. The argument then runs, in virtue of the demand for imaginative identification with B's desires, on committing A to wanting for himself what another person wants should happen, with the same strength as this person wants this something to happen, at the time that wanting this is relevant to deciding what he morally ought to do.[21]

Now, to have a desire that something be done is, according to Hare, to accept a prescription that it be satisfied.[22] In committing A to identify with B's desires, A is expected, in virtue of prescriptivity, to prescribe

that that desire should be satisfied. A can, of course, not accept one prescription for the hypothetical situation and another for the actual situation. For universalizability and prescriptivity jointly set an independent constraint on what he can consistently accept. This is that, if he accepts the prescription that he should imprison B, and thinks he is morally justified in doing so, he cannot consistently reject the prescription that C should imprison him, unless he can show that there is a relevant difference between his position and B's.[23] This is because, if identity conditions obtain both with respect to circumstances and to persons, the demand for universalization will compel him to accept that, whatever course of action he chooses to adopt, the same thing should be done to himself, irrespective of which position he occupies in the circumstances.[24] As I judge, this, together with the demand for imaginative identification, must be thought to determine that, in deciding what he ought to do regarding his problem with B, his treatment of the desires respectively associated with being creditor to B, and being debtor to C, must run along lines analogous to an agent's treatment of incompatible desires in a purely prudential context.

It is perhaps important to stress that I take Hare's view to be that identity-conditions, in conjunction with the demand for role-reversal and imaginative identification, interpreted as constraints of universalizability, must be thought to constrain A to treat B's desires 'as if they were his own'. By this Hare means that A must give those desires the same consideration as he gives his own, i.e. he must treat them on equal par with his own desires.[25] Hare's argument is simply that, if there are no morally relevant differences between A and B in the actual and imagined situations, A cannot give less consideration to B's interests than what he gives to his own interests. Differential treatment of this sort

calls for justification, and can be justified only with reference to differences in respect of the universal descriptive features of their respective positions in (allegedly) identical situations. But, in the case at hand, there are ex hypothesi no such differences. Therefore, in this case differential treatment of the sort just mentioned cannot be justified.

The demand that A compare and balance all relevant desires 'as if they were his own', is the demand that he must give them equal consideration, irrespective of the fact that he knows which position in the circumstances he in fact occupies. As I understand, this is the demand that nothing be made to depend on the fact that A plays the particular role which he does actually play, and that in this role he is the person who desires that a certain thing be done. The point seems to be that A cannot claim that the role he actually performs (as prescriber) accords a privilege which his hypothetical role (as recipient) does not accord, because a distinction between his roles on grounds that one is actual and the other hypothetical is not morally significant, i.e. being 'actual' and being 'hypothetical' are not 'right-making' or 'wrong-making' features of anything.[26] Though A may know which role is the actual one in which he desires that a certain thing be done, this is of no special significance, for this gives him in his role as prescriber no privileged position vis-á-vis his recipient, i.e. his own desires get no extra consideration because they are his in this role.

The requirement which I have just briefly outlined touches on a point crucial to Hare's defence of the import and scope of the principle of universalizability, which I shall examine in Section 2. Here I merely note its significance in the argument under consideration. Hare says:

> 'All that is essential to the argument is that [the prescriber] should disregard the fact that he plays the particular role in the situation which he does, without disregarding the inclinations which people have in situations of this sort. In other words, he must be prepared to give weight to [his victim's] inclinations as if they were his own.'[27]

Hare is here attempting to address the question of how it is possible for A to take account of what he would want - as his own current desire - if he were to be the person in B's position.[28] I have suggested that appeal to role-reversal and the demand for imaginative identification, must be thought to constrain A to treat the desires he would have, if his position were to change and become like B's in the relevant respects, on equal par with his own (though his own actual motivational state for what should happen to himself in the hypothetical situation may differ from this). It is worth noting here that the constraints just mentioned are in themselves not evidence that A is committed to a balancing procedure as a way of settling his problem with B. I think a balancing procedure is tacitly assumed in the argument. This is a point I take up below.

It may be thought that (1) the requirement that A must disregard the fact that he plays the role which he in fact plays, and (2) the requirement that he must treat B's desires 'as if they were his own', amount to the same thing, or at least that accepting (1) entails acceptance of (2) on the grounds that, if his <u>role</u> is disregarded, it follows <u>that it is irrelevant whether he is A or B</u>, in which case he must treat his own desires and B's in the same way. My own view is that this is unacceptable on the grounds that (1) and (2) are distinct, and that (2) is not entailed by (1). Hare thinks that A's deliberation about what he morally ought to do must remain unaffected by considerations concerning (a) who is actually who in each

position he occupies, and (b) which position he in fact occupies in the circumstances. This is a matter which touches on Hare's defence of the import and scope of the universalizability thesis to which I return in Section 2. Here I wish to note that (a) and (b) are different requirements of universalizability. (a) is the necessary ground of the requirement that A must treat B's desires 'as if they were his own', and (b) is the necessary ground of the requirement that hypothetical situations count as actual.[29] I shall argue later (Sections 2 and 3) that the role-reversal procedure - at least the interpretation which I think Hare endorses - demands that A must identify the person in the imagined situation as himself, but that this gives rise to a problem about making coherent sense of his identity in that role.

I understand Step 5 of the argument as requiring that A must treat B's desire not to be imprisoned as his on balance desire for what should happen to himself in the circumstances, for otherwise how is the argument to yield the desired conclusion? I have suggested that on one count, imaginative role reversal enables Hare to fulfil the identity conditions which are needed for universalizability and prescriptivity to get a grip in the argument. Now, in this there appears to be a suggestion that if A adequately represents to himself what it would be like to suffer, as B is about to suffer, his desire not to be imprisoned is likely to emerge as stronger than his desire to imprison B.[30] It then appears that A must be willing to act on the stronger desire, for it may be suggested that this is what his rationality requires. This means that Step 5 of the argument requires that A, in his role as prescriber, must be rational, in the prudential sense of 'rational'.

It may, however, be asked what there is in the argument which determines that a prudential calculation is the appropriate response for the case in which A treats himself as one who has incompatible desires. This is to ask whether there is anything in the argument which commits A to acting on the stronger desire. As I judge, the strategy Hare employs to give the desires and other relevant features of another person a place in A's deliberation is intended to show that his decision about what he ought to do regarding his problem with B, will turn, qua rational agent, on what he most desires should happen to himself overall.

There are three points to note about this strategy. The first concerns Hare's appeal to fact. So far in my exposition this appeal has been confined to stressing the significance of identity-relations in the argument.[31] There is, however, another aspect to this appeal. This is that A cannot be said coherently to be making a decision about what he ought to do regarding his problem with B, without his knowing what it is he would be doing, if he decided to do one or other of the things he could be doing in the situation in question.[32] For Hare this means that the determination of what he is willing to prescribe universally, qua rational agent, is a matter of weighing one set of consequences to himself, should he satisfy one desire, against an alternative set of consequences to himself, should he satisfy the other desire.[33] If his desire not to be imprisoned on grounds of non-payment of debt is stronger than his desire to imprison B on those grounds, as Hare thinks it will be, he must, if he is rational, choose to satisfy the former. In other words, Hare thinks A must be willing to concede, as rationality requires, that he is not willing to override his stronger desire.

I suggested above that there is a suppressed premiss in Step 5 of the argument. This, as I noted, is simply that A's desire not to be imprisoned must be thought to be stronger than his desire to imprison B. Now, I think Hare's appeal to fact, and in particular to consequences, supplies indirectly one answer to the question why, in reversing roles with B, A is likely to accept his desire not to be imprisoned as the stronger desire overall. This brings me to the second point I wish to stress. The acceptability of Hare's answer rests on the acceptability of an empirical generalization. Hare says:

> 'People's inclinations about most of the important matters of life tend to be the same (very few people, for example, like being starved or run over by motor-cars) ...'[34]

I think this generalization sits at the heart of Hare's claim that A will be inclined to reject the prescription that he should be imprisoned in the case of non-payment of debt, and hence that he will reject the moral judgement that he ought to imprison B. If I am right, this generalization accounts for the greater weight Hare thinks A will assign to himself staying out of prison, as against exacting the penalty to which the law entitles him. But, now, does it follow from this <u>in itself</u> that A will be led to accept a form of rational prudence? And, supposing it does, does rational prudence entail that A should <u>not</u> give extra weight to his own desires simply because they are his? I shall examine Hare's appeal to rational prudence below. The point to note here is simply that unless one assumes that A accepts his desire not to be imprisoned by C as his <u>on balance</u> desire for what should happen to himself in the circumstances, there is no reason to believe that trapping A in self-contradiction is sufficient to compel him to accept Hare's conclusion.

The third point to note about Hare's strategy is that it works on a contentious supposition. This is that just as it is rational for A to do what he, prudentially speaking, most desires should be done in his own case, so it is rational for him to do what he most desires should be done overall, in a situation in which he has, in addition to his own desires, other desires he treats 'as if they were his own'. I shall maintain below that there is in this no reason to believe that B's suffering commits A, or gives him a reason to act, as Step 5 of the argument requires.

1.3 The multi-lateral argument: first thoughts

As a preliminary step to reconstructing the argument for the multi-lateral case, I confine myself to a case involving two persons who have different inclinations, using Hare's 'trumpeter'-example. I shall then extend this argument to a case involving more than two persons, using Hare's 'judge-criminal' example. My object is to establish by what conceptual moves the argument for the 'hypothetical' bi-lateral case may be generalized to cover multi-lateral cases as well.

I begin by isolating the requirements Hare thinks must be met to accommodate genuine differences in inclinations.[35] Suppose A wants to play jazz on his trumpet, but B, who lives next door, likes to listen to classical music only. Should A strike up on the trumpet? Suppose it occurs to A (1) that B dislikes jazz and (2) that if he were now to strike up on the trumpet, B would be bored beyond endurance. Hare says:

> '... it is obviously of no use for [A] to ask himself whether he is prepared to prescribe universally that people should play trumpets when they live next door to other people who are listening to classical records. For if [A] himself were listening to classical

records (which bore him beyond endurance) he would be only too pleased if somebody next door started up on the trumpet.'[36]

The point is simply that, in imagining himself to be the person who would be discomforted, A must ascribe to himself that person's likes and dislikes. This is a point Hare justifies with reference to the requirements of universalizability, in particular with reference to the demand for role-reversal and imaginative identification.[37] Identity conditions require that their situations be alike in all relevant respects. I reconstruct Hare's reasoning thus: any person who has some relevant desires of his own defines a case which is identical to another case in which he imagines that some other person plays his role in the same set of circumstances. 'A in S' presents a different case from 'B in S', because they have different relevant inclinations. But the situation A has to imagine himself in must be the same situation that B is actually in, i.e. he has to imagine himself standing to someone else in the imagined situation in exactly the same relation as B stands to himself in the actual situation. This implies that, though there are different cases, the actual and imagined situations will be the same situations for both A and B, i.e. they will differ only in this morally irrelevant respect that A and B have reversed roles.[38]

Now, A can be said to be imagining himself to be in B's position, if he ascribes to himself all and only the relevant features of B's person and situation. These features must include all relevant personal characteristics and, in particular, the relevant desires of B's case. The relevant characteristics and desires must be those which characterize B's position as <u>recipient</u> of a prescription (i.e. the one entailed by the judgement A initially makes in the actual situation).[39] For Hare, however, it is not possible to say that A, i.e. the person with his individual

identity and name, can reverse roles with B, because this implies that A can be himself in this sense while imagining himself to be someone else, viz. B. If Hare is right, it is possible to say that he can imagine himself to be B, if this is understood as meaning that he loses his identifying features, including his individual identity, and acquires another set.[40] A then imagines himself to be the person who acquires all and only the relevant universal features of B's person and situation. For Hare role-reversal turns on a notion of knowing what it would be like to experience what another person, in his role as recipient, is about to experience. Suppose, as I have assumed, that A knows B is about to be made to suffer. If Hare is right, A cannot know what it would be like for himself to suffer, as B is about to suffer, unless he shares, in the imagination, B's desire not to be made to suffer.[41] As I read Hare, to say that A has reversed roles with B, it is necessary that he now (i.e. at the time he considers whether he ought to strike up on the trumpet) acquires a desire for what should happen to himself in the imagined situation, which must be of the same content and intensity as B's desire.[42] Hare's point is that unless he responds, in the imagined situation, in a way that rules out remaining unmoved by what he is about to suffer, if he were B, he cannot be said to have reversed roles with B.

Now, this way of putting the requirement Hare has in mind raises an interesting point. Hare rightly observes that the moral judgement A makes about the imagined situation has to remain 'his own', if it is to have any significance in the argument.[43] Hare says we should phrase the question A is expected to consider in the hypothetical situation thus:

> '... What do you say (in propria persona) about a hypothetical case in which you are in your victim's position?'[44]

It is of no use to ask him what he would say, or how he would like it, if he were B. For this apparently permits A (among other things) to evade expressing an opinion at the time that he contemplates striking up on the trumpet. The crucial consideration, however, does not touch this point. As I judge, Hare is suggesting that it is possible for A to say of himself that he desires what B actually desires for the situation in which A is about to strike up on the trumpet. For, in reversing roles with B, he has to imagine that he is the person who is about to be made to suffer. The requirement is that he must think of himself as identical with the person in that position, though in that position he will have a different set of universal features. This requirement is an essential ingredient of the so-called 'in propria persona' requirement,[45] which I think makes up a part of the identity-conditions Hare needs for universalizability and prescriptivity to get a grip on the argument.

It is important to note that A's transition from lacking a desire that B not be made to suffer, to acquiring a desire that he not be made to suffer, if he were B, is effected through the role-reversal procedure. The contention that A is identical with the person in the imagined situation is the crucial point, though Hare does not make this explicit. This contention is the claim, noted in my discussion of the 'hypothetical' bi-lateral argument, that A is being constrained to want for himself what another person wants should happen, at the time that wanting this is relevant to deciding what he morally ought to do. Now, this claim, which assumes an application of the principle of universalizability,[46] apparently helps Hare to establish another claim crucial to generalizing the argument outlined in the 'hypothetical' bi-lateral case. Consider what Hare says about how the argument is supposed to run.

> 'The natural way for the argument then to run is for [A] to admit that he is not prepared to prescribe universally that people's likes and dislikes should be disregarded by other people, because this would entail prescribing that other people should disregard his own likes and dislikes.'[47]

How does this enable Hare to generalize the argument? Suppose, as I think Hare assumes, that a moral judgement applies equally to anyone who falls within its scope, i.e. to anyone who, like A, lives next door to someone else who likes classical music.[48] Just how A must be thought to be constrained might be made clear by saying that he is expected to balance the advantages to himself of disregarding B's desires, against the disadvantages to himself, if he were the person in B's position of having the desires he ascribes to himself in that position disregarded by someone else. It seems that the version of role-reversal sketched above, and particularly the demand for imaginative identification, interpreted as requirements of universalizability, must be thought to constrain A to give weight - prudentially speaking - to desires he would have, if his position were to change and become like B's in the relevant respects.

Now, in prescribing universally, A must be prescribing for all situations which are like the actual one in the relevant respects, irrespective of which position he in fact occupies in the circumstances. For Hare the universalized form of a prescription expresses a universal desire that something be done in all relevantly similar situations, actual and hypothetical, which differ only in this morally irrelevant respect that different persons perform the various roles in them. How does a prescriber arrive at the universal desire which is presupposed in accepting a universal prescription? If the demand for role-reversal is interpreted as

demanding that he identify with his recipient in his role as recipient, and this is interpreted as demanding that he identify with the desires this person has in that role, then what he wants should happen universally will depend on what he and his recipient actually want should happen to themselves in their respective roles. If this be granted, A is expected, in deciding what he wants should happen universally, to consider all relevant desires.

Now, as I have reconstructed the 'hypothetical' bi-lateral argument, A need only balance out the relevant desires in order to arrive at a moral conclusion. In this, I suggested, A must be thought, qua rational agent, to be deciding on the basis of a principle of rational prudence. However, it appears that in generalizing the argument, the matter is not quite so straightforward. Consider how Hare thinks the argument is supposed to run.

> 'It does not follow ... that [A] will conclude that he ought never to play the trumpet when [B] is at home, but only that he will not think that he ought to have no regard at all for [B's] interests.'[49]

The conclusion is that A, the trumpeter, should not sacrifice his desire to play the trumpet, as A, the creditor, should sacrifice his desire to exact the penalty. Hare thinks A should co-operate and reach some sort of compromise with B, which he suggests would consist in a fair apportionment of time during which each may alternatively inflict his likes upon the other. What requires the compromise rather than the sacrifice?[50]

As I judge, the determination of what A wants should happen universally is not settled simply by appeal to the 'balancing' procedure. This procedure, as I suggested in my discussion of the 'hypothetical' bi-lateral argument,

assumes a particular view about how the interests of the parties concerned are to be combined in order to yield a moral conclusion. In this regard I noted Hare's reliance on a strategy which allows their interests to be treated along lines analogous to a purely prudential calculation. It may, however, be thought that this strategy is less likely to work in cases in which inclinations are genuinely different, and that something needs to be added to show how it can be made to work. I am here referring to a link Hare sees between universalizability and a formal impartiality,[51] i.e. an impartiality which takes equal account of A and B's desires, irrespective of their content.

It seems that role-reversal, and particularly the demand for imaginative identification, interpreted as requirements of universalizabilty, must be thought to constrain A to give equal consideration to everyone who would be affected by what he proposes to do, which is to say, if my reconstruction of the argument is correct, that he must be thought to be constrained to take equal account of all relevant desires. In this there is a demand for at least a formal impartiality, which Hare thinks is secured by the demand that a universal prescription has to apply to all similar situations, including hypothetical situations in which the prescriber plays a different role.[52] This is to say, impartiality is secured by the demand that the same moral judgement must apply to A, irrespective of whether he is jazz-player or classical-music lover. Hare's appeal to a formal impartiality tries to make the point that a prescriber is committed to treating the inclinations of his recipients as relevant to the moral appraisal of the act he contemplates performing, no less than his own, irrespective of the intrinsic worth, the rationality and possible causes of the desires he is expected to admit as relevant in the reversed situation. I am suggesting that the appeal to impartiality must be thought to address the question

about the proper treatment of A's acquired desires. It is important to note here that Hare is supposing that A is not apathetic about interests he imagines himself to have in his role as recipient.[53] This is the point, noted above, that A must be thought to be constrained to give weight, prudentially speaking, to the desires he ascribes to himself in that role. As I judge, this has a direct influence on the point Hare tries to carry via the appeal to impartiality, which is that A is expected to give positive weight to all relevant desires.

Now, how does the appeal to impartiality help Hare to establish his conclusion? I think impartiality requires that the 'balancing' be done in a certain way.[54] If A treats the frustration of his actual and hypothesized desires as of equal disvalue to himself, as I think the appeal to impartiality and rational prudence requires, it is not unreasonable to suppose that he will be willing to adopt a line of action which least frustrates them, i.e. that he will be inclined to opt for a compromise with B. The demand for the compromise is, in effect, the demand for a principle that requires the action which maximizes the satisfaction of the desires of all affected persons, or the demand for a moral judgement consistent with a form of utilitarianism.

1.4 The multi-lateral argument: second thoughts

I turn now to the final moves Hare makes to generalize his argument. In the multi-person case the prescriber has to consider everyone who will be affected by what he proposes to do because their desires affect which universal prescription he can in the end accept.[55] Suppose, as Hare's 'judge-criminal' example goes, a judge wishes to imprison a duly tried thief. But the thief objects thus:

> 'You wouldn't like to be sent to prison, if you were me, so how can you universalize your prescription to send me to prison? But if you can't, then how can you maintain that you ought to send me to prison?'[56]

Hare makes clear that the situation in which the judge imagines himself to be performing B's role, is but one of many imagined situations he has to consider. According to Hare's conception of the import and scope of the universalizability thesis, a moral judgement has to apply to all relevantly similar situations, irrespective of who is actually who in those situations, and irrespective of which role the prescriber actually plays in them. The judge then is apparently entitled to reply thus:

> 'If it were just you and I, then of course I might not feel obliged to send you to prison. But I am considering the people whom you will rob, and whom other people, encouraged by your example, will rob, if I don't put you in prison, and I find it easier to universalize the maxim that thieves should be put in prison.'[57]

Now, suppose that A, the judge, is committed to considering all relevant desires in virtue of the moves outlined in (1.3) above, and that the relevant desires are those of all persons whom B might rob. As I read Hare, the judge is justified in prescribing that his desire not to be imprisoned should be frustrated, if he were the person in B's position. It is clear that the justification of the conclusion that the judge ought to imprison B presupposes a reference to the interests of everyone who would be affected. Hare is suggesting that, <u>in sum</u>, the weight of every citizen's desire not to be robbed outbalances B's desire not to be imprisoned. As I judge, appeal to the balancing procedure carries the argument. Two questions arise

here. What is there in the argument which commits A to summing and balancing desires? How, on what basis, are weights to be assigned to the competing interests?

I think it is by now clear that the demand for imaginative identification with the desires of all persons concerned is the most significant move Hare makes to give the interests of distinct individuals a place in the prescriber's deliberation. Hare notes that

> '... in a multilateral situation, the agent has to consider the interests of <u>every</u> person who is affected. This seems in accord with a way in which we do sometimes argue, but it leaves unanswered the question of how, when we have considered all these interests, we combine the consideration of them into a single answer to our moral problem.'[58]

Hare suggests applying a corollary of the principle of universalizability. If two people ought to be treated differently,

> '... some difference must be cited as the ground for these different moral judgements.'[59]

Now, what is there in this which commits A to the 'balancing'-procedure? I think the requirement that A treat all relevant desires 'as if they were his own' sits at the heart of the demand for summing and balancing.[60] The demand is simply that A should combine each person's concern for the satisfaction of his desires <u>in his own person</u>, and then treat this on a par with a conflict of different interests in his person.

If appeal to the 'balancing'-procedure carries Hare's conclusion, there arises a need for some constraint which must be thought to commit A to Hare's principle of rational prudence, i.e. to acting on what he most desires should happen overall, in all relevantly similar situations, including hypothetical situations in which he identifies, in turn, with the desires of each person concerned. Hare thinks that rational prudence is required by universalizability, and particularly by the demand for role-reversal and imaginative identification.[61]

But Hare is assuming that just as it is rational for A to act on the basis of what he, prudentially speaking, most desires should be done in his own case, so it is rational for him to act on what he most desires should be done overall, for the case in which he has, in addition to his own desires, other desires he is willing, through an act of the imagination, to treat 'as if they were his own'. Without this assumption it is difficult to see why universal prescriptivism commits A to summing and balancing desires in a way that preserves a connection with his motivation to act on the stronger desire. If this assumption be granted, the argument leading to the conclusion that A should sum and maximally satisfy desires would then run on the appeal to rational prudence. It is, however, not clear that a request for rational prudence is justified in this context. It is not enough to say that A is constrained to treat all relevant desires 'as if they were his own'. For, besides the objection that this assumes that good sense can be made of A calling the judgement he makes in the reversed situation 'his own', Hare's assumption that a prudential calculation is the appropriate response in two-person and multi-person cases, itself assumes that A acknowledges a reason to act on the stronger desire, simply on grounds that such action accords with his own interests.

I have just suggested that the appeal to Hare's principle of rational prudence is an essential ingredient in the argument to the conclusion that A should sum and maximally satisfy desires. Hare's argument for a maximizing principle runs on the strategy he employs to combine all relevant desires in the person of the decision-maker. Hare treats the demand for imaginative identification with the desires of all affected persons, together with the appeal to impartiality and rational prudence, as implying a commitment to balancing all competing interests on the basis of the principle of equal interests, i.e. on the basis of the principle which bids that a prescriber give equal weight, strength for strength, to all relevant desires.[62] At bottom the argument is simply that anyone who is prepared to universalize his prescriptions, and who is not apathetic about interests he is expected to identify as his own, will be constrained, in virtue of this commitment, to treat the competing interests of different persons as a conflict of different interests in his own person, and hence, as one who is rationally prudent, to give the same weight to other interests as he gives to his own equal interests.

If this be accepted, it is not unreasonable to suppose that A, qua prudent universalizer, will opt for that course of action which satisfies more rather than less of his desires. This, of course, suggests that the determination of what A desires should be done universally, will be the result of summing like desires of different strengths, and allowing like desires of greater strength to override different, weaker ones. This is to say, the final outcome in any conflict of interests will depend, ultimately, on balancing desires for different courses of action solely in proportion to their strength, and accepting as right that course of action which is expected to yield the greatest possible sum of desire-satisfaction over all the possible situations - taken together - in which the prescriber

performs various roles. The desires in question may, but need not, correspond to his own actual desires for what should happen to himself in each hypothetical situation. It is the satisfaction of these desires he should seek to maximize by finding a moral principle that requires the action which ensures that the sum of desire-satisfaction will be maximized, as utilitarianism requires.

Hare sums up what I have here presented as the conclusion of his argument thus:

> '... when I have been the round of all the affected parties, and come back, in my own person, to make an impartial judgement giving equal weight to the interests of all parties, what can I possibly do except advocate that course which will, taken all in all, least frustrate the desires which I have imagined myself having? But this (it is plausible to go on) is to maximize satisfactions.'[63]

It is, however, not clear why A should accept a maximizing principle. Specifically, it is not clear what constrains A to act on the stronger desire, all things considered. Consider the following adaptation of Hare's 'creditor-debtor' example. Suppose that A, who is creditor to B, is also B's judge. Assume that B's desire to avoid imprisonment is stronger than A's desire to imprison him on grounds of non-payment of debt. Assume also that A, in his role as judge, considers the following courses of action.

(a) He considers imprisoning B, for the maximum period which the law allows, which action he expects will result in significant frustration of B's desires, but also significant satisfaction of his own desires.

(b) He considers imprisoning B, for the minimum period which the law allows, which action he expects will result in less significant frustration of B's desires, but also some significant satisfaction of his own desires.

(c) He considers imposing a suspended sentence, which action he expects will result in less than minor frustration of B's desires, but only small satisfaction of his own desires.

(d) He considers imprisoning B, but with provision for release as soon as members of B's family had redeemed his debt, which action he expects will result in more than minor frustration of B's desires, and more than small satisfaction of his own desires.

Now, if A balances all relevant desires on the basis of the principle of equal interests, but wishes only to minimize frustration of desires, he should opt for (d). On the other hand, should he seek to maximize satisfaction of desires, (c) would be the logical option. What is there in the argument presented which commits A to (d) rather than (c)? As I judge, the case for maximizing the average has as strong a claim, on the argument Hare presents, to what is morally right, as maximizing the sum which he attempts to run on the basis of giving equal weight to equal interests.

Section 2. The defence of the universalizability of moral judgements reconsidered.

I have in the course of presenting an outline of Hare's theory of moral reasoning drawn attention to the role he assigns to the thesis of universalizability. In this section I shall examine a renewed attempt to defend this thesis. My main concern will be to show how Hare understands

the claim that moral judgements are universalizable prescriptions. The defence of this claim proceeds via the defence of universalizability with reference to the notion of a descriptive meaning-rule. I shall, accordingly, first have something to say about this.

2.1 Universalizability and the descriptive-meaning thesis

In Section 1 of Chapter 2 I drew attention to Hare's view that it is not possible without logical offence to derive a prescriptive conclusion to the effect that a particular act ought to be done, from the mere statement that the act has certain descriptive features, unless one also accepts a universal prescriptive principle which says that acts of the given description ought to be done. I also indicated that for Hare some set of the act's descriptive features are, or can be, the ground of the prescriptive conclusion, if one accepts the principle in Hare's evaluative or prescriptive sense. I tried to show, by means of an examination of Hare's account of supervenience, that the features of the act in question are necessarily relevant to the judgement one makes about the act, in virtue of one's acceptance of the principle, but that Hare's account of supervenience provides no direct support for the claim that the principle must be universal in the sense advanced in U.

In Freedom and reason Hare reiterates the claim, advanced in LM, that in so far as moral words have descriptive meaning, they behave much like ordinary descriptive terms. It is part of what it means for a term to be a descriptive term that it applies to all things which are similar in the relevant respects.[64] In FR Hare says that what this similarity is, is laid down by a descriptive meaning-rule, which being a rule governing the use of the term, is a universal rule.[65] There is in this an innovation which I wish to draw attention to. This is the explicit claim - which I shall call the

descriptive meaning thesis - that the universalizability of singular moral judgements is the logical consequence of the fact that they carry descriptive meaning. Hare expresses this by saying that the claim that moral judgements have descriptive meaning, as one element of their meaning, entails the claim that they are universalizable, which implies that they are universalizable in much the same way as ordinary descriptive judgements are.[66]

But, Hare is careful to point out that though descriptive meaning-rules govern the use of moral terms, just as descriptive meaning-rules govern the use of purely descriptive terms, the other element in the meaning of moral terms, the evaluative or prescriptive meaning, determines that their meaning-rules are more than just descriptive meaning-rules. This 'more' is an agent's commitment to a particular substantive moral principle.[67] What Hare has in mind here might be clarified thus: if an agent accepts a moral judgement, say about some act, this implies that he also accepts the moral principle which is presupposed in making the judgement, since it entails the judgement when asserted in conjunction with a statement of those descriptive features of the act which he deems to be relevant to its moral appraisal. Now, his decision to treat these features as morally relevant logically depends on his decision to apply his principle to the act, or to make a moral judgement about it. If he is using the moral term employed in the judgement in its evaluative or prescriptive sense, he implies that he accepts a prescription to the effect that the act ought or ought not to be done because it exhibits the features in question, i.e. he in effect offers moral instruction. Hare thinks that this fact alters the logical character of the descriptive meaning-rule to which he is committed whenever he offers his judgement about the act, which change is reflected in the fact that by

applying a moral term to some set of the act's descriptive features, he in effect accepts a moral principle of substance.

Now, the descriptive meaning thesis says that whenever an agent makes a moral judgement about something, he is committed to a universal rule, which in turn commits him to the

> '... proposition that anything exactly like the subject of the first judgement, or like it in the relevant respects, possesses the property attributed to it in the first judgement.'[68]

As I interpret Hare, in using a word like 'ought' in a moral judgement about some act, an agent is committed not merely to a universal rule, but also to a moral principle of substance which gives certain descriptive features of the act their relevance, and which says that those features are the logical ground of his ascription of exact or relevant similarity to all acts conforming to the given description. Hare is careful to point out that the thesis of universalizability is a purely logical thesis, in the sense that nothing of substance can be derived from it alone. Hare says:

> 'By a "logical" thesis I mean a thesis about the meanings of words, or dependent solely upon them. I have been maintaining that the meaning of the word 'ought' and other moral words is such that a person who uses them commits himself thereby to a universal rule.'[69]

As I understand, universalizability is logically dependent on a meaning-rule. This is the property of being governed by a universal quantifier and containing no references to individual constants (other than by description). To exemplify, using one of Hare's own examples,[70] the

judgement that it is illegal to marry one's own sister cannot be universalized in the sense in which the meaning-rule requires, since legal judgements refer, implicitly or otherwise, to the jurisdiction of some country. It is interesting to note that the judgement is not universalizable in the sense that it commits an agent to the view that it is illegal for anyone

> '... in any country that was otherwise like England'[71]

i.e. 'otherwise like England' in the relevant respects, to marry his sister. For this still assumes a reference to the jurisdiction of some country, however similar the countries in question may be in the relevant respects, i.e. in respect of the prohibition on marriages between siblings.[72] In contrast, the moral judgement 'No one ought to marry his own sister' is universal in the sense Hare requires.

Hare's reason for maintaining that the legal judgement referred to above cannot be universalizable is simply that it is not possible to substitute for 'like England' a description in purely universal terms. If I am right, Hare's sense of 'similar' (which is similarity only with respect to the universal descriptive features of whatever a proposition is about), is critical for Hare's defence of moral judgements as universalizable prescriptions. This is a matter I shall attend to shortly. I think it is a direct consequence of the fact that this is the sense of 'similar' required by a meaning-rule, that the following claim, which I shall call the inconsistency thesis (IT), is possible.

(IT) An agent, A, contradicts himself if he allows that it is legitimate for himself, or anyone else in relevantly similar circumstances, to

make different moral judgements about two acts or situations he concedes to be exactly or relevantly alike.

To forestall a possible misinterpretation of the import of (IT), it is important to be clear on what it commits an agent to. On Hare's view, A cannot intelligibly judge that he ought to do x in S while refusing to commit himself to the view that anyone else in relevantly similar circumstances, say B, ought also to do x. To be consistent, A must judge that C does wrong, if C decides differently in those circumstances. But there is no requirement binding on C which commits him to the view that he contradicts himself, if he decides differently.

As I read Hare, it does not matter <u>who</u> is in circumstances relevantly similar to S, for a charge of self-contradiction to stick to A, should he allow that this person ought not to do x in S. It is clear that if A can be said to be guilty of self-contradiction, should he allow that C ought not to do x in relevantly similar circumstances, it must be the case that he can justifiably claim that what is not a morally relevant difference for him, can also not be a morally relevant difference for C, though he and C are different persons. As I judge, the defence of a claim to the effect that A contradicts himself in the sense made clear in (IT), would have to show that no morally relevant difference can arise <u>for A</u> from this simple fact. It is worth noting that (IT) defines one sense in which moral judgements are universalizable. This, of course, implies that the defence of moral judgements as universalizable prescriptions must show how the problem of relevance is to be addressed. To this matter I now turn.

2.2 The universalizability of moral judgements

In Section 1 of Chapter 1 I tried to show that Hare's principle of supervenience is much too weak to support the sense in which he claims moral judgements are universalizable. I argued that on Hare's principle of supervenience, it does not follow that A contradicts himself, should he allow that doing x may be wrong for C, for this principle allows that the same act in the same set of circumstances may without logical or linguistic error be judged differently by different persons.

Now, does it follow from the meaning of the word 'ought', that in accepting the moral judgement 'X ought to be done in S', that A is thereby committed - on pain of self-contradiction - to the view that C ought also to do x in S? I begin with some considerations concerning what is involved in committing A to a universal moral judgement.

I have in Section 1 of this chapter briefly shown how Hare thinks the role-reversal procedure can establish similarity in the relevant respects. I indicated that for Hare relevant similarity between two situations is always similarity in respect of some set of their universal descriptive features. A relevantly similar situation, however, need be only an imagined situation. For, as Hare observes, it is unlikely that two situations will arise in practice in which the parties concerned stand to each other, respectively, in exactly the same relations.[73] Now, granted that this is so, it may be asked which universal features of the persons concerned and of their circumstances are the ones that make up the imagined set in order to secure the identity with respect to persons and circumstances which Hare needs to commit A to a universal judgement. As I tried to show above in my discussion of the role-reversal procedure, Hare thinks that the relevant set of features is the set which characterizes the relation between

prescriber and recipient in his position as recipient of the prescription entailed by the judgement the prescriber initially offers in the actual situation. Hare thinks that identity in the relevant respects between the actual and imagined situations simply is similarity with respect to the relations between prescriber and recipient in each situation, and with respect to the desires which characterize their relations in each.

On what grounds might it be claimed that Hare is entitled to this view of what shall count as 'relevantly similar'? It is necessary here to distinguish between two separate requirements of universalizability. These are that, in universalizing a prescription, a prescriber must treat all relevant desires independently of who is actually who in each position he occupies, and independently of the fact that he knows which position he in fact occupies in the circumstances. The first is the requirement that references to certain features, for instance, to particular individual identities, which cannot be specified without the use of singular terms, be granted no weight in his deliberation about what he morally ought to do. The point seems to be that he cannot claim that the fact that he is identical with himself is a feature which distinguishes his case from his recipient's, because this is a feature of his recipient's case as well, i.e. being identical with oneself is not the kind of feature which introduces a morally significant difference. The second is the demand that nothing be made to depend on the fact that he plays the particular role which he does in fact play, and that in this role he desires that a certain thing be done. The point is that he cannot claim that the role he actually performs (as prescriber) accords a privilege which his hypothetical role (as recipient) does not accord, because a distinction between his roles on grounds that one is actual and the other hypothetical is not morally significant, i.e. being actual and being hypothetical are not 'right-

making' or 'wrong-making' features of anything.[74] The point is that though A may know which role is the actual one in which he desires that a certain thing be done, this is of no special significance, for this gives him in his role as prescriber no privileged position viz-á-viz his recipient, i.e. his own desires get no extra consideration because they are his in this role.

Now, the requirements just mentioned should be distinguished from the requirement which demands that a prescriber must ascribe to himself the relevant universal features of his recipient's person and situation. Hare's strategy for dealing with the problem of relevance runs via this demand, which is, of course, the by now familiar demand that he must identify with the person in his recipient's position. As I have interpreted this demand above, a prescriber is expected to identify with his recipient's desires. This is part of what Hare needs in order to maintain that a prescriber is situated like his recipient in the relevant respects, i.e. that he is the recipient of the prescription which directs that he be treated in the way in which his recipient does not want to be treated. It is obviously important that the judgement a prescriber accepts in the imagined situation must be thought to be 'his own', for otherwise how is it to have a bearing on the claim that he contradicts himself should he accept different singular judgements for situations which are relevantly alike? The crucial point is that one and the same person must be thought to be occupying different positions in the circumstances and desiring that different things be done in those circumstances. If I have understood Hare correctly, then the following version of the universalizability thesis must be the one he presupposes in his attempt to establish the required identity conditions. I shall call it the stringent version (SUP).

(SUP) If some agent, A, accepts the moral judgement that x ought to be done in S, this implies that he is willing to prescribe x in each position he occupies in the circumstances, indifferently of the fact that he prefers to be in one position rather than another, and indifferently of the fact that in that position he desires that x be done, which is to say that the object and intensity of his desire must be independent of which role he in fact plays and of the fact that in this role he is the person who desires that x be done.[75]

As I understand the import of (SUP), a judgement qualifies as a moral judgement if and only if an agent accepts that the action it enjoins is right for himself, and from his perspective, right for anyone else who is in circumstances relevantly similar to those in which the judgement applies. If this is right, there can for A, and from A's perspective, for anyone else in a relevantly similar set of circumstances, be no action which is right and wrong. But, it being right for A, on penalty of self-contradiction, as this is presented in (IT), presupposes that no morally relevant difference can arise for A and C from the fact that they are different individuals.I think this view about the import and scope of universalizability reflects Hare's confidence about the adequacy of the role-reversal procedure, and the demand for imaginative identification, to accommodate relevant differences between persons, and to establish that it is possible, in principle at least, to get agreement about what the morally relevant facts of a situation are. I shall argue later that the suggested extension of the scope of (SUP) to include a reference to 'anyone else', as this is presented in (IT), is strictly illegitimate.

Now, if, in making a moral judgement, an agent is committed to a universal meaning rule, the content of the rule will depend on what features of an

action or situation he will be constrained to treat as relevant in virtue of his acceptance of the judgement. If this is correct, Hare needs a criterion of moral relevance. I formulate this criterion as follows:

(MR) Any universal feature of an action or situation is a feature relevant to their moral appraisal, if an agent treats the feature as relevant in any position he occupies in a given set of circumstances, indifferently of which position he in fact occupies and indifferently of who is actually who in each position.[76]

In what follows I shall test this criterion against one of Hare's own examples. My object is to establish whether Hare can be said to have succeeded in establishing that moral judgements are universalizable prescriptions in the sense explained in (IT). Hare's defence of this matter runs <u>via</u> the test of relevance. This is to say, in effect, that Hare's attempt to establish (IT) depends on whether he can make the test of relevance work.

Suppose that A, who is white, wants to discriminate against B, because B is black. Now, in reversing roles with B, it follows, if Hare is right, that no morally relevant difference can arise between A and B in the actual and imagined situations. The contention that A is identical with himself in the imagined situation, though in that situation he ascribes to himself a different set of universal features, which include B's desires, helps to establish this claim. But we may ask: what prevents A from disregarding his acquired desire on grounds that the situation in which he imagines himself not to want to suffer, as he proposes to make B suffer, is <u>merely</u> hypothetical? And what prevents A from disregarding B's desire not to be made to suffer on the grounds that, if he were to acquire the relevant

universal features of B's person and situation, including B's black skin, 'he' would not be A, and hence that the contention that he has to choose between satisfying two incompatible desires is false?

Hare's answer to the first question is simply that being hypothetical cannot be a morally relevant feature because it cannot be a 'right-making' or 'wrong-making' feature of anything. Hare says:

> '... whenever we desire anything, we desire it because of something about it; and, since being hypothetical and being actual are not, in the required sense, "things about" objects or events ... it is impossible for there to be anything about the hypothetical similar situation which makes us desire something different concerning it.'[77]

Part of Hare's answer to the second question is contained in the assumption that there is nothing logically amiss with A imagining himself to be the person who has all and only the relevant universal features of B's person and situation. It is, however, clear that for Hare to be able to maintain that A contradicts himself should he accept different singular moral judgements for relevantly similar situations, it must be possible to make good sense of the claim that he is identical with himself in each position he occupies in the circumstances, though this means ascribing to himself different sets of universal features. I shall allow, for the moment, that this claim makes sense on the supposition that the actual and imagined situations exist within a logically possible world in which it is possible that A is identical with himself in alternate positions in which he is successively white and black, and in which the desires respectively associated with being white and being black are both for him desires of the self. It might now be suggested that A's treatment of the desires

respectively associated with being white and being black will run along lines analogous to an agent's treatment of incompatible desires in a purely prudential context. One might say that A is the decision-maker who contemplates realizing one of two possible situations, and that his choice between the alternatives depends on which one of his desires he feels himself constrained to satisfy. Hare clearly needs this supposition in order to make his criterion of relevance run, all other things being equal, on the logical impossibility that the proposed act of discrimination can have different and opposite values for A.[78] If this is right, A is, within Hare's possible world, being constrained to accept that, whatever universal rule he adopts, the same moral judgement must apply to himself, irrespective of whether he is white or black. This is a crucial point. This way of putting the matter restates what I have called the stringent version of universalizability (SUP).

For Hare it follows from the constraint of universalizability just referred to, and the thesis of prescriptivity, that A cannot coherently reject the singular judgement which entails the prescription that he should be discriminated against in the case in which he is black, and not also reject the universalized form of this judgement. The test, then, whether A is putting forward a moral argument is whether he is willing to treat B's black skin as relevant in the case in which he is black. If it be granted that there is a possible world in which A is the person who desires that different things be done in alternate positions in which he is successively white and black, he would contradict himself if he prescribes that both desires be satisfied. A is then in a position in which he can avoid self-contradiction only by universalizing one of two possible judgements, which is to say that he can avoid self-contradiction only by making a decision of relevance.

I noted above that A's choice between realizing one of two possible situations depends on which one of his desires he feels himself constrained to satisfy. The decision of relevance will run - all other things being equal - on A being constrained to balance the desires respectively associated with being white and being black as if they were, for him, incompatible desires of varying strength. This amounts to saying that the decision of relevance turns on the respective strengths of these desires. If A is inclined to reject the prescription that he should be discriminated against in the case in which he is black, as Hare supposes,[79] he must also reject the moral judgement which entails this prescription. This means that he cannot then accept that having a black skin is a morally relevant feature. The point is simply that the prescription which he is willing on balance to accept must be thought to constitute a moral judgement. I have suggested that A needs to make a decision of relevance - which amounts to saying that he needs to balance two incompatible desires of varying strength - in order to establish which one of two possible universal rules he should adopt. For Hare A's final decision depends on him conceding that he is not willing to override his stronger desire, i.e. the desire that he should not be discriminated against in the case in which he is black. Though it would not be self-contradictory for him to prescribe the frustration of this desire, Hare thinks he is unlikely to do want to do so. I think Hare here appeals to the empirical generalization to which I have drawn attention above. Hare also thinks that role-reversal and imaginative identification would fail to be effective as a bar to him prescribing this, only if he fails to make it sufficiently clear to himself what his experiences at the receiving end would be like. This, of course, is to contend that given the required depth of representation of what it would be like, the desire not to be made to suffer is likely to emerge as the stronger desire.[80]

If I have correctly reconstructed the line of reasoning which leads to the test of relevance, the following three premisses are the crucial ones.

(a) A can want for himself what another person wants should happen, with the same strength as this person wants something to happen, as the demand for role-reversal and imaginative identification require, at the time that wanting this is relevant to deciding what he morally ought to do.

(b) A must accept, as universalizability requires, that whatever universal rule he adopts, the same moral judgement must apply to himself, irrespective of whether he is white or black.

(c) A must be willing to concede, as rationality requires, that he is not willing to override his stronger desire that he should not be discriminated against in the case in which he is black.

As I suggested above, (b) is a restatement of what I have called the stringent version of universalizability (SUP). It is obvious that (b) provides a basis for the test of relevance, and that (b) plays a role in committing A to (a) and (c). Now, this test will run only if the strategy Hare employs to give the desires and other characteristics of distinct individuals a place in the decision-maker's deliberation, works in the way suggested, viz. that one and the same person must be thought to be making a decision, which affects another person, on the assumption that he is identical with himself in each role he performs. This is essential to maintaining that the proposed act of discrimination cannot, without inconsistency, have different and opposite values <u>for A</u>. I have also suggested that the test of relevance runs, in part at least, on an appeal

to the 'balancing' procedure and Hare's principle of rational prudence as applied in bi-lateral cases. The application of a prudential rule in these cases works on the supposition that A will not be apathetic about interests he deems to be 'his own' in hypothetical situations. This supposition works with an assumption of rationality. This is the assumption, implicit in (c) above, that just as it is rational for A to do what he, prudentially speaking, most desires should be done in his own case, so it is rational for him to do what he most desires should be done in a situation in which he treats himself as one who has incompatible desires.

Now, for Hare a descriptive meaning-rule is a universal rule. Hare notes, however, that any principle can be rendered properly universal simply by substituting universal terms for all terms that function as individual constants. This is what Hare calls the 'first stage' of universalization.[81] However, a principle which conforms only to this stage of universalization does not necessarily qualify as a moral principle. To qualify as such, the principle has to pass what he calls the 'second stage' of universalization, which involves meeting the requirements of the test of relevance. From a principle which meets the requirements of this test, a moral judgement may be deduced, together with a statement of relevant fact. To qualify as moral, A's desire to discriminate against B must be deducible from the universal desire that every white man should discriminate against every black man. However, to say that A has this universal desire, is to say that he desires its satisfaction in all cases subsumable under the corresponding universal principle, including the case in which he is black. This means that the desire to discriminate against B would be universal in the sense required by (SUP), if and only if what he desires for himself qua white is also what he desires for himself qua black. The action which this desire motivates would then be morally right.

Here I wish to note two things. First, Hare's account of why A must be thought to be constrained to accept one moral judgement for the situation in hand needs an argument to the effect that it is coherent to suppose that the person in the imagined situation is A, and hence that he cannot without self-contradiction allow that the act of discrimination can have different and opposite values for himself qua white and qua black. Second, it is not difficult to see that Hare's sense of 'universal' places extraordinary strain on his account of supervenience.

In Section 1 of Chapter 1 I tried to show that on Hare's account of supervenience, it is possible for different persons, without logical or linguistic error, to evaluate the same act in the same set of circumstances differently. For Hare supervenience works on the presupposition of a universal prescriptive premiss, which together with a secondary substantive premiss (a statement of relevant fact), yields the conclusion that something is right or good. But from this it does not follow that the same thing is right or good for anyone else who does not subscribe to the same principle. To put this differently, it does not follow from supervenience alone that, if A judges the act of discrimination to be right and anyone else in relevantly similar circumstances judges it to be wrong, that A can say this person is in error, simply by appeal to consistency. For this result to obtain, it has to be the case that 'anyone else' is 'anyone else relevantly similar to A', i.e. anyone else who accepts the same substantive moral principle about the rightness of the act.

Now, to get to a point where it can be allowed that there is a place for supervenience in Hare's response to the problem of relevance, A must be thought to be the same person, qua white and qua black. It can then be allowed that there is something about what A desires for himself qua white,

and what A desires for himself qua black, which makes it appropriate for him to attach different moral predicates to whatever this something is in each case. This, of course, will be the quality of A's experiences qua white and qua black. It must also be accepted that one logically possible world is at issue here, for Hare's notion of supervenience requires only that within this world there not to be two things which are exactly or relevantly alike, and which differ solely in their value.[82] It can then be allowed that there is a universal premiss which holds for A qua white and qua black, and which together with a secondary substantive premiss, entails the conclusion that the proposed act of discrimination has a certain moral value for A.

But, in the context of Hare's logically possible world it strains to speak of supervenience, at least of the weak version Hare claims to accept. For, as I judge, there is nothing in supervenience itself which requires that the universal premiss Hare presupposes must conform to what he calls the second stage of universalization.[83] How, then, does Hare make space for the sense of 'universal' which his meaning-rule requires on the strategy he employs to deal with the problem of relevance in the two-person case? (b) above, of course, implies that differential treatment must be justified, and can be justified only with reference to differences in respect of the relevant universal descriptive features of A and B's positions in the actual and imagined situations. This conception of what universalizability requires must be thought to constrain A to accept only one moral judgement for experiences he would have qua white and qua black, as if being white and being black make no difference to who he is.[84] The important feature in this is the demand for consistency. This demand, however, presupposes that good sense can be made of the claim that A is identical with himself in his role qua black. Role-reversal and imaginative identification, however, do

not help Hare to establish that this person must be thought to be A, because in themselves these things set no bar to A thinking of that person as B. There can, of course, be no contradiction between the different moral judgements which different persons may accept for the same set of circumstances, a point which my formulation of Hare's principle of supervenience allows. Hare's response to the problem of relevance in the two-person case requires that A must not disregard the knowledge that it is he who occupies different positions in the circumstances. I think this is a point which recourse to a logically possible world cannot sustain, though Hare nevertheless thinks that it is appropriate to deal with the problem of relevance in the two-person case along lines appropriate to a decision of relevance in the single-person case. But, to proceed in this way, is to assume that the constraints of supervenience have the same force as the constraints of universalizability, in the sense in which (b) above and (SUP) make this clear.

Now, Hare concedes that (SUP) will not commit everyone to accepting (c). The fanatical racist is Hare's prime example of the person who can without offence to rationality escape from the argument. If I understand the fanatic's position correctly, he can reject (c) on grounds that the fulfilment of his ideal requires that nothing be made to turn on the respective strengths of the desires of prescriber and recipient, i.e. that nothing be made to turn on a 'balancing' procedure which works on the basis of desire-strength alone.[85] The difference between him and any other moral agent just is that he is willing to make some radical sacrifices, which include prescribing the frustration of his desires not to be made to suffer, desires which he would come to have, if his position were to change and become like his victim's in all relevant respects. If I read Hare correctly, the fulfilment of his ideal justifies his preference for

satisfying the weaker desire to disadvantage blacks, rather than the stronger desire that he not be disadvantaged, if he were black. We might say that he universalizes independently of how anyone's interests may be affected, including his own interests.[86]

Hare, of course, is saying that his attempt to make the test of relevance work on the basis of (b) above, will work only if ideal-dependent desires are left out of the reckoning.[87] As I judge, it works only for persons who accept (a) and (c), and particularly only for persons whose inclinations will be as Hare supposes in (c). I have in my discussion of the 'hypothetical' bi-lateral argument suggested that this view smacks of an empirical generalization. This assumes a premiss very suited to Hare's test of relevance, one which presupposes the kind of motivation Hare needs to make the test work.

Hare's generalization about what people are likely to want and not to want suggests that the extension of the scope of (SUP) to include a reference to 'anyone else', as this is presented in (IT), may be rejected as illegitimate. As I judge, there is a clear difference in import between (UP) on the one hand and (SUP) and (IT) on the other, which is just the difference between how Hare's moral agent and the fanatic must be thought to be universalizing.[88] If the assumption that people's inclinations will be as Hare supposes in (c), is set aside, the reference to 'anyone else' can be construed only as meaning 'anyone else relevantly similar to A in relevantly similar circumstances', i.e. anyone else who accepts as right what A accepts as right.[89]

Yet, I think the descriptive meaning-rule thesis must work on the unrestricted sense of 'universal', if Hare is to be able consistently to

maintain that legal judgements are not universalizable, and if Hare is to succeed in generalizing the argument he outlines for bi-lateral cases. Hare clearly needs something like (SUP) and (IT) above, if his contention that, in prescribing universally, a prescriber is committed to the view that his prescription applies to anyone else in relevantly similar circumstances, is to have any meaningful import. It is, however, difficult to see how the unrestricted sense of 'universal', may be thought to cover the vast number of cases which an agent's universal principle subsumes. In this Hare's generalization about what people are likely to want, or not to want, plays the crucial role. There is a peculiar anomaly in Hare's account of this matter, which is best brought out by noting his contention that judgements which express the ideal-dependent desires of the fanatic are universalizable in exactly the sense in which judgements which express mere desires, are universalizable.[90] But it is obvious that if the fanatic can ignore (c) without offence to rationality, it follows that the judgement he universalizes, will not be limited in the scope of its application to the restricted sense in which the non-fanatic must be thought to be universalizing. Now, I do not think that judgements which express ideal-dependent desires are universalizable in exactly the sense in which other judgements are universalizable. Hare's contention to the contrary, however, reinforces my view that there are different notions of universalizability in his theory of moral reasoning, viz. the unrestricted sense in which the fanatic universalizes, and a restricted sense in which the non-fanatic must be thought to be universalizing, but which assumes Hare's generalization about what people are likely to want or not to want.

Section 3. From universal prescriptivism to utilitarianism: two problems

In this section I shall maintain that the strategy Hare employs to combine all relevant desires in the person of the decision-maker gives rise to two

problems which undermine his attempt to derive a form of utilitarianism from his universal prescriptivism. The strategy in question is that by which Hare attempts to reduce a conflict of interests between distinct individuals to a conflict of interests in the person of the decision-maker, the resolution of which Hare interprets as demanding a mere prudential calculation. Hare's claim that a prescriber is committed to balancing all relevant desires in his own person depends crucially on the acceptability of this strategy. I think however, that the contention that all relevant desires are a prescriber's 'own' desires, is one central weakness in the argument to the effect that a prescriber is committed to summing and maximally satisfying desires.

What exactly is this weakness? Suppose, again, that A, the judge, is considering whether he ought to imprison B, the thief. Now, in settling his problem with B, as morality demands, A must be willing to adopt a universal principle, which means that he must be willing to accept that the moral judgement it implies, applies equally to himself (1) irrespective of which position he in fact occupies. Hare suggests that if A is willing to concede the moral irrelevance of (1), it follows, in deciding what he morally ought to do, that (2) it is irrelevant whether he is A or B, in which case he is committed to treating all relevant desires in the same way, i.e. on equal par with his own desires. There is, regarding (2), a question about the coherence of the demand that A must be thought to be the person who accepts a judgement <u>in propria persona</u> in the situation in which he imagines <u>himself</u> to be the person in B's position,[91] which raises a question about the acceptability of Hare's strategy for the resolution of inter-personal conflicts.

How does Hare understand the demand that A think of the person in the imagined situation as himself? This is to ask how Hare understands the so-called in propria persona requirement. Hare says:

> '... we shall make the nature of the argument clearer if, when we are asking [someone] to imagine himself in the position of his victim, we phrase our question, never in the form "What would you say, or feel, or think or how would you like it, if you were he?", but always in the form "What do you say (in propria persona) about a hypothetical case in which you are in your victim's position".'[92]

As I understand Hare, the in propria persona requirement demands that A address the question 'What do you say ...?' to himself, as one whose strongest current desire is not to be imprisoned. Hare thinks that his theory would fail to trap A in self-contradiction if the 'What would you say ...?' question is all A need ask. The reason seems to be that this question loses the contact Hare is trying to preserve between desiring, prescribing and acting over a range of situations, including hypothetical situations in which the prescriber has to ascribe to himself different current motivational states, other than those he might actually have for those situations.

Now, as Hare observes, in order to trap A in self-contradiction, it must be the case that he now holds an opinion about the hypothetical situation which is inconsistent with the opinion he now holds about the actual situation.[93] I am in doubt whether making this point, via the demand for role-reversal and imaginative identification, as Hare understands this in terms of the in propria persona requirement, is at all necessary in order to force A into a position of self-contradiction. Consider, for instance,

that A might reason along the following lines. A might say that since B is a thief, he ought to be imprisoned, and offer as reason 'Anyone who is a thief ought to be imprisoned.' This would commit A to conceding that, if he were a thief, he ought to be imprisoned. It is not necessary to fulfil the requirements of role-reversal to make this point. So, what could the point of role-reversal be? One answer, which I have suggested in Section 1, is that role-reversal and imaginative identification are needed to establish identity in all relevant respects between A and B in the actual and imagined situations. These moves, however, raise the question about who the person in the imagined situation might be. But, it may be argued that these moves are not essential in order to trap A in self-contradiction. If it be supposed, for instance, that A believes that thieves ought to be imprisoned, he would, given certain conditions, fall prey to a charge of inconsistency, should he also believe that he ought not to be imprisoned, if he were a thief. Now, this inconsistency is not dependant on role-reversal requirements, and so is not dependent on settling identity-questions.

What Hare takes the point of role-reversal to be links with his conception of what constitutes adopting the moral perspective. Hare standardly offers something along the following lines. If A is to make a <u>moral</u> judgement, he has to subject the judgement he <u>initially</u> offers to the role-reversal test, i.e. he has to imagine himself performing, in turn, the role of each person who would be affected by the action which his initial judgement enjoins, and then determine whether he is willing to assent, <u>in propria persona</u>, to that action being performed in each role in which he ascribes to himself, in turn, all and only the relevant universal features of all concerned persons and their situations. The determination of whether he is willing to assent depends on whether he chooses to perform that action, if he knew he

might be, in turn, the recipients of the proposed action. This is the crucial test of the moral value of an action.

Hare takes care not to identify this interpretation of what constitutes adopting the moral point of view with utilitarianism, for he allows that it is possible, without logical or factual error, to accept non-utilitarian principles, as his 'fanatics' indeed do. There are two incompatible interpretations of what constitutes adopting the moral point of view. On the one hand, the demand that A universalize his prescriptions reads as a demand for an interest-dependent decision, one which A might be willing to accept, if he is also willing to accept that action which he expects will maximize overall desire-satisfaction, on the assumption that he might be, in turn, the persons whose interests would be adversely affected by the action which his initial judgements enjoins. On the other hand, the demand that he universalize his prescriptions reads as a demand for an ideal-dependent decision, supportable by reasons which are independent of how anyone's interests may be affected, such as Hare's 'fanatics' may offer. This is a tension I noted in my discussion of Hare's defence of moral judgements as universalizable prescriptions in Section 2. Interest-dependent reasons turn out, if I am right, to have restricted force, in so far as the scope of the universal prescription which underlies the reason, will be limited to other agents who accept as right what the decision-maker accepts as right. Interest-dependent reasons, on the other hand, are not similarly restricted.

Now, the important assumption about the possibility of imaginative identification in the role-reversal procedure Hare attempts to defend is, as stated above, that A can intelligibly be said to be imagining himself to be performing B's role. There are, however, at least three ways in which

the in propria persona requirement may be understood.[94] In asking the 'What do you say ...?' question, A might be asking one of three questions:

(1) 'What do I say about the hypothetical situation in which I am in my victim's position?'

(2) 'What does B say about the situation in which he stands to me (A) in the relation of recipient to prescriber?'

(3) 'What would I say about the hypothetical situation in which I imagine that I have a different set of universal features?'

(1) demands that A imagine what it would be like for himself to be the person in B's position, as if he (A) were the person who is about to experience what B is about to experience. This may be understood as analogous to A imagining what it would be like to be in some future state of himself, and now desiring what he currently believes he will then want. But, as I judge, (1) demands what is logically impossible, viz. that A can imagine himself to be the person who can have B's experiences.[95] On Hare's view, the 'himself' in question here must refer to A, for in order to trap A in self-contradiction it must be the case that he and the person whose experiences he imagines himself to be having, must be the same person. But how is this possible? It is obvious that the person he identifies as himself and the person he imagines being do not share the same set of universal features. If I am right it is not possible that A can intelligibly treat his representations of what it would be like to be B, as representations of what it would be like for himself to be in B's situation. For the representations in question cannot be the representations of one and the same person. I think it is clear enough that

what I have just said here applies equally to the 'What do I say ...?' and the 'What would I say ...?' questions, i.e. to (1) and (3). If this is right, then in the imaginatively reconstructed situation (2) above must be the question A must be asking.[96] I am suggesting that Hare's strategy for settling interpersonal conflicts suffers from this weakness that the desires A is expected to ascribe to himself in the hypothetical situation, cannot be 'his own' in the sense in which the 'What do you say ...?' question requires. This means that the demand for imaginative identification, as Hare understands this demand in terms of the in propria persona requirement, suffers from incoherence.

A second weakness in the argument that a prescriber is committed to summing and maximally satisfying desires, concerns an assumption of rationality to which Hare appeals. This is the assumption that just as it is rational for an agent to act, prudentially speaking, on what he most desires should be done in his own case, so it is rational for him to act on what he most desires should be done in a situation in which he treats all relevant desires 'as if they were his own'. My problem with this is simply that it is being supposed that an agent, in his role as decision-maker, recognizes a reason to act on a prescription offered by a recipient, and not just that he would recognize such a reason if he were forthwith to be in the position of his recipient.[97] This distinction is crucial. As I judge, Hare has the former in mind, for that is what he needs to run the strategy for settling conflicts between the interests of distinct individuals along lines of a purely prudential calculation as in the single-person case. We may ask: what constrains me to accept that your reasons for wanting something are also reasons for me to want this something?

The prescriptivity thesis says simply that if an agent sincerely assents to a moral judgement, this implies that he sincerely assents to the entailed prescription, i.e. that he is committed to performing a certain act, provided that it is within his power (physically and psychologically) to do what he commits himself to doing. Hare is careful to point out that from this it does not follow that it is logically impossible sincerely to assent to a moral judgement, yet fail to perform the action it enjoins. The point is that sincere agents may be morally weak.[98] But, barring cases of this phenomenon, it is a tautology to say that if an agent sincerely assents to a moral judgement, he is committed to acting on the entailed prescription. It is necessary to point out that for Hare prescriptivity is the logical ground of the claim that inclinations (wants or desires) are relevant to what people may choose to prescribe, simply because their prescriptions express in language what they may want or desire to do. As Hare puts it, to have a desire, in the sense of 'desire' in which any felt disposition to action counts as a desire, is to accept a prescription that it be satisfied, i.e. that something be done.[99]

Now, in Section 2 I argued that universalizability plays a crucial role in committing an agent to Hare's strategy for the resolution of inter-personal conflicts. Universalizability enters the argument via the requirement that A must give equal consideration to all affected persons, and via the demand for role-reversal and imaginative identification, to compel him to give equal consideration to all relevant desires. I think the demand for imaginative identification assumes an appeal to self-interest, i.e. to what an agent might reasonably be supposed to have reasons for wanting, if he were to be, in turn, the persons whose positions he occupies. [100] This appeal is Hare's way of preserving a connection between desiring, prescribing and acting over a range of situations, including hypothetical

situations in which the decision-maker has to consider having different current motivational states, other than those he might actually have for those situations.

On what grounds does Hare think the appeal to self-interest is compelling? Consider, again, what Hare thinks is involved in adopting a universal prescription. On Hare's account of what role-reversal requires, A, the creditor, cannot be said to have reversed roles with B, the debtor, if he remains unmoved by what he would have to endure in his role as recipient. Hare, of course, is here supposing that A will not be apathetic about how his interests would be affected, if his position were to change and become like B's in the relevant respects. The appeal to prudence, in this sense, is essential in supposing that A acknowledges a motivational state which is like B's in all relevant respects. One might say that if he applies the prescription he offers in the actual situation to himself, in his role as recipient of this prescription, he must, as one who is prudent about interests he deems to be 'his own', concede a desire that he should not be made to suffer as B is about to be made to suffer. This, as I have understood Hare, is how the interests of affected parties are accorded a place in a decision-maker's deliberation, and hence receive equal consideration in generating a universal reason which the acceptance of a moral judgement requires.

If A sincerely assents to the universal prescription that debtors ought to be imprisoned, this implies that he is prepared to accept this prescription for every case in which he performs the role of an affected party. Acceptance of the prescription under this condition implies the existence, from his perspective, of a universal desire that debtors should be imprisoned, and hence a universal reason for action which he regards as

binding on himself and anyone else in relevantly similar circumstances. It does not follow from this that other persons in relevantly similar circumstances have reason to act in accordance with the prescription he accepts. It merely follows that he is committed to the view that they have reason to act as required. Now, to say that A has this universal desire means that he desires its satisfaction in each position he occupies, which involves accepting (1) that the fact that some interest is B's and not his will always be irrelevant to a decision about which interests are to be admitted to the calculus, and (2) that anyone's interests generate reasons which rank on equal par with anyone else's, including his own, in support of particular judgements about what ought to be done in the circumstances.

Hare attempts to establish a connection between A's motivation to act as his original judgement enjoins him to act, and his motivation to act on other judgements which he is expected to accept and consider himself as having reason to act on. In this Hare appeals to what he has reason to do in each position he occupies. Now, the contention that A is constrained to think of the person in the hypothetical situation as himself, in the sense in which the in propria persona requirement demands this, is, I think, the key to understanding Hare's notion of moral motivation. As I judge, the appeal to self-interest rests on this contention. The justification of this appeal begins with Hare's conception of what it means to adopt the moral perspective, which derives, if I am right, from his definition of what is involved in prescribing universally that something be done, i.e. from premisses which the logical features of moral language commit an agent to in reasoning morally.

Universalizability, it seems, puts the stamp on what it means to adopt the moral point of view. Hare also sees the demand for universalization as a

criterion or standard of rationality. A cannot be said coherently to be deciding what he ought to do in ignorance of the relevant facts of the situation he is judging, which are facts about 'what it would be like' for himself to be the person in B's position, and how much he would like/dislike having B's experiences, in a 'felt'-sense of 'how much'. This is to say that he has to know the relevant facts of the situation he is judging in a sense of 'know' which assumes a 'felt'-disposition account of 'desire'.[101] But, it is difficult to see what there is in this which requires that a form of rational prudence be adopted in order to arrive at a final decision. Appeal to a form of rational prudence is an essential supplement to what Hare offers as a mere tautology. We may ask: just how does Hare's definition of what constitutes sincere assent to a moral judgement itself commit someone to conceding a motive or reason to act in a certain way? It looks as if Hare is saying that an agent will have reasons to act as morality requires simply in virtue of what it means to use moral language in the way he thinks is conceptually appropriate.

Conclusion

I conclude this chapter with some brief remarks concerning the direction in which Hare's thinking develops in Ethical theory and utilitarianism (ETU) and Moral thinking (MT). Hare sticks to the defence of universalizability which I examined in Section 2, but the argument for utilitarianism is significantly modified. The important change concerns the development of a set of rationality constraints which are implicit in FR, though not fully exploited in the argument for utilitarianism. The renewed argument for utilitarianism runs strongly on an appeal to these constraints. This argument is present, in undeveloped form, in ETU. I shall in Chapter 3 examine some of the moves Hare develops in ETU - particularly an attempt to

give a more prominent role to his rationality constraints - as a preliminary to examining the argument for utilitarianism in MT (Chapter 4).

Notes

1. This formulation follows that given in Chapter 1. In FR Hare operates with a stronger version of the universalizability thesis, the import of which exceeds that of the principle of consistency. I formulate this version in Section 2.

2. In FR Hare does not substantially alter the thesis of prescriptivity. Hare says: 'In the Language of morals I performed what some have thought an evasive manoeuvre by defining "value-judgement" in such a way that if a man did not do what he thought he ought, he could not be using the word evaluatively. I have in this book done something similar with the word "prescriptive" - only with the qualifications mode above.' The most important of these qualifications is that the '... sense of "imply" in which "ought" implies "can" is not that of logical entailment.' Hare explains as follows: '... if we say that somebody ought to do a certain thing, and "ought" has its full (i.e. universally prescriptive) force, then we give our hearers to understand that we think that the question arises to which this is a possible answer, which it would not, unless the person in question were able to do the acts referred to.' I have formulated the prescriptivity thesis with this qualification in mind. See Hare (1980): Freedom and reason. Oxford. Oxford University Press. p. 53-54 and p. 84. All subsequent references are to the 1980 reprint.

3. For Hare's account see Hare (1980) p. 90-93.

4. Hare (1980) p. 93-111.

5. For Hare's account see Hare (1980) p. 112-136.

6. Hare (1980) p. 92 and p. 112.

7. Acceptance of the prescription 'Let me imprison B, if B cannot pay' is conditional. The object of the argument is to establish whether A is willing to accept this prescription universally. Hare draws a parallel between his argument and the 'Popperian theory of scientific method'. Hare says: 'Just as science, seriously pursued, is the search for hypotheses and the testing of them by the attempt to falsify their particular consequences, so morals, as a serious endeavour, consists in the search for principles and the testing of them against particular cases. Any rational activity has its discipline, and this is the discipline of moral thought: to test the moral principles that suggest themselves to us by following out their consequences and seeing whether we can accept them.' See Hare (1980) p. 92.

8. I am here allowing that my initial formulation of the principle of universalizability is adequate. As formulated, it is equivalent to the principle of consistency. I take Hare to be saying that A cannot without inconsistency accept different prescription for situations which are identical in the relevant respects. Hare says: 'The demand for universalization ... compels [A] to make the same moral judgement, whatever it is, about both cases.' See Hare (1980) p. 102.

9. Hare says: 'Because moral judgements have to be universalizable [A] cannot say that he ought to put [B] into prison for debt without committing himself to the view that C, who is ex hypothesi in the same position viz-à-viz himself, ought to put him into prison ...' See Hare (1980) p. 92.

10. Hare (1980) p. 112.

11. Hare says: 'In the example which we have been using, the position was deliberately made simpler by supposing that [A] actually stood to some other person in exactly the same relation as [B] does to him. Such cases are unlikely to arise in practice. But it is not necessary for the force of the argument that [A] should in fact stand in this relation to anyone; it is sufficient that he should consider hypothetically such a case, and see what would be the consequences in it of those moral principles between whose acceptance and rejection he has to decide.' See Hare (1980) p. 93.

12. See note 13 below.

13. Hare's account of the role-reversal test requires that A imagines himself to be the person in B's position, with B's inclination and desires. Hare says: 'All that is essential to [the argument] is that [A] should disregard the fact that he plays the particular role in the situation which he does, without disregarding the inclinations which people have in situations of this sort. In other words, he must be prepared to give weight to B's inclinations and interests as if they were his own.' See Hare (1980) p. 94.

15. Hare (1980) p. 94.

16. By assuming that the cases are identical (in the relevant respects, i.e. in respect of what universal descriptive features they exhibit), '... it follows a fortiori that there are no morally relevant differences'. See Hare (1980) p. 106 and 107. I think the important point to stress is that the cases are identical in respect of the relations between A and C and A and B, and in respect of the desires which characterize these relations. In this I follow H.J. White. See White (1969): An analysis of Hare's application of the thesis of universalizability in his moral arguments. Australasian journal of philosophy, vol. 47, p. 174-175.

17. Here I follow D.L. Norton. Norton says: 'The action which is in itself or in its principle universalized is the product of the conjunction of two factors, namely persons and circumstances.' Norton notes regarding an 'internal disparity in universalizability-criterion formulations' that '[parity] will be achieved by acknowledging relevant differences in persons no less than in circumstances'. He goes on to show that Hare offers a 'revised formulation of the universalizability criterion' which covers this. The revised formulation runs as follows: '[A] has got, not to imagine himself in [B's] situation with his own [A's] likes and dislikes, but to imagine himself in [B's] situation with [B's] likes and dislikes.' Hare (1980) p. 113. See Norton (1980): On an internal disparity in universalizability-criterion formulations. Review of metaphysics, vol. XXIII, p. 519 and p. 521.

18. See note 9 above.

19. Hare says: 'Logic does not prevent me wanting to be put in a gas chamber if a Jew ... " to be put in prison" and "to be put in a gas chamber if a Jew", are not prescriptive expressions; and therefore these things can be wanted without offence to logic. It is, indeed, in the logical possibility of wanting anything (neutrally described) that the "freedom" which is alluded to in my title essentially consists.' See Hare (1980) p. 110.

20. Roxbee Cox argues that (1) does not figure as a component of the 'hypothetical bi-lateral' argument. He says: '... the inclination of the agent to perform the action, if he has such an inclination, is not mentioned as one of the ingredients in this simplest case ...' i.e. the 'debtor-creditor' case. He sees this as inconsistent with the requirements of utilitarianism. He says: 'Any inclination that X may have to a does not enter into the UP reasoning; while for a utilitarian X's inclination will be relevant.' See Cox (1986): From universal prescriptivism to utilitarianism. The philosophical quarterly, vol. 36, p. 3 and 4. My own view is that Cox's reconstruction of the argument is incorrect. I think the 'balancing-procedure' is tacitly assumed in step 5 of the argument. Cox argues later, in his discussion of the 'trumpeter'-case, that the agent's inclination to perform x plays a role, and that the 'balancing'-procedure is assumed in this case. He says, regarding the 'connexion between UP and utilitarianism', that the 'defence of the position starts with the trumpeter example, which [Hare] takes to show (wrongly, I have argued) that the UP approach will proceed by weighing the competing inclinations of those involved ...' See Cox (1969) p. 8-11.

21. See again notes 13 and 14 above.

22. Hare says: '... wanting is like assenting to a singular imperative, not to a moral or other value-judgement.' See Hare (1980) p. 71.

23. See again notes 8, 13 and 16.

24. See again note 9.

25. See again note 13.

26. See Hare (1980) p. 197.

27. See Hare (1980) p. 94.

28. Roxbee Cox says the following: '... Hare contrasts the agent's motivational states with those of the person who would be affected by the action and who does not want to be so treated, and it is natural to include the agent's desire among his motivational states. It is possible, however, that this contrast is mentioned [1] not because the agent's inclination to perform the action is to feature in the reasoning, but because [2] it poses a question: how am I to take account of what I should want if I had the motivation of another person, if my own actual state is so different?' See Cox (1986) p. 5. I agree with Cox regarding [2], though not regarding [1]. My own view is that the 'hypothetical' bi-lateral argument assumes the 'balancing' procedure.

29. See Hare (1980) p. 219. I take Hare's view to be that universalizability requires the role-reversal procedure and that hypothetical situations be treated as actual.

30. I am indebted to K.A. Milkman for this point. Milkman raises this point in his criticism of Hare's article 'Relevance' (Reprinted in Hare (1989): Essays in ethical theory. Clarendon Press. Oxford.) See Milkman (1982): Hare, universalizability, and the problem of relevant descriptions. Canadian journal of philosophy, vol. XII, p. 27.

31. See Hare (1980) p. 90-93.

32. Hare says: '... if [the person making the decision] is ignorant of the material facts (for example about what is likely to happen to a person if one takes out a writ against him), then there is nothing to tie the moral argument to particular choices.' One of the choices he has to make is whether he is willing to allow that the same thing be done to himself in the case in which he performs the other's role. Hare says: '... if the person who faces the moral decision has no

imagination, then even the fact that someone can do the very same thing to him may pass him by.' See Hare (1980) p. 94-95.

33. One consequence of prescribing a line of action universally is that we are thereby committed to the view that the same thing be done to ourselves in any case in which we are in the position of 'the weak, who went to the wall'. Hare regards 'appeal to universalized self-interest' as the 'foundation' of the argument he is presenting. See Hare (1980) p. 105 and p. 194.

34. Hare (1980) p. 97.

35. Hare (1980) p. 112-113.

36. Hare (1980) p. 112-113.

37. See again note 17 above. I agree with Norton that generalizing the argument to cover cases in which the inclinations of the parties differ requires the revised formulation of the principle of universalizability.

38. For the distinction between 'cases' and 'situations' see M. McDermott (1983): Hare's argument for utilitarianism. The philosophical quarterly, vol. 33, p. 389.

39. See again note 16 above. White says: 'The set of qualities defining a person's role, I take as including everything about that person relevant to one's moral reasonings about the situation in question.' See White (1969) p. 175.

40. This point is not explicit in FR. I consulted Hare (1981) Moral thinking. Clarendon Press. Oxford p. 119-121. These pages clearly express how Hare understands the requirements of the role-reversal procedure.

41. Hare (1980) p. 94.

42. Hare (1980) p. 112.

43. Hare (1980) p. 113.

44. Hare (1980) p. 108.

45. Here I follow White. White says: 'My phrase "as his own" is intended to be equivalent to Hare's phrase "in propria persona".' See White (1969) p. 174.

46. I have Norton's revised formulation in mind. See again note 17 above.

47. Hare (1980) p. 113.

48. I think Hare here assumes that moral judgements are 'U-type' valuations, in the sense explained in Chapter 1.

49. Hare (1980) p. 113.

50. Here I follow Cox. Cox says: 'This compromise arrangement might arise in the special case where two agents initially propose an action that will be unwelcome to the other and then engage in bargaining.' He says further: 'The question of what [one] ought to do when [one] has been bargaining is a more specific one, not discussed by Hare.' See Cox (1986) p. 9 and p. 10.

51. Hare (1980) p. 123.

52. See Hare (1980) p. 123 and p. 125.

53. Hare says: 'If, in becoming impartial, [A] became also completely dispassionate and apathetic, and moved as little by other people's interests as by his own, then ... there would be nothing to make him accept or reject one moral principle rather than another.' See Hare (1980) p. 94. This point is more clearly expressed on p. 94. Hare says: '... if he is not a fanatic, and is guided solely by self-interest, together with a willingness to universalize the maxims it suggests, then he will not be willing to prescribe that even he himself, were he to come to dislike the trumpet, should have his dislike of it disregarded by his neighbours.'

54. This is suggested by Cox. Cox says: 'It is in order to solve the problem of how the weighing should be done that Hare gives the argument that leads to a utilitarian position in FR ...' See Cox (1986) p. 10 and note 52 above.

55. Hare (1980) p. 116.

56. Hare (1980) p. 115-116.

57. Hare (1980) p. 117.

58. Hare (1980) p. 117.

59. Hare (1980) p. 118.

60. Cox argues that there is 'a large gap' between the corollary and Hare's attempt to arrive at a utilitarian conclusion. See Cox (1986) p. 11 and note 52 above. I agree with Cox, but, as I read the phrase 'as if they were my own', a decision-maker is expected to combine all relevant desires in his own person and then to treat this as a conflict of interests in his person. White proposes the following: '... we ask that [A] exhibit rationality by proposing that moral judgement which eliminates so far as possible conflicts among the prescriptions he makes in the various roles.' By 'rationality' he means '... the practice of choosing courses of action which best lead to the satisfaction of one's various desires'. Regarding the 'judge-criminal' case, he says that the judge is '... exhibiting rationality, because this action [to imprison the thief] best satisfies the desires of all parties, which he came to think of as in him through his consideration of hypothetical cases.' See White (1969) p. 182.

61. See notes 33 and 53 above. I am here anticipating a move - Hare's appeal to hypothetical self-interest - which I discuss at greater length in Chapters 3 and 4.

62. Hare (1980) p. 120-121. I agree with Don Locke that there is '... no conclusive argument to the PEI [Principle of Equal Interests] from Universal Prescriptivism alone'. Locke notes that in FR there is an argument to support the PEI, but this argument is not '... a consequence of the logic of moral discourse or the nature of moral reasoning as such'. This argument is rather a consequence of '... the nature and function of morality itself' which is to '... provide impartial, mutually acceptable solutions to conflicts of individual interest, solutions which all the parties can accept ...' See Locke (1981): The principle of equal interests. The philosophical review, vol. XC, p. 541 and p. 556. Locke regards the PEI as a substantive moral principle. See Locke (1981) p. 559. A similar point is made by L. Versényi. Versényi argues that the requirement of role-reversal turns on appeal to a '... substantial principle - give weight to others' interests as if they were your own - that for all his disclaimer Hare has introduced by a sleight of hand as if it had been a merely logical one.' See Versényi (1972): Prescription and universalizability. The journal of value inquiry, vol. 6, p. 30. I am inclined to think that Hare appeals to rational prudence in his argument for the PEI, and that the problem is to account for this appeal in reversed-role situations.

63. Hare (1980) p. 123.

64. Hare (1980) p. 11.

65. Hare (1980) p. 13-15.

66. Hare (1980) p. 16.

67. Hare (1980) p. 22-23.

68. Hare (1980) p. 12.

69. Hare (1980) p. 30.

70. Hare (1980) p. 35-36.

71. Hare (1980) p. 36.

72. I think the sense of 'universal' at issue here is that advanced in U. Compare Hare's 'balance of payments' example discussed in Chapter 2.

73. Hare (1980) p. 93. W.G. Lycan charges Hare with having '... avoided the problem of finding criteria of "relevant similarity" by basing his moral-argument paradigm on the agent's ability to imagine himself in an identical case ...' See Lycan: (1969) Hare, Singer and Gewirth on universalizability. The philosophical quarterly, vol. 19, p. 138.

74. Hare (1980) p. 197.

75. This formulation assumes that moral judgements are 'u-type' valuations, i.e. that they make no reference to individual constants.

76. This criterion is implicit in FR. See Hare (1980) p. 107. The crucial point is that any universal feature of an action or situation may be morally relevant provided that we treat it as relevant in any position we occupy in a given set of circumstances. If A treats B's black skin as a reason why he ought to discriminate against B, he has to be prepared to accept that having a black skin in the case in which he performs B's role is also a reason why he ought to be discriminated against.

If he does not, he has not '... met the demand for universalizability, and cannot claim to be putting forward a moral argument at all'.

77. Hare (1980) p. 197.

78. Here I follow Jaegwon Kim. Regarding the 'supervenience of A on B', Kim says that weak supervenience '... only requires that within any possible world there not be two things agreeing in B but diverging in A ...' See Kim (1984): Concepts of supervenience. Philosophy and phenomenological research, vol. XLV, p. 159-160.

79. See again note 30 above. The clearest statement of this point is to be found in Hare (1989): Relevance. Reprinted in Essays in ethical theory. Clarendon Press. Oxford. p. 210. Hare says: 'Having this desire [not to be discriminated against in the case in which he were black], which ex hypothesi is greater that his original desire to do the act, he cannot, if he is to make the same judgement about the actual situation as about this hypothetical similar one ... give greater weight to his own original desire than to his victim's; for to do so would be to give greater weight to the lesser desire among desires which he himself has ...' See also Hare (1980) p. 193.

80. See notes 30 and 79 above.

81. Hare (1980) p. 219.

82. See again note 78 above.

83. Kim judges that Hare's concept of supervenience corresponds 'roughly' to what he calls 'weak supervenience'. See Kim (1989) p. 161.

84. Don Locke notes that the requirement that judgements be made in propria persona presupposes a special sense in which moral judgements are universalizable. This is the sense in which it is always possible '... to apply any moral principle to anyone, at least in a hypothetical case'. Locke says: '... the plain fact is that not all moral principles do apply to everyone. "Wives ought to honour and obey their husbands" does not apply to me, and "Thou shalt not commit adultery", taken strictly, does not apply to anyone who is not married.' See Locke (1968): The trivializability of universalizability. The philosophical review, vol. LXXVII, p. 41.

85. I am assuming that Hare's 'fanatic' is not committed to the PEI. The 'fanatic' may, without offence to rationality, prescribe that the desires he has in the case in which he performs someone else's role, be frustrated. See Hare (1980) p. 110.

86. Hare says: '... he [the fanatic] thinks that, even if the other interests of people (including his own) are sacrificed, the ideal state of society ought to be pursued ...' Hare notes that 'A person who was moved by considerations of self-interest, and was prepared to universalize the judgements based on it, but had no ideals of this fanatical kind, could not think this ...' Hare (1980) p. 161.

87. Hare (1980) p. 160.

88. I think the fanatic's moral judgement is universal in the sense in which moral judgements are 'u-type' valuations. He is committed to the view that his moral principles apply to anyone indifferently of who they are. It is worth noting that Locke charges Hare with failing to notice the difference in meaning between 'universal' and 'u-type'. '"All Locke's are muddle-headed philosophers" is a universal proposition, but it is not u-type because it contains the proper name "Locke"'. See Locke (1968) p. 34 and p. 36. I agree with Locke. On my view (UP) is simply a restatement of the principle of consistency and indistinguishable from Hare's principle of supervenience. (UP) allows reference to particular individuals - (SUP) and (IT) do not.

89. The sense of the word 'universalizable' at issue here is the trivial sense Locke has in mind. Locke says: '... Hare has established that if something ought to be done, then anything to which the same reason applies ought also, ceteris paribus, to be done. He has not established that different people ought in the same situations to do the same thing, for the fact that they are different people might mean that the reasons which apply to them are different ...' See Locke (1968) p. 33-34.

90. Hare (1980) p. 162. Hare says that the fanatic is '... making prescriptive universal judgements, and the only difference between himself and his opponent is that the Nazi sticks to his judgements even when they conflict with his own interest in hypothetical cases (for example the case where he himself is imagined as having the characteristics of Jews)'.

91. For this point I am indebted to C.C.W. Taylor. Taylor remarks that '... it is necessary for Hare that his opponent should be obliged to assent 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 certain cases, however, it is logically impossible for Hare's opponent to be in the victim's position, since anyone who had the characteristics of the victim would be a

different person from the man who is now Hare's opponent'. See Taylor (1965): Review of Freedom and reason. Mind, vol. 74, p. 286.

92. Hare (1980) p. 108.

93. Hare (1980) p. 108.

94. Here again I am indebted to Taylor. Taylor notes that the role-reversal procedure may be understood in different ways: '... (a) imagining that one is some person other than the person one in fact is, (b) "putting oneself in someone else's place", i.e. imagining what one would oneself feel were one in the same situation as someone else is in, or in a similar situation, and (c) imagining what someone else feels in a particular situation.' See Taylor (1965) p. 288.

95. Here again I agree with Taylor. See Taylor (1965) p. 288.

96. A can imagine what it is like to be B (or to be having a similar experience) though not what it would be like for himself to be B. The former, Taylor says, '... is never logically impossible, though it may be practically very difficult or even impossible.' See Taylor (1965) p. 289.

97. Here again I am indebted to Taylor. Taylor observes that Hare's '... way of convincing someone on a moral issue is to show him that he actually has overriding reasons for acting in a certain way, not merely that he would have such reasons were he in some different situation'. See Taylor (1965) p. 287.

98. Hare (1980) p. 67-85.

99. Hare (1980) p. 170.

100. Taylor says the prescriber has to '... regard the suffering of his victim as a reason for him not to do the act which will cause the suffering'. I agree with this view, but I think that Taylor overlooks Hare's appeal to self-interest. See Taylor (1965) p. 289.

101. I am indebted to E.F. Crowell. Regarding the connection between '... wanting and moral reasoning ...' Crowell observes that '... to regard desire as a felt disposition...' makes it more difficult to give an adequate account. See Crowell (1974): Hare on desire and moral reasoning. Theoria, vol. XL, p. 161.

CHAPTER 3

AN ANALYSIS AND CRITIQUE OF 'ETHICAL THEORY AND UTILITARIANISM' (ETU)

Introduction

In this chapter my main concern is to examine the place and function of a set of rationality constraints in the argument Hare advances in support of utilitarianism. Hare's argument appeals to a knowledge requirement of rationality, which embodies a certain view of what it means for a moral agent to have knowledge of someone else's situation. The peculiarity of his view is that it includes certain motivational constraints as a crucial component of the knowledge requirement. As I read Hare in ETU, the force of these constraints depends crucially on the force of the formal constraints of universalizability, and particularly on certain formal restrictions of the role-reversal procedure. I shall endeavour to show how these constraints work, and what is involved in the main thesis Hare advances, viz., that we are qua rational agents, committed to endorsing act-utilitarian solutions to moral disputes. I shall argue that we may reject this thesis on grounds that the knowledge requirement and the motivational constraints which it subsumes do not offer a coherent view of what constitutes a rational decision.

Section 1. An outline of the argument for utilitarianism

In ETU Hare advances the following thesis: A rational agent, i.e. anyone whose reasoning obeys the constraints which the formal/logical features of the moral concepts (universalizability and prescriptivity) impose on his thinking, and who has an adequate grasp of the relevant facts of a situation, will always endorse, as morally correct, that judgement which is required by act-utilitarianism. The suggested contention is that a fully

rational deliberator would not endorse any other judgement as morally correct, unless he has made some logical error or is not apprised of all the relevant and available facts of the situation he is judging.

I begin with a general statement of the argument Hare advances in ETU in support of this thesis. If I say that x ought (morally) to be done in a given situation, I thereby commit myself to the prescription that x be done in all situations which are just like the original (actual) one in the relevant universal respects. Hare thinks that the prescription to which I commit myself expresses, in linguistic form, the desire that x be done.[1] Now, since, to be moral, this desire must be universal, in the sense that I must desire that it be satisfied in all situations in which I occupy (in turn) the positions of the persons who would be affected by its satisfaction, to say that x ought (morally) to be done is to say that I have a desire of overriding strength that x be done in all situations of the given kind. If this desire is really overridingly strong relative to any other desires I may have for what should be done in any of the relevantly similar situations at issue, its satisfaction would be rationally required on grounds that this would maximally satisfy, in sum, all my rational desires, including those I treat 'as if they were my own' in hypothetical reversed-role situations. The significant normative claim, which I shall be contesting in Section 2, is that a fully rational deliberator will, on the basis of full knowledge of the relevant facts of the situation he is judging and a complete understanding of the reasoning to which he is committed in virtue of the formal/logical features of the moral 'ought', endorse the prescription that x be done, and therefore that he will be adopting a position consistent with that required by the preferred form of utilitarianism.

In what follows, I shall attempt to explain what I take Hare's argument to be. My main concern is to show the place and function of Hare's rationality constraints in this argument. I quote at length the passage from ETU in which Hare explains the role of his rationality constraints in committing his agent to act-utilitarianism.

> 'When I do the judging [in the various positions I occupy in hypothetical situations], I have to do it as rationally as possible. This, if I am making a moral judgement, involves prescribing universally; but in prescribing (albeit universally) I cannot, if rational, ignore prudence altogether, but have to universalize this prudence. Put more clearly, this means that, whether I am prescribing in my own interest or in someone else's I must ask, not what I or he does actually at present wish, but what, prudentially speaking, we should wish. It is from this rational point of view (in the prudential sense of 'rational') that I have to give my universal prescriptions. In other words, it is qua rational that I have to judge; and this involves at least judging with a clear and unconfused idea of what I am saying and what the actual consequences of the prescription that I am issuing would be, for myself and others. It also involves, when I am considering the desires of others, considering what they would be if those others were perfectly prudent - i.e. desired what they would desire if they were fully informed and unconfused. Thus morality, at least for the utilitarian, can only be founded on prudence, which has then to be universalized.'[2]

Here I wish to raise and discuss two questions about the view expressed in the passage just quoted. Regarding the claim that in making a moral judgement, I am, qua rational agent, committed to prescribing universally,

it may be asked what is involved in this claim, and what sense to attach to the appeal to prudence. A brief answer to the first of these questions runs along the following lines. We begin our moral thinking by asking what we are committed to in our reasoning by our understanding of the moral concepts we employ whenever we make a moral judgement. Our understanding of the logic of these concepts, particularly of the moral 'ought', brings with it an understanding of the requirement, in moral reasoning, to prescribe universally for all situations of a given kind. The important constraint to which our reasoning is subject, is that our prescriptions have to apply to all situations of the kind in question, including situations in which we are in the less advantageous position which our recipients in fact occupy. We are, because our prescriptions have to be universal, prevented from prescribing differently for the situations in which we occupy these positions. As I read Hare, the significant point is simply that since we have to decide which prescription we are willing to universalize (1) in ignorance of which positions we in fact occupy, and (2) in ignorance of who is actually who in each of these positions - i.e. in ignorance of any information about particulars which enable us to identify particular roles as our own - that we have to allow that it is equally probable that any of the situations at issue might be the actual situation, and that it is equally probable that we might actually be any of the persons whose positions we occupy qua affected parties.[3] It is because of this that, in deciding what we morally ought to do, we cannot, without offence to the constraints of universalizability, accord our own interests extra or special consideration because they are our own, on grounds that we are the particular individuals whom we in fact are, or on grounds that we occupy the positions which we do in fact occupy. If Hare is right, this means that we will have to decide in the interests of everyone who would be affected by our choice of universal prescription, and that, in making this choice,

we will be willing to accord equal consideration to the interests of all concerned, i.e. the same consideration as we accord our own interests.

It is a major contention of Hare's argument for utilitarianism that we are, qua rational agents, committed to treating all affected persons as deserving of equal consideration. This contention is part of what is involved in the claim that the requirement to universalize our prescriptions, is a requirement of our rationality. But much more is involved here. In coming to a decision about which prescription we are prepared to universalize, we need to know the facts of other people's situations. Hare sees this requirement to know the facts as a logical requirement stemming from the requirement to universalize our prescriptions. He offers an account of how these facts are relevant to our decisions. Briefly, we cannot be said to know what it is we are doing, should we universalize any one prescription from the set of prescriptions which make up the alternatives available to us, in ignorance of what its acceptance entails for ourselves and others.[4] This knowledge of what acceptance of a universal prescription entails depends crucially on what we want for ourselves and what others want for themselves. The facts of our own and other people's situations to which we should be attending are facts about how we and others would be affected by its acceptance, particularly facts about how much we and others would like/dislike the effects of accepting this prescription.

It is important to note that Hare has in mind a 'felt' sense of 'how much', or 'how strongly' we would like/dislike the effects of universalizing a prescription.[5] This sense of 'how much' or 'how strongly' works from the perspective of the first person, on the basis on an appeal to the imagination. My knowledge of the facts of your situation is acquired by

imagining 'what it would be like' for myself to be in your precise situation. But the decision I make in the hypothetical reversed-role situation regarding what you want for yourself, is a decision I construe as affecting myself, i.e. the decision is made in propria persona, on the basis of the supposition that I might actually be the person in that situation.

As I understand Hare, the requirement to universalize my prescriptions brings with it the requirement that I must know the facts about your situation. We might put the point at issue here in terms of what it means to have a universal desire, and say that since, to be moral, a desire has to be universal in the sense that I must desire, on balance, that it be satisfied in all situations of the given kind, the requirement to universalize my prescriptions cannot be met in ignorance of what things are desired, and how strongly they are desired. The point is simply that since the universal prescription which I am prepared to accept expresses, in linguistic form, a desire strong enough to override all other competing desires for what should be done in the various situations for which I am prescribing, I must consider, as counting among the relevant facts of my own and your situations, the strengths of all competing desires.

There is, however, an important restriction on which desires should be admitted for consideration. As Hare makes clear in the passage quoted above, we are, in prescribing universally, required to have regard for the constraints of prudence.[6] To be prudent is to want to satisfy only those desires which best promote our own interests to the greatest extent possible. In ETU Hare does not offer an account of prudence,[7] but it is possible to interpret the request for prudence as making a demand on an agent's motivation, which I shall treat as the demand that he must have a

desire of overriding strength to do only that which best serves his own interests over time.[8]

To see just how the requirement to universalize our prescriptions bears on the request for prudence, it is necessary to point out that the argument for utilitarianism turns upon the application of universalizability to prudential reasoning.[9] As I understand, in reversed-role situations the request for prudence translates to an appeal to hypothetical self-interest.[10] If the motivational interpretation of the request for prudence is correct, we are required, before we are allowed to universalize our prescriptions, to criticize our own and other people's desires by asking whether they are the ones we would want to satisfy, given that we are, qua rational agents, motivated to satisfy our own interests to the greatest extent possible. The basic idea is simply that just as I decide what I, qua prudent, most desire should happen in my own case, so I decide, from the perspective of first-personal rationality (i.e. prudential thinking), in the case in which I am in your precise situation, what I most desire should happen to myself in that situation. And, just as I decide, qua prudent, what is best for myself in my own case, so I decide, in the case in which I deem myself to be one who has incompatible desires, which include desires I treat 'as if they were my own', what is best for myself overall. The significance of criticism from the perspective of prudence is simply that we are expected to identify only with the prudent desires of others, and to admit only our own and other people's prudent desires for consideration at the time we choose the prescription we think we can universally endorse.[11] For only these desires have a claim, if satisfied, to what is best for ourselves and others, and hence only they give the alternatives between which we are to choose. This is part of what is meant by saying that we are required to know what it is we are doing when we universalize our

prescriptions. The idea that there is a place for prudential calculations in making a decision to universalize a prescription depends crucially, as suggested, on an appeal to universalizability, which is needed to constrain prudence to the demands which rational decision-making in morals imposes on agents.

The points raised above concerning the sense in which the requirement to universalize my prescriptions is a requirement of my rationality, imply that I am fully rational

(1) if I reason in accordance with the constraints which the formal/logical features of the moral 'ought' impose on my thinking, and make no logical errors in deciding which universal prescription to endorse;
(2) if I decide what I morally ought to do in the light of all relevant (and available) information, and admit only prudent desires for consideration at the time I make this decision.

Given the general claim that reasoning in accordance with (1) and (2) leads us to accept moral judgements consistent with those required by act-utilitarianism, we might express Hare's notion of what constitutes a rational course of action in terms of the maximization of desire-satisfaction. A course of action is rational if and only if it maximizes the satisfaction, in sum, of the (rational) desires of all concerned parties. A rational course of action will also be the morally right course of action if and only if there is no other course of action which - according to the available evidence - is likely to yield a greater net sum of desire-satisfaction for all concerned.

I indicated above that the requirement that I must know the facts about the situation I am judging is a logical requirement which stems from the

requirement to universalize my prescriptions. I shall now endeavour to show in what sense of 'know' I am required to know these facts. Following Hare, to say that I am acquainted with the relevant facts of the situation I am judging, i.e. that I know what it is I am doing should I universalize a prescription for this situation, turns on a notion of knowing 'what it would be like' for myself to be at the receiving end of this prescription. For Hare this knowledge derives from the imagined experience of 'what it would be like'. I shall call this request for an imagined experience the knowledge requirement of rationality. (KR)

Two things about KR are here worth noting. First, KR can be satisfied only by appeal to the demand for imaginative role-reversal, which involves imaginative identification with the prudent desires of all affected parties. These are desires I ascribe to myself in hypothetical reversed-role situations, and which I accept as desires I would have, were I actually to be in those situations. According to KR, my present full knowledge of your desires, were I to be in your precise situation, entails my now having the same desires as you have for what should be done in the hypothetical situation in question.[12] I do not fully know these hypothesized desires unless I accept not only that I would have them, but also 'what it would be like' to have them. As I understand Hare's notion of 'what it would be like', I am expected to endorse your prudent desires as 'my own', which involves wanting for myself, in the hypothetical situation in which I occupy your position, what you want should happen to yourself. Add to this the contention that I have to allow, as counting among the relevant facts of your situation, the relevance of facts about experiences I would have, were I to be in your position, particularly facts about how much I would like/dislike having these experiences (albeit only imagined experiences treated as actual), satisfying KR would seem to be a matter of adopting

your concern for the satisfaction of your desires.[13] As I judge, a strong sense of this concern is required, one akin to the sense - appropriate in the sphere of prudence - in which a single person experiences concern for the satisfaction of his desires.[14] One significant point which I wish to stress is that, since I am required to identify only with your prudent desires, KR would seem to bear only on these desires, which means that KR cannot fully be satisfied without the appeal to prudence. With this point in mind, I formulate KR as follows.

KR: If I know that you desire x, as one who is perfectly prudent as far as his own interest goes, this entails that I desire x in the hypothetical situation in which I am in your precise situation.

Second, KR itself entails a claim which I shall call the suffering thesis (ST). If it is correct to say that Hare's theory requires a strong reading of KR, as suggested, and supposing that my experiences, in the situation in which I occupy your position, are experiences of suffering, then the suggested claim which KR entails might be formulated as follows:

ST: I do not fully know what it would be like for myself to be in your position, unless I am equally averse to myself suffering as you are suffering, or are about to be made to suffer.[15]

As I judge, ST implies that there is a requirement to the effect that the things other people have good prudential reasons for wanting are, or must be construed as, reasons for a prescriber to want the same things. Hare's notion of knowing 'what it would be like' supplies an answer to the question, 'Why are the things you have good prudential reasons for wanting also reasons for me to want the same things?' This is that I might actually

be any of the persons whose positions I occupy in hypothetical situations, and that the quality of my experiences in these situations is itself a reason for me to want what my recipients want.

I said above that KR can be satisfied only by appeal to the demand for role-reversal. It is worth noting that KR requires a particular account of what the demand for role-reversal involves. There are two significant moves in this account. Hare clearly indicates which desires shall count as relevant in hypothetical reversed-role situations, and he amplifies some suggestions offered in FR which show how he sees the relations between prescribers and recipients.[16] I begin with the distinction Hare draws between an agent's role qua author of a moral decision and the various roles he performs qua recipient of the decision. Qua author, I am expected to disregard my own actual (current and future) desires for what should happen to myself in a hypothetical situation, and consider only the desires of the person whose role I perform in this situation. My own desires are considered only in my role qua affected party, which is simply my own actual role so understood. The important stipulation is that all desires which do not pertain directly to what a recipient wants should happen to himself are irrelevant to a prescriber's representation of 'what it would be like' to be in his position. I quote at length the passage from ETU in which Hare presents this view. Where Hare refers to a desire, a liking or a wish, this must be understood as a desire, liking or wish of one who is perfectly prudent. Hare says:

> '... when we speak of the "situations" of the various parties, we have to include in the situations all the desires, likings etc. that the people have in them - that is to say, I am to do for the others what I wish to be done for me were I to have their likings etc. and not those

which I now have. And, similarly, I am not to take into account (when I ask what I wish should be done to me in a certain situation) my own present desires, likings etc. There is one exception to this: I have said that one of the situations that I have to consider is my own present situation, I have to love my neighbour as, but no more than and no less than, myself, and likewise to do to others as I wish them to do to me. Therefore, just as, when I am considering what I wish to be done to me were I in x's situation, where x is somebody else, I have to think of the situation as including his desires, likings etc. and discount my own, so, in the single case where x is myself, I have to take into account my desires, likings etc. In other words, qua author of the moral decision I have to discount my own desires, etc., and consider only the desires, etc., of the affected party; but where (as normally) I am one of the affected parties, I have to consider my own desires, etc., qua affected party, on equal terms with those of all other affected parties.

It will be asked: if we strip me, qua author of the moral decision, of all desires and likings, how is it determined what decision I shall come to? The answer is that it is determined by the desires and likings of those whom I take into account as affected parties (including, as I said, myself, but only qua affected party and not qua author). I am to ask, indeed, what I do wish should be done for me, were I in their situations; but were I in their situations, I should have their desires etc., so I must forget about my own present desires (with the exception just made) and consider only the desires which they have; and if I do this, what I do wish for will be the satisfaction of those desires; that, therefore, is what I shall prescribe, so far as is possible.'[17]

It seems that the point of separating an agent's role qua author from the various roles he performs qua recipient, is to put him in a position in which his role, qua author, cannot influence his decision about what he morally ought to do. The position Hare's agent occupies, qua author, is in this respect similar to that explained in FR. Qua author, he may know which role is the actual one in which he desires that a certain thing be done, but this knowledge gives him no privileged position vis-à-vis his recipients, i.e. his own desires get no extra or special consideration because they are his in this role. The important constraint operative here, which derives from the demand that he has to bracket or put aside all information about particulars which enable him to identify a particular role as his own, is that all affected persons (among whom he counts himself), are deserving of equal treatment.

It is worth noting that qua author of the moral decision he is expected to adopt a position of complete impartiality.[18] For Hare impartiality is secured by the requirement that a universal prescription has to apply to all situations which are exactly or relevantly alike, including hypothetical situations in which the role of the parties concerned are precisely reversed. Put more clearly, the request for impartiality draws on the constraint that whatever line of action he chooses to endorse universally, the same thing should be done to himself in any of the various positions he occupies qua affected party. And since he is expected to prescribe in full awareness of the relevant facts of the situations of the persons who actually occupy these positions (which are, for him, facts about experiences he would have, were he to be in their positions, particularly facts about the quality of these experiences, i.e. how much he would like/dislike having them), it seems that he must be prepared to be impartial in the formal sense which Hare's theory requires. This is an

impartiality with respect to the content of all relevant rational desires. Briefly, he is expected

(1) to make no judgement about the intrinsic worth (moral legitimacy) of the relevant desires;

(2) to make no judgement - initially - about the rationality of these desires;

(3) to make no judgement about their (psychological) causes, even if such a judgement affects the issue about the rationality of these desires.

The important requirement of the reversibility test is, however, the one which demands that he must bracket or put aside his own actual desires for what should be done in the hypothetical situations in which he occupies his recipients' positions, and identify only with their (prudent) desires. How is this requirement to be accounted for? Hare's explanation touches on the requirement that hypothetical situations must be identical (in the relevant respects) to actual situations. This requirement, which derives from universalizability, has the effect of restricting the desires, likes/dislikes to be counted as relevant in a reversed-role situation to a recipient's desires, likes/dislikes. I think this restriction is a necessary condition of the identity which Hare's theory requires, which is simply identity with respect to the relations between a prescriber and his recipients, and with respect to the desires, likes/dislikes which characterize these relations. The upshot is simply that the desires, likes/dislikes which a prescriber is expected to ascribe to himself in a reversed-role situation must be identical to his recipient's for otherwise the claim that he would not want what his recipient in fact wants must

dictate that the actual and imagined situations are not identical in the relevant respects.

It is not difficult to see that to violate the requirements of the reversibility test is also to violate Hare's rationality constraints. This is obvious enough, if it is correct to say that these requirements derive from universalizability, and that reasoning in accordance with the logic of this feature of the moral 'ought' is a requirement of my rationality. As I understand, the requirement that I, qua author of the moral decision, must suppress all information about particulars as irrelevant to the moral decision, constrains me to allow - since I have no way of telling which situation is the actual one in which I desire that a certain thing to be done - that any of the situations at issue might be the actual situation in which I am the affected party. This is necessary if I am to be constrained to accept - as a step towards according the same consideration to every one who would be affected by my decision - that whatever line of action I wish universally to endorse, the same thing be done to myself in any of the various positions I occupy qua affected party. And the requirement that I, qua author of the moral decision, must, in reversing roles with you, put aside all knowledge of my own actual desires, compels me to consider, as counting among the relevant facts of your situation, the relevance of facts about experiences I would have, were I actually to have your desires for what should happen in that situation. I am assuming that it is correct to say that I have not reversed roles with you unless I want for myself in the hypothetical situation what you want for yourself in the actual situation, and that an adequate representation of 'what it would be like' for myself consists, not in having an aversion to you being subjected to an experience you would rather not have, but rather in having an aversion to myself being subjected to the experience. This way of putting the matter is, of course,

consistent with the in propria persona requirement, which is an essential feature of Hare's argument for utilitarianism.

With these points in mind, I am in a position to formulate a specific requirement of my rationality. Any prescription which I endorse as applying to myself in a hypothetical reversed-role situation, must, to be rational, be endorsed in the light of the relevant facts of this situation.[19] Rational endorsement proceeds on the assumption that I know these facts, in the required sense of 'know'. Allow, for the moment, that the desires with which I am expected to identify must be only those desires of my recipient which can be rationally defended, as desires he has good prudential reasons to want to satisfy (or would have, if he were perfectly prudent). Now, one condition of saying that I know what these desires are, is that I must identify with them. If I genuinely identify with these desires, I must assent - initially at least - to the prescriptions which are their expression in linguistic form.[20] I formulate this request for assent as a requirement of my rationality.

R: Rationality requires that I must assent - initially at least - only to those prescriptions which express in linguistic form, the dominant rational desires I would have in each situation in which my position were to be different.

One implication of R is worth noting here. In ETU Hare casts his argument for utilitarianism in terms of desires, for this he thinks is required by his prescriptivity thesis. Desires, says Hare, are in the generic sense in which any felt disposition to action counts as a desire, 'assents to prescriptions'.[21] This move helps Hare to establish a connection between his rationality constraints and the prescriptivity thesis. We might say that not to assent to my recipient's (prudent) prescriptions implies that I do

not know the relevant facts of the situation I am judging, contrary to what KR and ST require, in which case my acceptance of any other prescription in the hypothetical situation in which I occupy his position would run counter to R, and therefore be irrational.

I have above indicated that there are similarities between FR and ETU regarding Hare's treatment of the role his agent performs qua author of a moral decision. What, then, are the differences? It is worth recalling that in FR it is possible for an agent to prescribe the frustration of his recipient's desires purely on grounds of an ideal which favours the satisfaction of an ideal-dependent desire over any other desire. Hare's problem in this regard seems to have been that the constraints of universalizability and prescriptivity cannot compel the 'fanatic' to grant other person's desires the same consideration as he grants his ideal-dependent desires. For, as Hare maintained in FR, ideal-dependent desires have a universalizability of their own,[22] i.e. they qualify as moral simply because there is always something about an ideal which makes it legitimate for the 'fanatic' to override other competing desires. Indeed, Hare seems to have conceded that applying a maximizing desire-satisfaction rule to a situation in which an ideal-dependent desire is a feature, is always likely to yield the prescription that it should be satisfied, simply because the 'fanatic's' commitment to the ideal allows his concern for the satisfaction of his ideal-dependent desire to take precedence over his concern for the satisfaction of any other desires. This is in line with a view, expressed in FR, that there is nothing logically amiss with an agent prescribing the frustration of desires he would have, if his position were to change and become like his recipient's in all relevant respects.[23] In ETU, however, this is no longer possible. What accounts for this?

Hare's explanation touches on his appeal to the rationality constraints. I said that it is not possible to violate the requirements of the reversibility test without also violating these constraints. The point raised above concerning the separation of an agent's roles qua author and qua recipient, which I said gives an agent, qua author, no privileged position vis-à-vis his recipients, is intended to prevent the 'fanatic' from taking his commitment to an ideal into account in his role, qua author, of the moral decision.[24] As I judge, Hare thinks that he is, in effect, being prevented, qua author, from ranking the value to himself of satisfying the various desires in the calculus in terms of the value he places on the pursuit of the ideal. Hare expresses this constraint as the demand that, qua author, the 'fanatic' must treat his ideal-dependent desire simply as one desire among others, i.e. as a desire which has an equal, initial claim to satisfaction, along with all other desires in the calculus.[25] Moreover, since the 'fanatic' has to make his decisions in ignorance of any information about particulars which enable him to identify a particular role as his own, and on the basis of the supposition that he might actually be any of the persons whose roles he performs in hypothetical situations, he would, if Hare is right, be constrained by this supposed fact, and the requirement that he has to know the relevant facts of the situations he is judging, in the required sense of 'know', to grant positive weight to desires he is expected to ascribe to himself in these situations.[26] As I have understood Hare, he can be said to know the facts of the situations he is judging if and only if he is willing to prescribe - albeit only initially - that these desires be satisfied. Add to this requirement of his rationality the request for prudence, and the constraint to which he is subject might be expressed as the demand that he must have regard - prudentially speaking - for the (prudent) desires of others, and hence prescribe their satisfaction strictly in proportion to their

strengths. The suggested point is that he cannot, without offence to his rationality, accord his ideal-dependent desires any extra weight because of the value he places on its satisfaction, which is to say that he cannot, qua rational agent, disregard the principle of equal interests.[27] I shall shortly present Hare's argument for this principle. Here I merely note Hare's confidence that the constraints just considered are sufficient to block the kind of 'fanaticism' which presented a problem in FR.[28] Hare concedes, however, that another form of it may yet survive. It might still be possible for the 'fanatic' to override the balance of other competing desires in the calculus, but only if his ideal is pursued with overriding intensity. On the argument for utilitarianism which Hare presents in ETU, this ideal-dependent desire ought to be satisfied. But, though this result is theoretically possible, it is unlikely to present practical problems for the simple reason that this sort of case is unlikely to arise in practice.

I am, in effect, maintaining that the force of the rationality constraints mark a difference between FR and ETU. In FR, Hare's concession that utilitarianism does not cover the whole of morality is a consequence of the problem which the 'fanatic' presented. As Hare argued in FR, a deliberator may, on his method of reasoning, without logical or factual error (i.e. without offence to rationality), accept moral judgements inconsistent with those required by utilitarianism. In ETU, however, this is not possible. The crucial move, as suggested above, is the one which binds a deliberator, in virtue of his rationality, to the principle of equal interests.

It is now time to state Hare's argument for act-utilitarianism. This argument proceeds via an attempt to derive the so-called principle of equal interests (PEI) from universalizability, prescriptivity and the

rationality constraints. I begin by stating the argument for this principle, in summary form, in Hare's own words.

> 'In this position [i.e. in the position I occupy qua author of the moral decision] I am prescribing universally for all situations just like the one I am considering; and thus for all such situations, whatever role, among those situations, I might myself occupy. I shall therefore give equal weight to the equal interests of the occupants of all the roles in the situation; and, since any of these occupants might be myself, this weight will be positive.'[29]

Now, to comply with the PEI is, as Hare suggests in the passage just quoted, to prescribe that the dominant rational desires I would have in each of the various positions I occupy qua affected party, should be satisfied, strictly in proportion to their strengths. This means that I have to give equal weight to the equal interests of all concerned parties, or equal weight, strength for strength, to all relevant rational desires. The basic line of reasoning in Hare's defence of this principle might briefly be presented as follows. There are three main points.

First, in any conflict situation all relevant rational desires are deserving of equal, initial consideration. The argument for this is by now familiar. I have to allow, since I am constrained to make a decision concerning what prescription to universalize in ignorance (1) of the fact that I play the role which I do in fact play, and (2) of who is actually who in each of the various positions I occupy - i.e. in ignorance of any information about particulars which enable me to identify a particular case as my own - that all cases which have been put up for consideration are deserving of equal consideration, i.e. the same consideration I accord to

my own case. There is, as I have suggested above, a crucial contention in this brief statement of the argument. I have to suppose, since I have no way of telling which situation is the actual situation in which I desire that a certain thing be done, that it is equally likely that any of the situations at issue might be the actual situation, and that it is equally likely that I might actually be any of the persons who occupy the various positions in these situations. These suppositions, which Hare takes to be justified with reference to the elimination of all irrelevant information about particulars, constrains me to accept that all affected persons, among whom I count myself, are deserving of equal treatment. If we construe this constraint in terms of how Hare understands the identification requirement, then the point is simply that I must treat all relevant rational desires as equally worthy of satisfaction, albeit only initially.

Second, I cannot rationally decide what the morally correct course of action would be in ignorance of the relevant facts about my recipients' situations. This demand for correct factual information, which is the by now familiar demand to know 'what it would be like' for myself to be in their situations, can be satisfied only by appeal to the role-reversal procedure. If the suggested interpretation of what this procedure involves is correct, as I have amplified it with reference to the rationality constraints, I am, in reversing roles with you, committed to fulfilling a condition of rational assent to your prescriptions. This is, as KR and ST imply, that I am not in possession of correct factual information about your case, unless I myself want, were I to be in your situation with your desires, that these desires should be satisfied.[30] As I understand, I cannot rationally deny you the satisfaction of your desires, unless I know what it is I am doing. Knowing this involves knowing what it is like for you to be experiencing the frustration of your desires. Suppose that this is causing

you to suffer. Hence, knowing what it is I am doing involves knowing what it is like for you to be suffering. I cannot, however, know what it is like for you to be suffering without knowing that this involves your wanting the suffering to stop. Now, since the facts of your situation to which I should be attending are facts about experiences I would have, particularly facts about how much I would dislike having your experiences, and since an adequate representation of 'what it would be like' for myself consists in having an aversion to myself suffering as you are suffering, I am committed to prescribing that, were I in your situation with your desires, my suffering should stop. So, putting myself in your position involves accepting the prescription 'Let me not be denied relief from my suffering', as my rationality requires.

As I understand the argument so far, by entertaining the thought (1) that some state of another person might actually be a state of myself, I thereby (2) acquire a concern for the satisfaction of desires I would have, were I to be in his precise situation. (1) draws on the supposition that this situation might be the actual situation in which I am the affected party, which derives from the demand for universalization. If construed in terms of how Hare understands the identification requirement, I am, in my role qua affected party in this situation expected to identify with the (rational) desires of the person who is actually the affected party. Now, a full account of the justification of this move, and of (2), involves an account of how the facts of his situation are relevant to the moral decision. Briefly, what he wants should happen to himself, and how strongly he desires this, are relevant because these things affect what universal prescription I think I can rationally endorse. Rational endorsement proceeds on the assumption that I know these things, in a specific sense of 'know'. I have to know 'what it would be like' for myself to be in his

precise situation - i.e. how much I would dislike having his experieces, in a 'felt' sense of 'how much' - in order to know what it is I would be doing should I universalize a particular prescription, for this prescription would otherwise be faulted for lack of information. If I am correct in maintaining that Hare's theory requires a strong reading of KR, one in which the sense of my concern for the satisfaction of your desires is akin to the sense, appropriate in the sphere of prudence, in which a single person experiences concern for the satisfaction of his desires, then my initial assent to the prescription 'Let me not be denied relief from my suffering' is a requirement of my rationality in the prudential sense which Hare's theory requires. The significant points to note are, I think, that assent to this prescription presupposes that I know fully the relevant facts of your situation, and that the requirement to universalize my prescriptions constrains me to have regard, prudentially speaking, for desires I am expected to deem 'my own' in hypothetical reversed-role situations.

The latter point brings me to the third leg of the argument for the PEI. This concerns Hare's appeal to the requirement of prudence.[31] In hypothetical reversed-role situations, this appeal translates to an appeal to <u>hypothetical</u> self-interest,[32] and affects my treatment of desires I am expected to deem 'my own' in these situations. As I understand, I have to decide, on the basis of the supposition that any of these situations might be the actual situation in which I am the affected party, what I want should happen to myself in each situation, as one who is perfectly prudent about interests I deem to be 'my own'. It is obvious, as I have suggested above, that this supposition, taken together with the requirement that I must know the relevant facts of the situations I am judging, in the required sense of 'know', functions as a constraint on what universal

prescription I think I can rationally endorse. It is, however, equally obvious that without this supposition Hare's appeal to self-interest cannot be extended to cover hypothetical situations in which I am expected to ascribe to myself motivational states other than those I might in fact have for what should happen to myself in them, and that it goes some way towards explaining how prudence might be constrained to the demands which rational decision making in morals impose on agents, at least as Hare conceives of these demands. If, then, we grant Hare this supposition, we might allow Hare the premiss that, to the extent I think of the motivational states of another person as motivational states of myself, to that extent I acquire a concern for the satisfaction of his desires, and hence am willing to prescribe, at least initially, that these desires should be satisfied, as my rationality requires. Then, as one who is prudent as far as his own interest goes, which includes interests I deem to be 'my own' in hypothetical reversed-role situations, I would give (positive) weight to all relevant rational desires, and strictly in proportion to their strengths, just as the persons who are actually in these situations would, if they were perfectly prudent.

Hare's theory requires that a prudential calculation be made in each position I occupy qua affected party. If, in each of these positions, I prescribe as one who is perfectly prudent, I must give positive weight only to those desires that can be rationally defended, as desires each affected party has good prudential reasons to want to satisfy. If I understand Hare correctly, two applications of prudence, at different levels, are needed to bring me to a final decision, first to determine the criticized desires of my recipients, which then become the desires with which I am expected to identify, and then to determine what, in the end, all things considered, I ought rationally to prescribe, which involves only the sub-set of

criticized desires.[33] I shall have more to say about the second application later on. It is worth noting that only the desires in this subset are to be accorded an equal, initial claim to satisfaction, strictly in proportion to their strengths, as the PEI requires.

I am now in a position to pull a few strings together. To decide, qua rational agent, what I morally ought to do, I have to be prepared to accept a universal prescription for all relevantly similar situations. Since my prescription has to be universal, I cannot, in choosing the one I think I can endorse, accord my own interests any special or extra weight, but have to choose in the interests of everyone who would be affected, considered impartially. This involves choosing on the basis of the contention that I have no reasons, which satisfy the constraints of universalizability, to treat other people's rational desires differently from my own. The test of this is simply whether I am willing to treat these desires as all equally worthy of satisfaction, and whether I am willing to assent, at least initially, to the prescriptions which express these desires in linguistic form. I am, however, expected to choose qua rationally self-interested person, i.e. as one who wants (or is motivated by a desire of overriding strength) to satisfy his own interests to the greatest extent possible. If my final decision is ultimately determined by an application of universalizability to prudential reasoning, I have to prescribe, on the basis of the supposition that it is equally probable that any of the situations at issue might be the actual situation in which I am the affected party, that the rational desires I would have in each of these situations, should be satisfied strictly in proportion to their strengths. Regarding the treatment of these desires, I am, qua rational decision-maker, in a position which is structurally similar to a single-person's treatment of his rational desires. This is to say that I am, in a way

which parallels his treatment of these desires, in the two-person or multi-person case, committed to the principle which bids that desires of equal strength must count equally.[34]

Before proceeding to the final stages of the argument for utilitarianism, I wish to note two points concerning Hare's contention that all the moves he needs to make in his argument for the PEI derive from a 'consideration of the logic of the moral concepts alone'.[35] In my reconstruction of the argument for the PEI, I have accordingly taken the view that the appeal to prudence in hypothetical reversed-role situations rests on the constraint, which derives from universalizability, that it is equally probable that any of these situations might be the actual situation in which I am the affected party. It is, however, not difficult to see that in this I am assumed to be motivated by self-interest. If this is correct, the argument for the PEI must be thought to <u>begin</u> with the appeal to prudence, i.e. that this appeal logically precedes all other moves in this argument.[36] The second point concerns an assumption implicit in the conditions for arriving at a universal prescription. I have above indicated that two applications, at different levels, of the requirement of prudence are needed to bring me to a decision about what I ought, all things considered, rationally to prescribe. Now, following Hare, a rational desire to do something qualifies as moral if and only if the desire is a universal desire, i.e. if and only if I desire, on balance, that that action be done in all positions I occupy <u>qua</u> affected party. I arrive at this on balance desire <u>via</u> two applications of the request for prudence. The second application, which parallels the single-person case, will affirm that the action ought morally to be done if and only if its performance is likely (on the available evidence) to secure the greatest possible sum of desire-satisfaction for all concerned. It is,

however, not difficult to see that in this the appeal to prudence must be thought to have overriding force.

Now, the two points just made concerning the place and function of the appeal to prudence in Hare's argument for the PEI supply the clearest evidence that this appeal exceeds what might reasonably be supposed can be justified with reference to the formal constraints of Hare's theory alone. It is clear enough that appeal to universalizability is needed to constrain prudence to the demands which rational decision-making in morals impose on agents, at least as Hare conceives on these demands. But it is equally clear that the justification of this appeal rests on the unproblematic contention, which might be posited at the outset of the argument, that it is rational to be prudent. I think it is correct to say that the request for identification with only the prudent desires of my recipients rests squarely on the prior assumption of prudence, and its overridingness, in the argument for the PEI. Hare might argue for these things as necessary additions to the method of reasoning provided by the logic of the moral concepts. However, it seems to me that such a line of reasoning would alter the overall structure of Hare's argument for the PEI, and the nature of the project he claims to be pursuing, viz., to set up an argument for utilitarianism which draws only on premisses derivable from 'the meanings of the moral words'.

Now, just how does being committed to the PEI lead to utilitarianism? Hare has the following to say:

> 'If I am trying to give equal weight to the equal interests of all parties in a situation, I must, it seems, regard a benefit or harm done to one party as of equal value or disvalue to an equal benefit or harm

> done to any other party. This seems to mean that I shall promote the interests of the parties most, while giving equal weight to them all, if I maximize the total benefits over the entire population, and this is the classical principle of utility.'[37]

Hare, however, points out that his theory does not work in terms of maximizing benefits (understood as including the reduction of harms).[38] Being committed to the PEI should be understood as committing an agent to the following interpretation of the principle of utility.

> 'Let us say ... that what the principle of utility requires of me is to do for each man affected by my actions what I wish were done for me in the hypothetical circumstances that I were in precisely his situation; and, if my actions affect more than one man ... to do what I wish, all in all, to be done for me in the hypothetical circumstances that I occupied all their situations ...'[39]

What I have just said here might be expressed differently, in terms of a condition Hare thinks must be satisfied in order to arrive at a universal prescription. If, supposing again that the things my recipients have good prudential reasons for wanting are also reasons for me to want the same things, I can rationally affirm that a particular action ought to be done if and only if its performance is likely (on the available evidence) to secure the greatest possible sum of desire-satisfaction for all concerned parties. Now, this clearly presupposes a constraint which must be thought to commit me first to sum and balance desires of different strengths and then to balance the result. What might this constraint be?

The answer is by now obvious though it involves many complications. We might begin by asking what constrains me to accept that all relevant rational desires are, for me, self-interested desires, which I have reason to balance against each other. The key to understanding the force of universalizability in Hare's argument for the PEI, is that it requires that I treat other people's rational desires 'as if they were my own'. The significant point is that I cannot reject the prescription 'Let me not be denied relief from my suffering', as 'my own' in the situation in which I occupy your position, for the objection is - put in terms of the supposition of equal probability referred to above - that I am not really accepting that this hypothetical situation might be the actual situation in which I am the affected party. Let us grant that this supposition, combined with the request for prudence, constrains me to have regard, prudentially speaking, for the satisfaction of your desires. The point then is that I am expected to treat the desires I associate with being in different positions as desires I have reasons to want to satisfy (albeit only initially). This is crucial, for this supplies the motivation to sum and balance which the principle of utility itself lacks.

Now, my suggestion is that the move to sum and balance runs on Hare's appeal to the notion of hypothetical self-interest in reversed-role situations. Let us grant that I am constrained to accept that any of these situations might be the actual situation in which I am the affected party. Suppose now that in all but one of the many situations in which I am the affected party, I desire, as one who is perfectly prudent, that x not be done (with a certain determinate strength, which is the same in all these situations). Suppose also that this one exception is the actual situation in which I desire that x be done. Now, I have to accept that there is greater probability that the desire that x not be done will be satisfied,

because there is greater probability that I will actually have this desire.[40] The point is simply that since there is but one situation in which I desire the contrary, the probability that any of the other situations might be the actual situation, will be much greater. I am here assuming that we may weight the probability of a situation actually occurring in the world as it is in terms of the number of instances that it actually occurs.[41] To this have to be added the strengths of the desires which accompany the occurrence. If, to be fully informed involves knowing both the probability of the situation occurring and the strengths of the desires, I must, in order correctly to represent the strength of the desire that x not be done, sum all the desires that it not be done.

The idea underlying the move to summing and balancing might be expressed as an assumption about my rationality in a two-person or multi-person case. This is that just as it is rational, in the prudential sense of 'rational', for me to act on what I most desire should be done in my own actual case, so it is rational for me to act on what I most desire should be done in the case in which, in addition to my own desires, I have other desires which I treat 'as if they were my own' through an act of the imagination. This assumption works on the basis of the contention that my final decision applies in one logically possible world in which it is equally probable that I might actually be any of the persons whose roles I perform qua affected party.[42] The significance of this contention is two-fold: it helps to retain the sense in which it is rational for me to be prudent in situations which differ from the actual situation in this respect that I have different universal features in them, and it prevents me from overriding the desires that conflict with my own actual desire, each taken separately (assuming, of course, that the strength of my own actual desire entitles me to do this). It is obvious that the prudential calculation

which I am then expected to make reduces to a choice between a greater and a lesser satisfaction. Since it is imprudent to forego greater satisfactions for lesser ones, I must, qua rationally self-interested person, opt for the greater satisfaction, and this is likely to be the satisfaction of the 'summed' desire that x not be done.

I have just suggested that the link Hare sees between the PEI and the principle of utility assumes that my final decision in a two-person or multi-person case parallel's a prudential decision in the single-person case. Prudence requires that I balance the 'summed' desire that x not be done against my own actual desire that x be done. There are two points worth noting regarding this request for balancing. First, I am expected to compare the value I assign to the consequences of acting on one of two prescriptions. In comparing these values, from the perspective of first-personal rationality, it has to be assumed that the reduction of inter-personal conflicts to intra-personal ones provides an adequate basis for the required comparison. The assumption is, in effect, that by construing all relevant rational desires as 'my own', I bring these desires onto the same scale of comparison. Second, since I would not, if prudent and fully informed, prescribe a course of action should it, over the range of situations at issue, yield less rather than more desire-satisfaction for myself, I would opt to maximize the satisfaction of my (rational) desires.

It is now time to show how the argument for utilitarianism works. The crucial claim is this: if I were apprised of all relevant information about a conflict situation, and made no logical errors in my reasoning, I would endorse an act-utilitarian solution to the conflict. This would be a solution which requires that I accept that line of action which is likely to yield the greatest net sum of desire-satisfaction for all concerned

parties. On the assumption that I have made no logical or factual error, I am entitled to conclude that this line of action is morally correct, since there would be no other course of action, among the alternatives open to me, which is likely to yield a greater net sum of desire-satisfaction.

How might we reconstruct Hare's argument for this claim? We begin with universalizability. Now, for Hare, universalizability is a property of moral judgements. Moral judgements entail identical judgements about all situations which are identical in what universal descriptive features they exhibit. The thesis of universalizability may be formulated thus:

> If I accept the moral judgement that x ought to be done in a certain situation, this implies that I am willing to prescribe x in each position I occupy in this situation, indifferently of the fact that I prefer to be in one position rather than another, and indifferently of the fact that in that position I desire that x be done, which is to say that my desire that x be done must be independent of which position I <u>in fact</u> occupy and of the fact that in <u>this</u> position I am the person who desires that x be done.

Let us grant that this thesis commits me to the supposition of equal probability which I have discussed above.[43] I am, then, by the thesis of universalizability, committed to the following premiss:

1. If I accept the singular moral judgement that I ought to do x to you, I am thereby committed to the view that, were I to be in your precise situation (which involves treating this situation as the actual situation in which I am the affected party), x ought to be done to me.

The prescriptivity of moral judgements is the thesis that moral judgements entail prescriptions for action. If I accept the singular moral judgement that I ought to do x, I am thereby committed to the prescription 'Let x be done'. Since accepting the singular moral judgement commits me to accepting the universal principle which supports it, I am committed to this prescription in all cases that fall under the principle. This is to say that prescriptivity, combined with universalizability, commits me to the following premiss:

2. If I accept the singular moral judgement that I ought to do x to you, I am thereby committed to the prescription 'Let x be done to me', were I to be in your precise situation.

Hare argues that I cannot be said rationally to be accepting this prescription, as the one I would be willing universally to endorse, in ignorance of the relevant facts of your case. I can be said to know these facts if and only if I know, in the required sense of 'know', 'what it would be like' for myself to be in your precise situation. Suppose I know that you are suffering, and that you want the suffering to stop. Given that these things are relevant to my appraisal of your situation, and that knowing them involves knowing how much I would dislike having your experiences, in a 'felt' sense of 'how much', I cannot be said to be acquainted with the relevant facts of your case unless I have equal motivation not to suffer as you are suffering. This is the point that an adequate representation of 'what it would be like' for myself requires that I adopt your concern for the satisfaction of your desires. I am suggesting that the requirement to universalize my prescriptions, which I must satisfy if I am to decide the issue between you and me rationally, commits me to the following premiss:

3. If I know that you desire that x not be done (because doing x to you is causing you to suffer), this entails that, in the case in which I am in your precise situation, I desire that x not be done, with the same strength or intensity as you desire this.

Now, I am rationally required to assent to your prescriptions on grounds that my final decision will otherwise be faulted for lack of information. Unless I assent - albeit only initially - to the prescription 'Let x not be done to me', it can be said that I lack the required knowledge of your case, because I have failed to make it clear to myself 'what it would be like' to be suffering as you are. My assent implies that I grant positive weight to desires I would have, were I to be in your precise situation, and that I am treating these desires 'as if they were my own', in the sense which the appeal to hypothetical self-interest requires. It is, for Hare, a necessary condition of my assent to this prescription that I desire (i.e. am motivated) to do what the prescription enjoins me to do in the case in which I am in your precise situation. Indeed, this would seem to be a necessary condition of my assent to <u>any</u> prescription, irrespective of whether I assent in actual situations or in hypothetical situations treated as actual. Hare says that desires are, in the generic sense in which any felt disposition to action counts as a desire, 'assents to prescriptions', and that prescriptions are simply the expression of desires in linguistic form. I formulate the premisses Hare has in mind as follows:

4. (a) I am fully informed about the facts of your situation only if I assent - albeit only initially - to the prescription 'Let x not be done to me' in the case in which I am in your precise situation.

4. (b) I assent to the prescription 'Let x not be done to me' only if I desire (i.e. am motivated) to do what the prescription enjoins me to do in the case in which I am in your precise situation.

2 and 4 are the crucial premisses. I am, by 2, committed to accept the prescription 'Let x be done to me'. By 4(b), this acceptance implies that I desire that x be done in the case in which I am in your precise situation. If I wish to universalize the prescription that x be done, it must be the case that I desire, on balance, that it be done in all situations of the given description. I am, however, by 4(a), committed to accept the prescription 'Let x not be done to me'. By 4(b), this acceptance implies that I desire that x not be done in the case in which I am in your precise situation. If I wish to universalize the prescription that x not be done, it must be the case that I desire, on balance, that it not be done in all situations of the given description. Now, if I were fully rational, in the prudential sense of 'rational', I would choose to act on what I most desire should be done. If, then, I desire, on balance, that x not be done, rationality will require that I act accordingly. It is not difficult to see that to extend this simple bi-literal case to multi-person cases involves the argument for summing which I set out above. In multi-person cases I am required first to sum and then to balance. If it is correct to say that this summing and balancing procedure runs, in multi-person cases, along lines analogous to a prudential decision in the single-person case, Hare would have shown that I am, qua rational agent, required to endorse, as morally correct, that line of action which is likely to maximize the satisfaction, in sum, of the (rational) desires of all concerned parties. This is to say, I shall not, if rational, endorse a moral judgement if it runs counter to the preferred form of utilitarianism. In this reconstruction of how Hare's argument works, I have assumed that the

argument begins with Hare's appeal to prudence, i.e. with the general motivational assumption that a rational agent wants to satisfy his own (prudent) desires to the greatest extent possible.

Section 2. A critique of the argument for utilitarianism

In this section my main objective is to contest the claim that we are, in any conflict situation, rationally required to accept an act-utilitarian solution. I shall concentrate on the moves Hare makes to commit his agent to assent to his recipient's prudent prescriptions. I formulated this request for assent as follows:

R: Rationality requires that I must assent, initially at least, only to those prescriptions which express in linguistic form, the dominant rational desires I would have in each situation in which my position were to be different.

Hare's point is that unless I assent to your prudent prescriptions, it cannot be said that I know the relevant facts of your situation, in the required sense of 'know'. As I have understood Hare, the knowledge I am expected to have of your situation is knowledge of 'what it would be like' for myself to want what you in fact want, and that this requires a strong reading of the sense in which I am expected to know the facts of your case. Briefly, I can be said to know these facts, if I know how much I would dislike to have a certain 'suffering'-experience, in a 'felt' sense of 'how much' which must correspond to how much you dislike having the experience. The crucial claim is that if I know that you desire x with a certain strength, this entails that I desire x with the same strength, and hence that I have equal motivation not to suffer as you are suffering in the case in which I am in your precise situation. It is from the first-person

perspective, made available to me by an imagined experience treated as actual, that I have to know 'what it would be like'. If we grant Hare the supposition that wanting for myself what you want for yourself works from this perspective, and in a sense of 'want' which is appropriate in the sphere of prudence, then the point is simply that the quality of this (imagined) experience must be thought to make a difference to my current motivational state, and hence is a reason for me not to want to suffer. This, of course, involves the appeal to hypothetical self-interest. As one who is assumed to be prudent about interests I deem to be 'my own', I am expected to identify only with those desires you would retain, if you were perfectly prudent, and to grant them (positive) weight, strictly in proportion to their strengths.

Now, I said that the crucial claim is that my knowledge of your (prudent) desire that x entails that I desire x, and with the same strength as you desire it. This claim rests on two considerations,

(a) that some state of another person might actually be a state of myself, and

(b) that I am, because I identify this state as a state of myself, and because I am prudent about interests I deem to be 'my own', concerned to satisfy the desires I associate with being in this possible state of myself.

If I am right, (a) derives from universalizability.[44] I have to accept that there is an equal chance that the hypothetical situation in which I occupy your position might be the actual situation in which I am the affected party. The move from (a) to (b) draws on an application of

universalizablity to prudential reasoning,[45] which involves the assumption of an antecedent motivational state to want to satisfy one's own (prudent) desires to the greatest extent possible. The point is simply that to the extent that I think of your motivational states as states of myself, to that extent I acquire a concern for the maximal satisfaction of desires I would have in the case in which I am in your precise situation, given that I am prudent about interests I deem to be 'my own' in this situation. If this is correct, Hare's argument in support of R above, as well as the PEI, rests crucially on the supposition of equal probability combined with the appeal to hypothetical self-interest. I shall accordingly, in what follows, concentrate on two problems which arise from this feature of the argument, and, in particular, on how these problems affect Hare's answer to the question 'Why should I assent to your (prudent) prescriptions?'

The first problem concerns the 'identity'-question which I discussed in Chapter 2. I am, in reversing roles with you, expected to concede that I am the person who occupies your position, and hence that I am, qua rationally self-interested person, concerned about the satisfaction of (rational) desires I deem to be 'my own' in this position. I shall argue that it is not possible to make coherent sense of the claim that the motivational states I am expected to deem 'my own' are properly states of myself, which is to say that it is not possible to make coherent sense of the claim that I am the person in your precise situation. The point is simply that it is not possible to make coherent sense of the claim that I can imagine 'what it would be like' for myself to be in your precise situation, and hence that the claim that I can know the relevant facts of your situation, in the required sense of 'know', in propria persona, is itself incoherent.

The second problem concerns the 'motivational'-question which I discussed in Chapter 2. This touches on the claim that the things you have good prudential reasons for wanting are also reasons for me to want the same things. I shall argue that your reasons are not also reasons for me on grounds that Hare's appeal to hypothetical self-interest does not work for him in the way he requires. The point is simply that my representation of what you have prudential reasons to want or to do in your own case cannot be construed as a representation of what I have reason to want or to do in that case, in the sense in which having those reasons makes a difference to my actual current motivational state.

I begin with the 'identity'-question. I have maintained that the rationality constraints work from the perspective of the first-person. It is worth noting that in ETU Hare presents his understanding of what is involved in the requirement that I must know the relevant facts of the situation I am judging in much the same idiom as that employed in FR. In a comment on Rawls' use of formal apparatus, he has the following to say:

> '... rather than put the argument in [Rawls'] way, I will do overtly what he does covertly - that is to say, I do not speculate about what some fictitious rational contractors would judge if they were put in a certain position subject to certain restrictions; rather, I subject myself to certain (formally analogous) restrictions and put myself (imaginatively) in this position, as Rawls in effect does, and do some judging.'[46]

The question is not what I would have reason to say, to want or to do, were I to be the person in your precise situation, but rather what I do have reason to say, to want or to do, in propria persona. It is worth recalling

that in this situation my position, qua affected party, is best understood as the position of the recipient of the action I wish universally to prescribe in the actual situation. In this position I am stripped of all information about the particulars of my own actual case. The information Hare supposes is available to me is information about how I would be affected, given that I am situated just as you are in your actual position, qua recipient, in all relevant respects.

There are two points which must be stressed regarding this notion of 'in all relevant respects'. First, for Hare there is nothing incoherent about supposing that I am identical with myself in the case in which I am in your precise situation, though in that case I will have a different set of universal features. This works on the contention that there is nothing in logic which prevents me from imagining that I may lose one set of universal features and, at a different time and in a different position, acquire another set, yet remain the same person in spite of this change.[47] Second, it must then be thought that I am the person who has different experiences in (supposedly) identical situations. For Hare the relevant sense in which I am expected to know 'what it would be like' for myself to have your experiences is the 'felt' sense of how much I would dislike having them. Hare thinks that I must be able at least to imagine what this would be like for myself. This is best understood as analogous to imagining what it would be like to be in some future state of myself, and now desiring what I currently believe I will then want. The idea is simply that my imaginative representation of what you have reason to count as relevant to a prudential decision in your own case serves as a representation of what I have reason to count as relevant to a prudential decision in the case in which I am in your precise situation. If this is correct, I must be thought to be having the same experience as you are having, albeit only an imagined experience

treated as actual. This sets one condition of knowing the facts of your case, in the required sense, which I must fulfil if I am to '... do some judging rationally'.

But, it may now be objected, that I cannot do this judging rationally, at least not in the way Hare understands what is involved in rationally deciding on a course of action.[48] It is worth noting that the request for correct factual information about your case may be understood in different ways. It may mean that I am expected, as Hare puts it, 'to do some judging', from the perspective of the first-person, in which case the request for correct factual information may be understood as demanding an answer to the following question:[49]

(1) 'What do I say about the hypothetical situation in which I am in your precise situation?'

(1) assumes that it makes coherent sense to say that I am imagining 'what it would be like' for myself to be in your precise situation. Recall that in the reversed-role situation, I am expected to make a decision on the basis of all and only the relevant features of your situation. The question then arises how I might make this decision from the perspective of the first-person, and the answer is that, stripped of all knowledge of the particulars of my own case, I merely have to imagine that I am the person who has your universal features, including your motivational states. However, it is not difficult to see that this must mean, not that I am imagining myself to be you, but rather (and simply) what it is like for you to be having a certain experience.[50] Imagining the feel of what it is like for you is possible on the assumption that I have access to your experiences, just as I have access to my own.[51] Hare, however, assumes, on

the basis of the assumption just mentioned, that it makes sense to say that I am the person standing in your shoes, i.e. that I can imagine the feel of what it is or would be like for myself, in a way which parallels imagining the feel of what it is or would be like to be in my own future shoes. But this makes no sense because that person, standing in your shoes, stripped of all the particulars of his own actual case, and who has only your motivational states and other relevant characteristics, cannot be myself. So the request that I '... do some judging', understood as demanding an answer to (1), is incoherent, because it requires what is logically impossible, viz, that I can imagine myself to be standing in your shoes.

I have suggested that it is possible to imagine the feel of what you are experiencing. So let us construe the request for correct factual information about your case as a request to know this, which knowledge I then construe as knowledge of what I would want, in some appropriate sense of 'want'. This is to say, let us construe this request as demanding answers to the following questions:

(2) 'What do you say about the situation in which you stand to me in relation of recipient to prescriber?'

(3) 'What would I say about the hypothetical situation in which I want what you in fact want?'

(2) assumes that I have, through an act of the imagination, access to the feel of your experiences, just as I have access to the feel of my own experiences. I can imagine, that is, feeling the intensity of your suffering and desiring that it be stopped, just as I can do these things in my own case. The question now arises how having this knowledge of what it

is like for you translates to a motivational state in me.[52] Recall that my having your motivational states, as ST formulates this, is a condition of my fulfilling KR. Now, Hare operates with a generic sense of 'desire' according to which any felt disposition to action counts as a desire. It is in this generic sense in which a desire is a felt disposition that I have to desire what you in fact desire. But, it may now be objected that this 'felt disposition' account of my knowledge of your desires does not work in the required way. This account, which is implicit in KR and ST, defines the bottom-line of what constitutes knowledge of your case. However, if I am right, I cannot construe the feel I have of what it is like for you as knowledge of what it is or would be like for me to be in your precise situation because it makes no sense to say that I can imagine myself to be you. And, if knowledge of your experiences is not available to me from the first-person perspective within the reversed-role situation, the crucial claim that my knowledge of your desires entails that I desire what you in fact desire, may be rejected as false. The upshot is simply that I (logically) cannot fulfil either KR or ST, since they run on a specific sense of 'know' which is not available to me in the required way.

Now, I have suggested that I can imagine what it feels like for you to be having a certain unpleasant experience by representing the feel of this experience to myself in a way which parallels how I represent the feel of a current or anticipated future unpleasant experience to myself. This way of putting the matter retains something of the 'felt disposition' account of my knowledge of your desires, though not the motivational implication Hare wants. If this is correct, (2) is the question I should be asking, and not (3). I can, however, not provide an answer to (3) if what I am expected to say must be thought to depend on my imagining that I am the person in your precise situation. Indeed, this appeal to the imagination is what is wrong

with (1), and if it is dropped, I could sensibly ask either 'What do I say about your situation?' or 'What would I say about it?', and my answers must correspond to what you say or would say about it. But, dropping this appeal to the imagination would mean dropping role-reversal as well, which leaves Hare without an account of the possibility of KR and ST.

I am arguing that (2) is the question I must ask, and that if this question is applied to myself, I must say what you say about your situation. My answer, however, is not what Hare requires. And I (logically) cannot provide the required answer for the simple reason that I cannot treat my representations of what it is like for you as representations of what it is or would be like for myself, i.e. as first-person representations within the reversed-role situation, since it makes no sense to say that I can imagine myself being you, with your universal features and other relevant characteristics. As I judge, Hare's mistake is to make the knowledge requirement of rationality depend on a move he builds into the role-reversal procedure. This is the move by which I am required to bracket or put aside all information about the particulars of my own actual case, which Hare interprets as demanding that I allow not only that the position I in fact occupy is morally irrelevant, but also that my individual identity is irrelevant to the moral decision. The demand for universalization itself requires this interpretation. But in so far as this interpretation demands that we must abstract from the identities of the persons involved in order rationally to arrive at a moral decision,[53] it is not possible to make coherent sense of what KR requires.

I have maintained that the demand for correct factual information about your case can be satisfied only by appeal to the role-reversal procedure. In reversing roles with you, I am committed to fulfilling a condition of

rational assent to your prescriptions. This is, as KR and ST imply, that I am not in possession of correct factual information about your case unless I myself want, were I to be in your precise situation with your desires, that these desires should be satisfied. I have also maintained that KR and ST require a strong reading of the sense in which I am expected to be concerned about the satisfaction of your desires. This sense is akin to the sense, appropriate in the sphere of prudence, in which a single person experiences concern for the satisfaction of his desires. The relevant point to stress is that to identify some other person as myself, and his desires as 'my own', is to desire what he in fact desires, as KR makes clear. As I understand, I have to treat your (prudent) desires as 'my own' in the sense that I have regard - prudentially speaking - for their satisfaction. This is necessary if my initial assent to your prescriptions is to be a requirement of my rationality in the prudential sense of 'rational' which Hare's theory requires. There is in this a requirement to the effect that the things other people have good (prudential) reasons for wanting are, or must be construed as reasons for me to want the same things.

Now, Hare's appeal to what I have reason to want or to do in the case in which I am in your precise situation rests essentially on an appeal to my self-interest - albeit only hypothetical self-interest. I have argued that the argument for utilitarianism begins with this appeal, and that it translates to an appeal to hypothetical self-interest in reversed-role situations. Regarding the justification of the latter move, my suggestion has been that universalizability must be thought to constrain prudence to the demands which rational decision-making in morals impose on agents, as least as Hare conceives of these demands. The suggestion is that universalizability must be thought to constrain me to allow that any of the situations in which I occupy the position of affected party might be the

actual situation, and that I might actually be any of the persons whose positions I occupy qua affected party. If we allow Hare this contention, we might also grant him the premiss that to the extent I identify the motivational states of another person as states of myself, to that extent I acquire the required concern for the satisfaction of his desires, and hence am willing to prescribe, at least initially, that these desires should be satisfied, as my rationality requires. The appeal to hypothetical self-interest rests squarely on the contention mentioned above, for without it, it is difficult to see just how the appeal to self-interest may be extended to cover hypothetical situations in which I am expected to ascribe to myself motivational states other that those I might actually or in fact have for what should happen to myself in them.

It is not difficult to see that the idea that universalizability constrains me to think of the person who has these experiences in the reversed-role situation as myself, supplies the key to understanding the motivational constraints of Hare's theory. The basic idea is that if I fully represent to myself my own (unpleasant) experiences, were I to be in your precise situation with your desires, I must acquire a motivational state which is just like yours in all relevant respects. If I do not, I have not fulfilled these conditions, and so have not become fully rational. Add to this the premiss that to have a desire is to accept a prescription that it be satisfied, and it seems that Hare is in a position to clinch the motivational issue simply by making it a requirement of my rationality that I must assent to your (prudent) prescriptions.

Now, the strength of this position derives ultimately from the contention that to become fully rational, I have to fulfil conditions which arise from the demand for universalization. As I judge, there is in this one clear

point: I have to treat you as deserving of equal treatment, which means that I have to grant equal consideration to your desires, the same as I grant my own. But, and this is the difficulty with this point, I am compelled by the demand for universalization to think of the person in the reversed-role situation as myself. This is to say, in order to become fully rational, the demand for universalization requires that I abstract from my own individual identity and assume the universal features of your person and situation. This is necessary if I am to fulfil KR. KR requires that I know how much you desire x, in the 'felt' sense of 'how much', and from the perspective of the first-person in the reversed-role situation. This means that KR cannot stand independent of universalizability, and hence that in fulfilling KR I have no reasons which satisfy the constraints of universalizability to treat your desires differently from my own.

I argued above that it makes no sense to say that I can imagine myself to be the person in your precise situation. KR fails on these grounds. It is not difficult to see that the appeal to self-interest in reversed-role situations fails on the same grounds. If I am right, I have to construe the person I imagine to be having certain unpleasant experiences to be you, and this leaves no room for an appeal to self-interest, albeit only hypothetical self-interest. So, it may be asked on what grounds my feel of your unpleasant experiences is a reason for me to assent to your prescriptions, i.e. why I <u>have</u> reason to assent.

If we abandon role-reversal, as I think we should, it is worth asking how far it is possible to push the fact that I can know the feel of what it is like for you as a motivational constraint to which I am subject, <u>qua</u> moral agent.

It is not difficult to see that we would have to allow substantive moral intuitions to intrude at a critical stage in Hare's theory. I might argue thus: I might say that if I am to suffer, and desire not to, I ought to be offered relief from my suffering, and offer as reason 'Anyone who suffers, and desires not to, ought to be offered relief'. This would commit me to conceding that, if you were suffering, and desire not to, you ought to be offered relief. It is not necessary to appeal to role-reversal to establish this point. I merely have to assent to the moral principle I have appealed to in this simple argument, and the singular prescription it entails. And if I assent to these, I would be appropriately motivated, though not strictly in the way Hare's theory requires.

If we drop the role-reversal procedure, as I think we must, critical thinking, as Hare conceives of it, cannot proceed. I wish to advance three points. I maintained that two applications of prudence, at different levels, are needed to bring me to a decision concerning what I morally ought to do, first to determine your criticized desires and then to determine what, in the end, all things considered, I ought rationally to prescribe. Regarding this final decision, I am expected to choose qua rationally self-interested person, i.e. as one who wants to satisfy my own interests to the greatest extent possible. If my final decision is ultimately determined by an application of universalizability to prudential reasoning, I have to prescribe - on the basis of the supposition that it is equally probable that I might myself actually be any of the persons whose positions I occupy - that the rational desires I associate with being in these positions, should be satisfied strictly in proportion to their strengths. I am then, qua rational decision-maker, in a position which is structurally similar to a single-person's treatment of his rational desires, i.e. I am, in a way which parallels his treatment of these

desires, in the two-person or multi-person case, committed to the principle which bids that desires of equal strength count equally.

Now, one point is that Hare's strategy to bring me to a final decision runs on the logical impossibility that an action can, without inconsistency, have different and opposite values for me, at least not at the same time.[54] This strategy will run only if the moves Hare employs to give the desires and other relevant characteristics of distinct individuals a place in my deliberation, works in the way suggested, viz., that one and the same person must be thought to be making a decision, which affects another person, on the assumption that I am identical with myself in each role I perform. But, as I have argued, I am not identical with myself in all of these roles. So the strategy Hare employs to resolve inter-personal cases of conflict, which is to reduce them to intra-personal ones, cannot come into play.

A second point is the assumption that the reduction of inter-personal cases of conflict to intra-personal ones provides a basis for inter-personal comparisons. The assumption is, in effect, that by construing all relevant rational desires as 'my own', in the sense that I have regard - prudentially speaking - for their satisfaction, I bring these desires on to the same scale of comparison.[55] I am supposed to be able, along lines analogous to intra-personal comparisons, to compare the values different persons assign to various expected outcomes, from the perspective of first-personal rationality. But, if I am right, the intra-personal response to the comparison problem is no solution at all because no intra-personal reduction is possible.

The third point follows from the ones just made. The major significance of the impossibility of intra-personal reduction as a solution to inter-personal comparisons is simply that it undermines the idea underlying Hare's argument for summing and balancing. I expressed this idea as an assumption about my rationality in a two-person or multi-person case of conflict. This is that just as it is rational, in the prudential sense of 'rational', for me to act on what I most desire should be done in my own actual case, so it is rational for me to act on what I most desire should be done in the case in which, in addition to my own desires, I have other desires which I treat 'as if they were my own' through an act of the imagination. If we reject this assumption, as I think we should, we in effect reject the idea that a single principle of rational choice can govern two different kinds of conflict situation, viz., conflicts between distinct individuals and conflicts within the lives of particular individuals.

It is difficult to see how Hare might retain the motivational constraints of his theory without appeal to the formal restrictions of the role-reversal procedure. It may be granted that the reason I have to be concerned about the satisfaction of (rational) desires I currently have and would have in future states of myself, which is simply that these states are and will be states of myself, is sufficient ground for me to allow that you have the same reason to be concerned about the satisfaction of your (rational) current and future desires. It is, however, less easy to grant that I have reason to be concerned about the satisfaction of your (rational) desires simply on grounds of what it means for me to be in your precise situation. I have suggested appeal to moral intuitions of substance as a motivating constraint. Hare, however, cannot avail himself of this

suggestion, at least not without abandoning the project he is attempting to defend.

Conclusion

I conclude this chapter with some brief remarks concerning the direction in which Hare's thinking develops in Moral thinking (MT). There are two significant developments in MT. Hare attempts to defend the knowledge requirement of his theory, not as a requirement of rationality, but rather as a conceptual truth which is independent of universalizability. He combines this requirement with a newly developed theory of prudence in a renewed argument for act-utilitarianism. These developments represent a clear departure from a claim which Hare advances in FR and ETU, which is that all the moves he need make to argue for utilitarianism derive from the logic of the moral concepts, universalizability and prescriptivity, alone. I shall examine these developments in Chapter 4 in order to establish whether Hare succeeds in defending this new project.

Notes

1. Hare (1980): Freedom and reason. Oxford University Press. Oxford. p. 170. Hare says: '... if we use the word "desire" in a wide sense, we can say that any evaluation, just because it is prescriptive, incorporates the desire to have or do something rather than something else. The wide sense in which we are here using "desire" is that in which any felt disposition to action counts as a desire ...' In ETU Hare says that '... desires, in the required sense, are assents to prescriptions'. See Hare (1989): Ethical theory and utilitarianism. Essays in ethical theory. Clarendon Press. Oxford. p. 217. (ETU was first published in 1976 in Contemporary British philosophy edited by H.D. Lewis.)

2. Hare (1989) p. 217-218.

3. Hare makes it clear that I have to accept that any of the 'occupants' of the roles I perform 'might be myself'. Hare (1989) p. 215.

4. Hare (1989) p. 218. Hare says: '... it is qua rational that I have to judge; and this involves at least judging with a clear and unconfused idea of what I am saying and what the actual consequences of the prescription that I am issuing would be, for myself and others'.

5. I take it that the 'felt'-disposition account of 'desire' is the account with which Hare operates in ETU. See again note 1 above.

6. Hare (1989) p. 218.

7. Hare (1989) p. 217.

8. Hare (1989) p. 218. I am here anticipating developments in MT.

9. This way of putting the matter due to W. Rabinowicz. See Rabinowicz (1989): Hare on prudence. Theoria, vol. LV, p. 147.

10. The term 'hypothetical self-interest' is due to Hare. See Hare (1988): Comments on Brandt. Hare and critics, edited by D. Seanor and N. Fotion. Clarendon Press. Oxford. p. 218. The point is made in the context of Hare's response to a criticism offered by R.B. Brandt. I am here again anticipating developments in MT.

11. Hare (1989) p. 218.

12. KR is not explicitly stated in ETU. Hare says: '... when we speak of the "situations" of the various parties, we have to include in the situations all the desires p [p = prudence], likings p, etc., that the people have in them - that is to say, I am to do for the others what I wish p to be done for me were I to have their likings p. etc., and not those which I now have.' See Hare (1989) p. 218. As a requirement of my rationality, KR is satisfied by appeal to the demand for role-reversal and imaginative identification with your prudent desires. As I understand Hare, this demand is required by the demand for universalization.

13. As I have understood Hare, the judgement I make in the hypothetical reversed-role situation is made in propria persona, i.e. on the basis of the supposition that I might actually be the person in your precise situation.

14. See again note 12.

15. ST is not explicit in ETU. I am here anticipating developments in MT. As I understand Hare, it is a condition of satisfying KR that I desire not to suffer (i.e. have equal aversion to suffering) as you are suffering. This is to be motivated not to suffer as you are suffering. I am here following H.J. McCloskey. According to McCloskey Hare '... must show that the moral point of view relates not simply to sincerely universalized prescriptions, nor simply to universalized prescriptions directed at maximizing satisfaction of interests (the position towards which he appears to have been feeling in Freedom and reason), but to maximum satisfaction of desires, in his sense of desire, a sense which encompasses reasons for action which are commonly contracted with desires.' McCloskey maintains that Hare's definition of 'desire' (note 1 above) is basic to the discussion in ETU. I agree with this view. See McCloskey (1978/79): Universalized prescriptivism and utilitarianism: Hare's attempted forced marriage. The journal of value inquiry, volumes 12-13, p. 69 and p. 71.

16. Hare (1989) p. 218-219.

17. Hare (1989) p. 218-219. Michael Gorr argues that this '... reformulation of the full reversibility test effectively jettisons the requirement that moral decisions be made in propria persona. Instead, Hare's present view seems to be that moral judgements, by their very nature as moral judgements, must be made from a completely impartial point of view rather than from the point of view of the agent himself.' However, it seems to me that Gorr misconstrues the import of the in propria persona requirement. Gorr says: '... to make a decision in propria persona, is, presumably,

to make it on the basis of one's knowledge of which desires, interests, and ideals one actually has.' As I understand this requirement, the judgement I make in a hypothetical reversed-role situation is made on the basis of the supposition that I might actually be the person in that situation with his desires. See Gorr (1985): Reason, impartiality, utilitarianism. Morality and universality edited by N.T. Potter and M. Timmons. Reidel. Dordrecth. p. 122.

18. Here I agree with Gorr. See Gorr (1985) p. 122. See also Hare (1989) p. 215 and p. 220-221.

19. Hare (1989) p. 217-218.

20. Hare (1989) p. 217.

21. Hare (1989) p. 217.

22. Hare (1980) p. 158.

23. Hare (1980) p. 196.

24. Hare (1989) p. 219.

25. Hare says: '... an ideal is a kind of desire or liking (in the generic sense in which I am using those terms ...).' See also note 1 above. Hare (1989) p. 219.

26. Hare (1989) p. 215. Hare says: '... since any of [the] occupants [of the various roles at issue] might be myself, [the] weight [I grant to their desires] will be positive.'

27. I think this represents a clear departure from Hare's position in FR.

28. Hare (1989) p. 219-220.

29. Hare (1989) p. 215.

30. I am here again anticipating developments in MT.

31. According to Don Locke Hare's argument for the PEI begins with his appeal to the requirement of prudence. Locke says: 'This argument begins from the psychological generalization - or logical necessity? - that people wish to satisfy their own desires to the greatest extent possible.' See Locke (1981): The principle of equal interests. The philosophical review, vol. XC, p. 546.

32. See again note 10 above.

33. Hare (1989) p. 217-218 and p. 219.

34. Don Locke argues that this argument fails. He says: '... if we are going to make moral judgements - and it is only then that Hare's arguments are intended to apply - we are committing ourselves to endorsing a particular course of action however it might affect us personally; and if we do this prudently we will refuse to commit ourselves through our moral judgements to any course of action which, over the range of possible cases, would be more against our interests than in them.' The crucial point is that if Hare's agent is rational '..."in the prudential sense of 'rational'", he will take account only of those desires and interests which are likely to become his and only to the extent that they are likely to become his.' See Locke (1981) p. 547. In my reconstruction of Hare's argument, I am treating the demand for imaginative identification and the demand that hypothetical situations be treated as actual as required by the demand for universalization. I interpret these demands in a strong sense: I must accept that it is equally probable that I might actually be any of the persons whose roles I perform, and that it is equally probable that any of the situations at issue might be the actual situation. I am here anticipating developments in MT.

35. Hare (1989) p. 215.

36. I am indebted for this point to Locke. See Locke (1981) p. 546.

37. Hare (1989) p. 215.

38. Hare (1989) p. 215.

39. Hare (1989) p. 215-216.

40. I am indebted for this point to I. Persson. See Persson (1989): Universalizability and the summing of desires. Theoria, vol. LV, p. 166.

41. See again Persson (1989) p. 167.

42. This again anticipates developments in MT. See also Hare (1989) p. 216. I occupy the various roles I perform, qua affected party, 'in random order'. Hare says: 'This way of putting the matter ... emphasizes that I have to give the same weight to everybody's equal interests; and we must remember that, in so far as I am one of the people affected ... my own interests have to be given the same, and no more, weight - that is to say, my own actual situation is one of those that I have to suppose myself occupying in this random order.'

43. I have said that this anticipates developments in MT. However, without this supposition it is difficult to see how Step 1 of the argument below is possible.

44. Hare (1989) p. 215.

45. See note 44.

46. Hare (1989) p. 214.

47. See Hare (1981) Moral thinking. Oxford University Press. New York. p. 119-121.

48. McCloskey remarks that 'Hare's whole notion of rationality needs explanation, clarification, justification'. See McCloskey (1978-79) p.75. McCloskey, however, raises this point for different reasons than those I have in mind. He says: '... Hare seeks to derive his new utilitarianism by moving from his interests - fanatics position to the contention that the criterion of the moral point of view is to be found in terms of universally prescribing the satisfaction of desires.' If McCloskey is right, the shift to 'desires' requires justification. He says: The only substantiation that seems to be offered is that in terms of an account of the normative ethic itself.' See McCloskey (1978-79) p. 68-69.

49. In what follows I adapt the critique presented in FR. I rely here on C.C.W. Taylor. See Taylor (1965): Review of Freedom and reason. Mind, vol LXXIV, p. 286 and p. 287.

50. See Taylor (1965) p. 289. Taylor says: 'State (c) [i.e. imagining what someone else feels in a particular situation] is never logically impossible, though it may be practically very difficult or even impossible.'

51. I am here assuming that the 'felt' disposition account of 'what it would be like' to have your desires, is the appropriate account. See again note 1 above.

52. I have McCloskey and Taylor in mind here. See again notes 15 and 49 above.

53. The phrase 'abstract from the identities' is Vendler's. See Z. Vendler (1988): Changing places? Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.p. 183.

54. See Jaegwon Kim (1984): Concepts of supervenience. Philosophy and phenomenological research, vol. XLV, p. 159.

55. See J. Griffin (1988): Well-being and its interpersonal comparability. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford. p. 79.

CHAPTER 4

An analysis and critique of 'Moral Thinking' (MT)

Introduction

In MT Hare develops a new thesis which stands independent of universalizability. This is the thesis that my knowledge of your preferences entails my having the same preferences for a similar hypothetical situation in which I am in your precise situation, and which universalizability compels me to treat as actual and 'my own'. Hare defends this thesis with reference to another thesis, viz, the thesis of the prescriptivity of 'I' and various conceptual claims he makes regarding my knowledge of other people's preferences. These things work with a concept of prudence and various rationality constraints in a new argument for utilitarianism.

I shall argue that the argument for utilitarianism fails on grounds that these theses do not commit me to the moves Hare needs to make to compel me to accept act-utilitarian conclusions. These moves require that we construe the decision-procedure he offers for a bi-lateral case of preference conflict as if it were a single-person case of preference conflict. I shall attempt to show that this move may be rejected, and that the argument for the maximization of preference-satisfaction fails.

Section 1: An outline of the argument for utilitarianism

In this section my object is to show how Hare thinks that a moral judgement consistent with act-utilitarianism may be derived from his premises. These are

(1) the logical principle of universalizability;

(2) the logical principle of prescriptivity and its relation to preferences;

(3) an analytic principle of knowledge, which claims that my knowledge of someone's preferences entails my having those preferences, and which specifies the conditions governing what shall count as knowledge of what it is like for someone to be having a certain experience;

(4) a set of rationality constraints, in particular the requirement of prudence.

According to the version of act-utilitarianism which Hare defends, an act is morally right if and only if no alternative is likely to yield a greater net-sum of preference-satisfaction. The central claim, on which his defence of act-utilitarianism rests, might be expressed thus: if we were apprised of all the relevant and available facts of a situation, and made no logical errors in our reasoning, we would not endorse any moral judgement which runs contrary to that required by the preferred form of utilitarianism.

I begin by stating the relevant versions of the universalizability and prescriptivity theses as employed in MT, as well as other premisses which relate to them. Hare formulates the universalizability thesis as a property of moral 'ought'-judgements. Moral judgements entail identical judgements about all situations which are identical in what universal descriptive features they exhibit.[1] I formulate the thesis as follows:

U: If I accept the moral judgement that x ought to be done in a certain situation, this implies that I am willing to prescribe x in each position I occupy in this situation, indifferently of the fact that I prefer to be in one position rather than another, and indifferently of the fact that in that position I prefer that x be done, which is to say

that my preference that x be done must be independent of which position I in fact occupy, and of the fact that in this position I am the person who prefers that x be done.

This thesis commits me to the view that, should I wish to do x to you, and think I ought to, I have to be willing to accept that x ought to be done to me in the case in which I am in your precise situation. This premiss works with the knowledge requirements of Hare's theory, which I must fulfil if I am to make a rational decision about how I would like to be treated in the case in which I stand in your shoes.[2]

Hare formulates the prescriptivity of moral judgements as the property of entailing at least one prescription for action. If I accept the singular moral judgement that I ought to do x, I am thereby committed to accepting the prescription 'Let x be done'.[3] Since accepting the singular moral judgement commits me to accepting the universal principle which supports it, I am committed to this prescription in all cases that fall under the principle. This is to say that prescriptivity, combined with universalizability, commits me to the following view:

P: If I accept the singular moral judgement that I ought to do x to you, I am thereby committed to the prescription 'Let x be done to me', were I to be in your precise situation.

Hare argues that prescriptions are the proper expression, in linguistic form, of preferences.[4] To have a preference that x be done is to accept the prescription that it be done, which we express by saying 'Let it be the case that x be done'. This relation between prescriptivity and preferences implies that, if I accept a prescription, then it must be the case that I

prefer to do what it enjoins me to do.[5] I formulate the premiss Hare has in mind as follows:

P_1: I accept the prescription 'Let it be the case that x be done' only if I prefer that x be done.

Hare maintains that the knowledge conditions which I have to fulfil, if I am to make a rational decision about how I would like to be treated in the case in which I am in your precise situation, rest on conceptual truths which concern the relation between knowing that someone is suffering and actually experiencing this suffering.[6] The main condition is that I have to know what it would be like for myself to be suffering as that person is suffering, and this condition can be satisfied only by appeal to the demand for role-reversal and imaginative identification with his preferences. The crucial claim Hare makes regarding my knowledge of his situation is that my present (full) knowledge of his preferences entails my now having the same preferences as he has for what should happen to himself in that situation.[7] I do not fully know these hypothesized preferences unless I accept not only that I would have them, but also 'what it would be like' for myself to have them. As I understand this notion of 'what it would be like', I am expected to want for myself, in the case in which I am in his precise situation, what he wants should happen to himself.[8] And, since in this situation, I have to allow, as counting among the relevant facts of his situation, the relevance of facts about experiences I would have, particularly facts about how much I would dislike having his experiences, knowing 'what it would be like' involves my having his concern for the satisfaction of his preferences. I understand this concern in a strong sense, akin to the sense, appropriate in the sphere of prudence, in which a single person

experiences concern for the satisfaction of his preferences. I formulate the knowledge requirement which Hare defends as follows:

KR: If I know that you prefer x, this entails that I now prefer x in the hypothetical case in which I am in your precise situation.[9]

It is clearly possible to defend KR as a requirement which rationality imposes on me.[10] But in MT Hare goes beyond this to claim that KR is a 'conceptual truth' and that it is distinct from universalizability.[11] Hare's case rests on the truth of a conceptual link which he sees between the following two propositions:

(a) 'I now prefer with strength S that if I were in that situation x should happen rather than not.'
(b) 'If I were in that situation, I would prefer with strength S that x should happen rather than not.'[12]

According to Hare I cannot know that (b), and what that would be like, without (a) being true, and this is a conceptual truth.[13] As I understand, it is a conceptual truth that I cannot know what it would be like for myself to be treated as I am treating you, unless I know what it is like for you to be treated like that. It is clear that a certain sort of knowledge is required, which is knowledge of how much I would dislike to be suffering like that, in the 'felt'-sense of 'how much'.[14] This is knowledge appropriate to the first-person perspective, i.e. knowledge I <u>have</u> of the feel of my experiences in the reversed-role situation, which I cannot have unless I not only know what you prefer, but also (actually) prefer the same thing. Hare's answer to the question concerning how it is possible that I can form a preference concerning how I would like to be treated in the case

in which I am in your precise situation, involves his thesis of the prescriptivity of 'I'. Hare says:

> '... by calling some person 'I', I express at least a considerably greater concern for the satisfaction of his preferences than for those of people I do not so designate.'[15]

The thesis is that I use the word 'I' of myself in the reversed-role situation, which I cannot do without preferring for myself what you prefer should happen to yourself in the actual situation. This, Hare thinks, follows from the prescriptive element in the meaning of the word 'I', when combined with the requirement that hypothetical situations must be identical to actual ones, which identity must extend to the preferences of the persons in them.[16]

Now, this thesis is extremely contentious. The account Hare offers of it runs on the truth of the following considerations, which concern a relation Hare sees between knowing that I am suffering, and (actually) experiencing this suffering.

(a) 'If I am suffering, I know that I am suffering ...'
(b) 'If I am suffering, I have a motive for ending the suffering.'
(c) 'It would be self-contradictory to report suffering but claim that one did not mind it, and had no motive for ending or avoiding it ...'[17]

The significant point about these considerations is that they are held to be true for actual and hypothetical reversed-role situations.[18] They have a peculiar relation to the knowledge requirement and, indeed, make clear that this requirement includes certain motivational constraints as a component

of knowledge. They therefore have relevance to the motivational assumptions of KR, and represent one step towards an account of how it is possible that my present reasons for action can be influenced by an adequate knowledge of the relevant facts of your situation. They can be rephrased to include a reference to my knowledge of the degree or intensity of my suffering, were I to be in your precise situation. As I read Hare, it is a condition of satisfying KR that I must now have an equal aversion to myself suffering as you are suffering. I formulate this as Hare's suffering thesis (ST).[19]

ST: Unless I now have equal aversion to suffering as you are suffering, in the case in which I am in your precise situation, it cannot be said that I really know the relevant facts of your case (i.e. I would be in factual error about what it is like for you).[20]

It is important to guard against a possible misinterpretation of KR. Does KR imply that if I have your preference that x, I can have no other preference which could outweigh this?[21] Hare's view is that I will form a preference for what should happen to myself, if I were now in your precise situation, which is identical in content and strength to your actual preference.[22] We might illustrate this by means of the following example.[23] B, the patient, prefers that the drilling into his decaying tooth be stopped immediately. By KR, A, the dentist, prefers that, if he were in B's position suffering that degree of pain, the drilling be stopped immediately. But though A must form this preference, he can have the possibly stronger preference that, if he were in B's position, his dentist should go on drilling, for otherwise he will not be able to avoid toothache in the future.[24] If A's own preference is really stronger than his acquired preference, and assuming that no other preferences are involved, there can on Hare's view be no bar to allowing it to override the weaker one. The

point is that KR does not assume that a recipient's preferences will always have a veto, irrespective of its strength.[25]

I have indicated that ST defines the bottom line of what constitutes an adequate knowledge of the relevant facts of your situation. The important point is that I am not in possession of correct factual information about your situation unless I now have equal aversion to myself suffering as you are suffering.[26] This, as Hare sees it, is a condition of rational assent to your prescriptions. It is, if Hare is right, a tautology to say that, if I really know the facts of your situation, in the sense of 'know' made clear above, I shall prescribe that the preferences I would have in the case in which I am in your precise situation, be satisfied, and strictly in proportion to their strength.[27] I am, however, required to prescribe this only initially, as a step towards finally endorsing a universal prescription. I formulate this initial requirement of my rationality as follows:

R: Rationality requires that I prescribe - initially - that the preferences I would have, were I now in your precise situation, be satisfied strictly in proportion to their strength.

However, Hare requires that the preference I form in the reversed-role situation must be the one you would retain, if you were perfectly prudent as far as your own interest goes.[28] This is part of what Hare understands by knowing the relevant facts of your situation, and part of the necessity, as he sees it, of rationally endorsing the prescription which I accept as applying to myself in the reversed-role situation. Hare expresses his notion of prudence as a current preference which I should have, and which

should always have overriding strength relative to other current preferences. The requirement of prudence is formulated as follows:

> '... we should always have a dominant or overriding preference now that the satisfaction of our now-for-now and then-for-then preferences should be maximized.'[29]

As I understand, prudence consists in satisfying my now-for-now and then-for-then preferences as far as possible, in proportion to their strength. Now-for-now preferences are preferences I currently have for what should happen at present, and then-for-then preferences are preferences I currently believe I will have at some time in the future, i.e. they are preferences I will have at that time for what should happen then. These future preferences may be represented in the present to generate what Hare calls resultant or surrogate now-for-then preferences, i.e. preferences I currently have for what should happen at the appropriate time in the future. Ideally now-for-now and surrogate now-for-then preferences should be compatible. But, they may conflict with other current preferences, particularly strong antecedent now-for-then preferences, i.e. preferences which bear little or no relation to other current and future preferences, and which, if satisfied, may lead to the non-satisfaction of my expected future preferences.[30] An example may help to make this clear. It is, for instance, conceivable that someone may now strongly prefer that x should then (in the immediate future) happen, though then he shall prefer that x not happen. The heroin addict knows that, should the pusher call again in a day or two, he will then want a shot. He now prefers that, should he then be in the grip of a craving for the drug, he be allowed a shot. But, he also knows that then he will prefer not to have succumbed, for then his fight for recovery will have taken a step backwards. As I understand Hare,

satisfying the antecedent current preference to be allowed a shot at the expense of the future preference that the shot be withheld, is a sign of imprudence, even if the former preference is the stronger one.[31]

It seems that we should understand prudence as requiring that strong antecedent current preferences be adjusted or modified in such a way that they no longer conflict with expected future preferences.[32] This works on the idea of putting oneself imaginatively in one's own future shoes. Thus, because the heroin addict now fully represents to himself the state of affairs he wants to realize in the future, he generates a surrogate current preference that the shot be withheld, which will be of the same strength as the corresponding future preference. The generation of these surrogate preferences helps to block imprudent actions, because the representation in the present of preferences for future states and events prevents us from discounting them simply on grounds that they are for the future, which Hare believes is irrational. (This is distinct from discounting because of unpredictability, which Hare thinks is rational.)[33]

One critical point Hare is trying to drive home regarding prudential calculations is that what is currently most strongly preferred may not yield maximal preference-satisfaction, in sum, over time, and hence that the requirement of prudence may demand that current preferences be adjusted or modified in the light of future preferences.[34] If Hare is right, recourse to the requirement of prudence is properly called for when strong preliminary or antecedent current preferences conflict with future ones.[35] A rational course of action is then determined by the balance of one's <u>current</u> preferences once all preferences in one's preference set have been exposed to all relevant facts and logic.

I shall attempt to explain what full exposure to all relevant facts and logic involves by constructing a case of preference conflict over time. I hope to illustrate Hare's view that where more than one rational alternative is available, prudence may point in different directions.[36] Consider the following case. A, the aspirant medical student, now prefers then (once matriculated) not to go to the army, since that will delay starting his university career. But this means that he will have to go as soon as he has graduated, which will then be particularly unwelcome, because then he will prefer to start his professional career.

The first preference is a preliminary present preference; he now prefers then (once matriculated) to be pursuing his university career, rather than being in the army. The second is a then-for-then preference; he will then prefer (once graduated) to be pursuing his professional career, rather than going to the army. His question is: what will <u>overall</u> be the best course of action to adopt, given that he believes he ought to do his military service?

Suppose that, of the two preferences, the latter is marginally weaker. Following Hare, one way of proceeding would be for A to ask what his present preferences would be if he exposed them to all relevant facts and logic.[37] If he fully represents to himself his then-for-then preference, without discounting, he will have two current preferences, the preference that x (going to the army) should then - in the immediate future - not happen, and the preference that x should then - some years later - not happen. Since the latter is weaker, the rational choice will be that he should delay going to the army.

But, as Hare observes, there is a different way of proceeding which does not yield to the same result.[38] A might consider that when the time comes for x to happen, his present preference that x then not happen will then be something of the past. Accordingly, he might decide to ignore this preference, and seek instead to maximize the satisfaction of other present and future preferences. This means that the balance of his present preferences will come down in favour of not delaying going to the army.

Hare suggests that what will be for A the rational course of action depends strongly on what he is really after.[39] Thus, if A thinks that, once matriculated, a delay in making a start to his university career will be the greater evil overall, the first option will prevail. On the other hand, if he thinks that, once graduated, a delay in making a start to his professional career will be the greater evil overall, the second option will prevail.

Now, as I understand Hare, if the relative strengths of A's preferences are as assumed, he should select the first option in order to secure maximal preference-satisfaction, in sum, over time. It is worth noting that Hare (following Brandt) favours the first option, for he holds that

> '... we shall in any case do what the balance of our present preferences requires ...'[40]

after exposure to facts and logic, and there may be cases in which preliminary now-for-then preferences may survive this exposure, which means that they need not be imprudent.[41] Accordingly, I take Hare to be saying that a charge of imprudence will stick to A if and only if, on either option, what A currently most strongly prefers (as determined by the

balance of the relative strengths of the preferences in his preference set), will not yield maximal preference-satisfaction, in sum, over time.

Two final observations are here in order. First, it need not always be the case that current preferences should be changed or adjusted. Hare does not mean to say that future preferences should never be overridden.[42] Let us grant that it is in order, at the time one decides which preferences are in one's own best interest to satisfy, to make allowance for possible changes in one's future preferences. A may, for instance, envisage the possibility that, once graduated, his aversion to x then happening will have become very weak relative to his current preference. We might suppose that, as graduate, he will receive an army post and the sort of experience much to his liking (e.g. as medical officer), which has the effect of boosting the strength of his current preference. If, then, his future preference is likely to be as weak as I am now supposing, choosing to frustrate it need not be imprudent. For, though satisfying the strong now-for-then preference first to complete his university career will have results which he will then prefer not to obtain, he will then not be faced with a case of serious preference-frustration, because the non-satisfaction of his weak future preference counts for less in the overall preference pattern.

Second, the request for prudence is best understood as making a demand on an agent's choice of preferences at the time he chooses.[43] This involves a demand on his motivation at <u>this</u> time, which may involve making adjustments to his future preferences.

Here I wish to note that the request for prudence extends to reversed-role situations. On the motivational interpretation of the request for prudence, I am required, before I am allowed to universalize my prescriptions, to

criticize my own and other people's preferences by asking whether they have a claim to what is best for myself and others. The significance of criticism from the perspective of first-personal rationality is, as mentioned above, that I am expected to identify only with the prudent preferences of others, and to admit only my own and other people's prudent preferences for consideration at the time I choose the prescription I think I can universally endorse. Hare presents this view as a simplifying assumption. He says:

> 'We are to assume, when we come to universalize our prescriptions, as morality demands, that we have to consider only those prescriptions and preferences of others which they would retain if they were always prudent in the sense ... defined.'[44]

Though this is Hare's preferred position, he allows that giving weight to imprudent (and evil or malevolent) preferences strictly in proportion to their strength is permissible because it is unlikely to have an effect on my final decision.[45]

Before proceeding to show how Hare's argument for act-utilitarianism works, it is necessary to take note of some methodological points which concern the fulfilment of the knowledge conditions on which the rationality requirements depend. Hare advises that we follow a direct route

> '... from knowledge of, and preferences regarding, my own present experiences, to knowledge of, and preferences regarding, what should happen to me in the hypothetical case in which I am to be, forthwith, put into the position of somebody else.'[46]

It is worth noting that this route is structurally similar to the procedure we follow in order to fulfil the knowledge conditions in the single-person case. As I judge, the crucial point is the contention, which is implicit in KR and ST, that an adequate representation of 'what it would be like' to be in your precise situation, consists not in my having an aversion to you suffering as you are, but rather in having an aversion to myself suffering like that.[47]

As I understand Hare, the direct route runs as follows:[48]

I must first

(1) determine all relevant preferences, my own and yours, all of which receive equal, initial consideration, and then

(2) determine, after criticism by facts and logic, in the context of the overall preference pattern, what I rationally ought to prescribe. This works on an application of the requirement of prudence which determines the final outcome on the basis of preference-strength alone, and on the assumption that imprudent (and evil) preferences will be outbalanced by other preferences.

Hare contrasts this with an indirect route.[49] On this alternative route, I must first

(3) determine your criticized preferences, which become the preferences I identify with and grant equal, initial consideration, along with my own prudent preferences, and then

(4) determine, on the basis of only our criticized preferences, what I rationally ought to prescribe. This, too, works on an application of the requirement of prudence to the set of criticized preferences.

The direct route requires that all relevant preferences must be criticized in the same context, as the preferences of a single person. This, unlike the indirect route, calls for only one application of the requirement of prudence to determine the final outcome on the basis of preference-strength alone. If Hare is right, no harm is done by including imprudent preferences in my deliberation, and granting them positive weight in proportion to their strength, for Hare thinks there never will be a case in which the sum of preference-satisfaction will be maximized by allowing this. Here I merely note that Hare's view is consistent with his rejection of an anti-utilitarian stance, i.e. the stance that imprudent (and evil or malevolent) preferences should receive no consideration, not even initial consideration.[50]

I shall illustrate an interesting point I wish to make about the direct route by means of a bi-lateral case of preference conflict. Suppose that A and B (the newly-weds) are considering starting a family. A $prefers_1$ to start a family now, but he knows that when they have a child he would $prefer_2$ not to have a child. Assume that $preference_2$ is stronger than $preference_1$. B $prefers_x$ to start a family now, but she knows that when they have a child she would $prefer_y$ not to have a child. Assume that $preference_x$ is stronger than $preference_y$, and that $preference_x$ is stronger than $preference_2$. Should they start a family now? Let A be the decision-maker.

Now, following Hare, A must adopt the following procedure. He must first identify all relevant preferences and then determine, after criticism by

facts and logic, what he rationally ought to prescribe on the basis of preference-strength alone. If Hare is right in his view that the inclusion of imprudent preferences is unlikely to affect the final outcome, then it should not matter whether A

(1) first sums like preferences, and then decides what the prudent course of action will be, or
(2) first sets aside imprudent preferences, and then sums like preferences in the subset of criticized preferences.

I said that recourse to the requirement of prudence is properly called for when strong preliminary or antecedent current preferences conflict with expected future preferences. Now, following (1), A will be led to the conclusion that starting a family now is the rational course of action. But it is obvious enough that the preference to start a family now is an imprudent preference <u>for A</u>. (2) yields the same result. When considered on their own as the preferences of two different persons, preference_1 and preference_y should be eliminated as imprudent. This leaves A with preference_2 and preference_x, and since preference_x is stronger than preference_2, he must rationally endorse preference_x. But <u>for A</u>, preference_x is an imprudent preference. I return to this point again in Section 2.

It is now time to show how Hare's argument for utilitarianism works. I shall present an outline of this argument, leaving my discussion of the justification of the various moves in it for the next section. The argument proceeds via an attempt to establish the principle of equal interests (PEI), which Hare regards as a crucial premiss in his argument. Hare says:

'It is in accordance with our method to assign equal weight, strength for strength, to all preferences alike, provided that they survive exposure to logic and the facts.'[51]

I begin by noting two things. First, as I understand the argument, we proceed by first satisfying KR and the knowledge conditions on which the rationality requirements depend. As I read Hare, the satisfaction of these things is independent of the constraints of universalizability. Universalizability enters the argument only once the relevant facts of the situation we are judging have been ascertained.[52] Second, the argument for act-utilitarianism then proceeds via appeal to the PEI and Hare's argument for summing and balancing in bi-lateral and multi-lateral cases of preference conflict.

I shall use the 'dentist'-example to illustrate my points. Given that A (the dentist) knows that B (the patient) prefers that x (i.e. that the drilling into his decaying tooth be stopped immediately), A acquires, by KR, the following preference.

(a) A prefers that x in the case in which he imagines that he is in B's precise situation.[53]

Now, it is a condition of fulfilling KR that A must have equal aversion to suffering as B is suffering, were he to be in B's precise situation. I have said that ST defines the bottom line of what shall count as knowing the relevant facts of another person's situation. If he really knows the facts of B's situation, he must, by ST (and the view that to have a preference is to prescribe that it be satisfied) prescribe that the preference for x be

satisfied, albeit only initially, and in proportion to its strength. A would then, by these things, be committed to the following move.

(b) A prescribes that the preference for x be satisfied in the case in which he imagines that he is in B's precise situation.

Regarding (b), it is important to recall that for Hare KR and ST are conceptual truths, deriving from the thesis of the prescriptivity of 'I' and the conceptual claims Hare makes regarding our knowledge of other people's preferences. It is conceptual truth that if he knows B prescribes x, he must be prescribing the same thing, and with the same intensity, in the case in which he imagines that he is in B's precise situation.[54] (b), of course, may be viewed as a requirement of A's rationality (R above). Rationality requires that he must assent to B's prescription 'Let it be the case that x', for his assent is the test of whether he really knows the facts of B's situation. Now, as I judge, Hare makes R depend on KR and ST, and in particular on the truth of the conceptual claims he makes regarding the relation between knowing that someone is suffering and (actually) experiencing this suffering.[55] It is also worth noting that the move to (b) does not depend on universalizability. For Hare maintains that a knowledge of the relevant facts of B's situation is a condition of rational assent to his prescriptions, and indeed, to all prescriptions, whether singular or universal.[56] If this view be accepted, there is no need to appeal to universalizability to justify the move from (a) to (b). I return to this point again in Section 2.

Now, given the relation Hare sees between prescriptivity and preferences, A can be said to accept a prescription only if he prefers to do what the prescription enjoins him to do. Hence it can be said that

(c) A accepts the prescription 'Let it be the case that x' only if he prefers that x in the case in which he imagines that he is in B's precise situation.

I said that universalizability enters the argument only once the request for correct factual information about B's situation has been met. Here I wish to note two points. First, for Hare it follows from universalizability that if A now thinks that he ought to keep on drilling into B's decaying tooth, he is committed to the view that the same thing ought to be done to him, were he in B's precise situation.[57] And since 'ought' is prescriptive, he would be committed to the following move.

(d) A prescribes that the preference for y (the preference, i.e. that his dentist keep on drilling into his decaying tooth), be satisfied in the case in which he imagines that he is in B's precise situation.

Given again the relation between prescriptivity and preferences, it can be said that

(e) A accepts the prescription 'Let it be the case that y' only if he prefers that y in the case in which he imagines that he is in B's precise situation.

Second, the key to understanding the force of universalizability in Hare's argument is that it requires that we treat other people's preferences 'as if they were our own'. This is essential to the basic strategy Hare employs for the resolution of bi-lateral and multi-lateral cases of preference conflict, which is the strategy of committing his decision maker to incompatible preferences.[58] Universalizability must then be thought to

constrain A to accept that the preference he has qua dentist, viz, the preference for y, and the preference he has qua patient, viz, the preference for x, are for him incompatible preferences. There is, if Hare is right, nothing logically amiss with A having incompatible preferences, but he cannot without logical inconsistency (i.e. self-contradiction) prescribe that both be satisfied.[59] Now, it is obvious that the prescription A must be accepting in (c) is inconsistent with the one universalizability constrains him to accept in (e). So, how might A proceed? Neither of A's preferences has a veto. Either may qualify as moral subject to a certain condition. This is that the qualifying preference has to be a universal preference, and it qualifies as such if and only if A prefers, on balance, that it be satisfied in each position he occupies qua dentist and qua patient.[60] The question then becomes: What constrains A to balance one preference against the other?

Two major constraints must here be noted. A must (1) grant equal weight, strength for strength, to all relevant rational preferences, and then, (2) assuming that the decision-procedure in the two-person case parallels a prudential decision in the single-person case, balance his preference for x against his preference for y. The argument for (1) has three legs. First, given that Hare is correct in maintaining that KR is a conceptual truth which rests on the thesis of the prescriptivity of 'I' and the truth of the conceptual claims Hare makes regarding our knowledge of other people's preferences, it follows that in using the word 'I' of himself in the reversed-role situation, he cannot without self-contradiction claim that he knows he is suffering, yet deny that he has no motive for ending this suffering. For, as I have understood Hare, he has to use the word 'I' of himself in the reversed-role situation, which means that he has to concede that some state of another person, B, might actually be a state of himself,

and that in conceding this he must concede that he is concerned about the satisfaction of the preferences he will have in that different state of himself. The point is simply that since to be in this state is to be in a state of suffering, he must know that he is suffering, and if he knows this, he must have a motive for ending his suffering. And, if he does not want the suffering to stop, or claim that he does not mind it, he cannot really be suffering.[61]

Second, I said that KR requires a strong reading of the sense in which A is expected to be concerned about the satisfaction of B's preference not to be made to suffer, one akin to the sense, appropriate in the sphere of prudence, in which a single person experiences concern for the satisfaction of his preferences. Now, I maintain that the motivational assumptions of KR run on an appeal to hypothetical self-interest in reversed-role situations.[62] I shall have more to say about this point in Section 2. Here I merely note that, if correct, we may say that as one who is rationally self-interested, A is expected to grant positive weight to his preference for x, and strictly in proportion to its strength, just as B would, if B were perfectly prudent.

Third, A must now prescribe, as his rationality requires (R above), that the preference for x be satisfied, albeit only initially, and strictly in proportion to its strength. But this he must do as if the preference for x were 'his own', as universalizability requires. The point is that since he has to make a decision concerning which prescription to universalize in ignorance of the fact that he plays the role which he in fact plays, and in ignorance of who is actually who in each of the positions he occupies - i.e. in ignorance of any information about particulars which enable him to identify a particular case as his own, there is an equal chance that any of

the cases which have been put up for consideration might actually be his own case, and therefore that all cases are deserving of the same consideration, the same consideration he grants to his own actual case.[63] If this is how universalizability constrains, and provided that he knows the relevant facts of B's situation - in the required sense of 'know', which includes knowing the strength or intensity of B's preference for x in that sense - he would be committed to the PEI.

Now, following Hare's direct method, A merely has to balance his preference for x against his preference for y in order to arrive at a final decision. As I judge, this move is justified by the requirement of prudence. If prudent, he will opt for that course of action which yields the greatest net sum of preference-satisfaction. And if his preference for y is his strongest considered preference, he must allow it to override the weaker preference for x. The indirect method yields the same result. It is obvious that we cannot read the preference A forms in the case in which he imagines that he is in B's precise situation, viz, the preference for x, as B's on balance or considered preference. This is clearly not what B would prefer, if he were fully informed and perfectly prudent. A must from this preference for x, for KR requires it. But KR is distinct from the requirement of prudence,[64] and so, qua prudent, he need not include this preference in his final set, which means that his preference, were he the patient, that his dentist should go on drilling into his decaying tooth in order that he may avoid toothache in the future, must win.

The resolution of multi-lateral cases of preference-conflict runs along the same lines. Hare says:

> '... in them too the interpersonal conflicts, however complex and however many persons involved, will reduce themselves, given full knowledge of the preferences of others, to intra-personal ones.'[65]

But the matter is not quite so straightforward. I shall use a modified version of Hare's 'car-bicycle' example to illustrate my points.[66] Suppose A, the dentist, wants to park his car in a convenient spot directly opposite the main entrance to his practice, but discovers that this spot (the only available one) is occupied by two bicycles belonging to two of his patients, B and C. They prefer to leave their bicycles exactly where they are and will mind if he moves them some distance to a vacant lot. Suppose also that A knows this, in the required sense of 'know', and assume that his preference to park his car in the occupied lot is stronger than their preferences to leave their bicycles where they are, taken separately, though not in sum. So, why should he sum their preferences before he balances in order to establish what he rationally ought to do?

How does Hare's argument for summing run? The demand for universalization requires that he accept a universal prescription for all relevantly similar situations. The relevant point to take is that being universal, the prescription has to apply in <u>one</u> logically possible world in which the individual identities of the person concerned, and the particular positions they in fact occupy, are irrelevant to the moral decision. So, let us accept that there is an equal chance that any of the three situations at issue might be the actual situation in which he is either the car-owner or one of the two cyclists,[67] and that he is constrained to allow that, whatever line of action he chooses to prescribe, the same thing should be done to himself <u>qua</u> car-owner and <u>qua</u> any one of the two cyclists. If we allow further that the things the two cyclists have good (prudential)

reasons for wanting, are also for him reasons to want the same things, as Hare requires, he can then, from the perspective of first-personal rationality, affirm that a particular line of action ought to be pursued if and only if its performance is likely to yield the greatest net sum of preference satisfaction overall. This presupposes a constraint to sum the preferences of the two cyclists. So, what might this constraint be?

We have to assume that we may weight the probability of any one of the three situations actually occurring in the world as it is in terms of the number of instances that they actually occur.[68] To this has to be added the strengths of the preferences which accompany the occurrence. Let us accept that to be fully informed involves knowing the probability of the situations occurring and the strengths of the preferences. Now, since there is but one situation in which A prefers to park his car, the probability of the other situations occurring is much higher.[69] This is to say, in effect, that A's preferences to leave the bicycles where they are, are in sum greater that his preference to park his car. And, if to be fully informed involves knowing how strongly he prefers to leave the bicycles where they are, he must sum the two preferences not to move them. The idea underlying the move to sum might be expressed as an assumption about his rationality. This is that just as it is rational, in the prudential sense of 'rational', for him to act on what he most prefers to do in his own actual case, so it is rational for him to act on what he most prefers in the case in which, in addition to his own preferences, he has other preferences which he treats 'as if they were his own' through an act of the imagination.

Section 2. A critique of the argument for utilitarianism

I have maintained that the argument for act-utilitarianism runs mainly on three things: KR, the request for prudence and universalizability. KR rests

on the thesis of the prescriptivity of 'I' and the conceptual claims Hare makes regarding our knowledge of our own future preferences, and our knowledge of other people's preferences. KR is independent of both universalizability and the requirement of prudence,[70] and is, as Hare maintains, an analytic or conceptual truth. It is a conceptual truth to say that I cannot use the word 'I' of myself in the hypothetical case in which I reverse roles with you or with my own future self, without currently preferring what I will prefer in these hypothetical cases. And it is a conceptual truth to say that, if I know I will prefer x in these cases, this entails that I now prefer x. For, if Hare is right, my knowledge of my own future states is not merely a cognitive matter.[71] KR includes a motivational constraint as a component of knowledge. This thesis, then, as Hare says, is

> '... analytic of the word "I" and of the word "knowledge" when applied to knowledge of preferences of other people or at other times.'[72]

Now, KR is the thesis Hare must be able to defend, for it plays a crucial role in establishing the general claim that I have to maximize the sum of preference - satisfaction, if I am to be rational. The important claims underlying KR are

(1) that some state of some person might actually be a state of myself, and
(2) that I am, in identifying this state as a state of myself, concerned about the satisfaction of the preferences I associate with being in this state of myself.[73]

In postulating the possibility of KR, Hare draws on the general point that any person who judges some person to be continuous with himself, will

presently desire that the preferences of that person be satisfied in proportion to their strength. I quote the passage in MT in which Hare offers this view. He says:

> '... it is probable that the word 'I' is attached to all identifying properties ... [bodily continuity, continuity of personal characteristics and linked memories] in varying degrees. I wish merely to suggest that to these we should add, not another identifying property, but another feature of the word 'I', namely that, by calling some person 'I', I express at least considerably greater concern for the satisfaction of his preferences than for those of people I do not so designate.'[74]

The thesis Hare is defending says simply that in identifying some person as myself, I do not merely affirm the (supposed) fact that continuity obtains, but also that I am concerned about the satisfaction of his preferences.[75] As I understand Hare's concept of personal identity, to suppose that all continuity requirements are satisfied, albeit only hypothetically through an act of the imagination, involves having this concern.[76]

Now, this thesis underlies KR and its subsidiary claims, as well as the role Hare assigns to the universalizability thesis. I have two main concerns: (1) to examine the defence of KR, and particularly the motivational assumptions which it implies. I shall accordingly have something to say about Hare's concept of personal identity. I begin with an analysis of how KR works in the single-person case, and follow this up with an analysis in a bi-lateral case. I shall then (2) attempt to assess whether there is a case for maintaining that a prudential rule may be

applied in the two-person case, and whether this move can constrain Hare's agent to maximize the sum of preference-satisfaction.

Let us return to the 'army'-case. I said that, qua rational, A, our aspirant medical student, should consider what it is he would be doing, if he chose one or other of the alternatives open to him. If he thinks that he should first go to university, he will not be able to start his professional career at the time he will want that to happen. If he rejects this course of action, this must be because of what delaying the start would be like for himself, i.e. some (supposed) fact about the anticipated experience, e.g., the fact that then he will not like it, supplies the reason for his rejection. Considering what it is he would be doing, then, involves taking cognizance of how that will affect what he will then want.[77] Hare maintains that he should do what he rationally most prefers at present, i.e. what will maximize the satisfaction of his present rational preferences, which are the ones he retains after criticism by facts and logic.[78] For Hare present actions must be compatible with what will maximize preference-satisfaction at other times, for otherwise his present preferences will not be rational in the prudential sense which Hare's theory requires.

Now, the thesis of the prescriptivity of 'I' plays a role in accounting for the possibility that a reason A will have to do something in the future can be a reason which presently constrains him to act in ways compatible with his own maximal welfare. According to this thesis, he uses the word 'I' of himself in the hypothetical situation in which he stands in his own future shoes, which he cannot do without currently preferring what he now believes he will then prefer and now prescribing that this preference be satisfied.[79] If he knows that there is something about an anticipated experience he will

then not like - say, because having the experience will cause him to suffer - he will now desire, i.e. have a motive that that experience be avoided. Hare's thesis is that there is a logically necessary connection between fully representing the anticipated experience to himself and presently desiring that that experience be avoided. The connection is, as he says, 'conceptual'. This is, as Hare puts it,

> '... because the experience that is being represented is a desire. If we try to represent to ourselves what it is like to have a certain desire, we have not succeeded unless there is something in our present experience to correspond to what we are trying to represent; and this has to be a desire too.'[80]

Hare's account of the rationality of present actions requires that he must satisfy KR as a step towards satisfying the principle which requires that he grant equal weight to all (considered) preferences of equal strength. KR itself, however, does not say which ones of the anticipated experiences it would be best for him to have, in a sense of 'best' compatible with what is likely to secure his own maximal welfare. This is strictly a matter for the requirement of prudence. It seems that a preference may be compatible with KR, but not be a preference that should be satisfied, all things told, i.e. it may not be one's considered preference for a particular contingency. So it seems that Hare is right in maintaining that KR is distinct from the requirement of prudence.[81]

But, we may ask: what constrains A to satisfy KR? Suppose Hare is right about personal identity. It then follows that if he has a reason at present to be concerned about the satisfaction of the preferences he envisages having in the future, this reason must be that the person who will have

those preferences will be himself.[82] We may, however, still ask: what reason do we have to believe that satisfying all continuity requirements involves having the suggested concern for the welfare of that future person?[83] Does Hare mean to say that the (supposed) fact of continuity itself has at present motivating force? If Hare is right, it follows on the assumption that all the continuity requirements of personal identity are satisfied, albeit only through an act of the imagination, that he has at present the required concern for the welfare of that person. As I understand Hare, we require as a condition of saying that he has the reason for the concern which this concept of personal identity demands, that all these requirements be satisfied. So, we have to assume, if the (supposed) fact of continuity has motivating force, that this must be operative at present. How might we account for this?

We need to suppose (1) that a distinguishing feature of A's present self, or of his present motivational set, supplies the reason for the required concern.[84] I think this works with another thesis which is closely linked to the thesis of the prescriptivity of 'I'. This is (2) Hare's concept of prudence. Hare says:

> 'To be prudent is to think of the future states of a certain person (normally the person whose body one's present body will be) as <u>oneself</u>, and thus to acquire a concern for the satisfaction of the future preferences of that person.'[85]

As I understand Hare, the point is that rational persons possess a strong degree of self-concern, and this we might express by saying that they want to satisfy their own preferences to the greatest extent possible. If this is correct, we might then say that anyone who uses the word 'I' to identify

some person as himself, and subject to the condition of its use with reference to that person, exhibits rationality in the required sense (though this is itself no guide to what he had best at present do to secure his own maximal welfare). If this be granted, we make room for the conceptual link Hare sees between fully representing to himself anticipated future experiences, and presently desiring what he will then want, to come into play.

Now, if I am right, allowing that continuity obtains does itself not constrain him to form a current preference compatible with what he will have reason to do or to avoid in the future. If, as we are assuming, that (1) and (2) serve as the main motivating factors, he has a reason, all other things being equal, to satisfy KR. Now, might we say that if he has a reason to satisfy KR, he also has a reason to satisfy the principle which demands that he grant equal weight to all (considered) preferences of equal strength?

Consider the following case. Suppose that A is strongly anti-conformist. At present, t_1, he prefers to serve time in prison rather than in the army. But he knows that at some time in the foreseeable future, t_2, he will not prefer to serve time in prison, i.e. that then he will have a preference he now regards as contemptible, rather to do his military service.[86] Let the preference to serve time in prison be x and the preference to do his military service be y. At t_1 he prefers that the preference for y be frustrated, even though he now believes that, should he prefer y at t_2, he will not then have the preference that y be frustrated, because then the preference for y will be very strong.

Suppose further that he thinks of the person who will have the preference for y at t_2 as himself. Now, what reason do we have to believe that regarding that person as himself involves having the required concern for his welfare? It seems plausible to say that if he knows he will want to do things in the future which he at present regards as contemptible, he will not now form preferences for those things.[87] And, if this be granted, we may ask what reason we have to believe that the (supposed) fact of continuity supplies a compelling reason why he should now form a preference that the preference for y be satisfied.[88]

Now, the obvious objection is that unless he satisfies KR, he does not exhibit rationality in the sense which Hare requires. If Hare is right, he cannot be said to know the facts of his own case, in a sense of 'know' which requires that he understands the circumstances in which he would have the preference for y, unless he now forms a preference that the preference for y be satisfied.[89] To be fully rational, he must know how strongly he would prefer that y, and how much he would then suffer if the preference for y were to be frustrated, and this he cannot know unless he represents these things to himself in the way KR requires. This means that he must now be equally averse to suffering as he would then be. As I pointed out above, Hare maintains that the experience he imagines to be having in the circumstances in which he would prefer that y, is a desire which has to be present in his current experience, if he is fully to understand those circumstances. So, now forming a preference that the preference for y be satisfied is rationally required.

Suppose, then, that he puts himself in his own future shoes, and suppose he understands the circumstances in which he would then prefer that y as follows. At t_2 he will be tempted, qua self-interested agent, to prescribe

that his preference for y be satisfied because then the preference to conform will be very strong. But I see no reason to accept the view that representing certain experiences to himself is simply a matter of desiring that certain actions be performed or not be performed. It is not implausible to suggest that if the person he envisages becoming does not resemble him in those respects he most values at present, he has a reason at present to disallow that the preferences of that person are worthy of satisfaction.[90] And, if (1) above is correct, it is not implausible to suggest that at present he lacks the motivation to desire that the preference for y be satisfied. Moreover, what reason does he have to accept that the strength of the preferences which will obtain at t_2, and not their content, should be decisive in deciding which preferences to allow to have an influence on present actions?[91] We may ask: are there sufficient grounds for allowing that he is rationally required presently to desire that the preference for y be satisfied? We may ask again: why should he now care whether at t_2 he will strongly prefer that the preference for y be satisfied, even though he now imagines and fully understands the circumstances in which he would prefer that y, just as the person he may become will then prefer that y will care and understand?

We may note that his current preference that the preference for y be satisfied is an antecedent now-for-then preference which the requirement of prudence might disallow.[92] It is also a preference corresponding to an ideal. We may, following Hare, call it an autofanatical preference. If Hare is right, anyone who is willing to allow strong antecedent now-for-then preferences to override strong and fully represented then-for-then preferences manifests a form of imprudence. Hare says:

'We might call him an autofanatic, because of his similarity in the sphere of prudence, to the fanatic in the sphere of morality.'[93]

The mark of the autofanatic is this. His current preferences are so strong that, though he thinks of the person he may become as himself, he is still prepared to prescribe that that person's preferences be frustrated. It is obvious that this runs counter to the requirements of KR, and this is perhaps the reason why Hare favours the exclusion of autofanatical preferences.[94] Hare does not say that these preferences are irrational. His point is rather that they are not rationally required in that they are not compatible with KR. We may ask whether they are also incompatible with the requirement of prudence. Appeal to the overriding strength of the current preference we should have that the satisfaction of our now-for-now and then-for-then preferences be maximized, is properly called for when strong antecedent now-for-thens conflict with strong and fully represented then-for-thens. But, it seems that in the case under consideration it does not help to appeal to this requirement, for the appeal will work only for antecedent now-for-thens which are not so strong that they become incompatible with the requirements of KR.

Now, I wish to draw attention to a point well worth considering. As I have understood Hare's theory of prudence, a charge of imprudence will stick to A if and only if what is currently most strongly preferred - as determined by the balance of all (considered) preferences in his preference set - is not likely to yield maximal preference-satisfaction, in sum, over time. The question then is how this might be established if KR is not satisfied. Hare's account of the rationality of present actions requires that KR be satisfied as a step towards satisfying the principle which demands that he must grant equal weight to all (considered) preferences of equal strength,

irrespective of their content. The reason Hare thinks is available to satisfy KR is not independent of the supposition that continuity obtains, and this supposition is necessary to account for the rationality of present actions. Now, we may ask: what theory of prudence requires that present aversions, of the kind I have been considering, should count for nothing in making prudential decisions? A could retain his current preference that the preference for y be frustrated, and still be rational.

I shall now attempt to show how KR works in a bi-lateral case of preference-conflict. I use the 'newly-weds' case to illustrate my points. Suppose again that A and B now prefer to have a child (A with strength 1 and B with strength x). But they realize that once they had started a family, they would prefer not to have done so (A with strength 2 and B with strength y). Let the preference to start a family be F and the preference not to be F^-. Suppose that F^-_2 is stronger than F_1 and that F_x is stronger than F^-_y. Suppose also that F_x is stronger than F^-_2. We might now reconstruct the critical steps in the initial stage of Hare's argument thus: A cannot defend denying B the satisfaction of starting a family now unless he knows what it is he is doing. Knowing what he is doing necessarily involves knowing what B is experiencing, i.e. what it is like for B to be denied this satisfaction. Suppose that this is causing B to suffer. Hence, knowing what it is he is doing involves knowing what it is like for B to be suffering. He cannot know this without knowing that it involves wanting the suffering to stop. The linguistic expression of this is 'Let it be the case that B not be denied the satisfaction of having a family.' If A uses the word 'I' of himself in the reversed-role situation, which he cannot do without now having the preference that, were he to be in B's precise situation, he should not be denied the satisfaction of having a family, he must prescribe that this preference be satisfied, at least initially, and

strictly in proportion to its strength. This involves accepting the prescription 'Let it be the case that I not be denied the satisfaction of having a family'. Accepting or assenting to this prescription is a requirement of his rationality (R above). If he does not assent, so Hare claims, he does not know the relevant facts of B's situation, which means that his final decision would be irrational, i.e. it would be faulted for lack of information.

We may note that F_x is what B would prefer, if she were prudent and fully informed about the facts of her own case. As I understand Hare, the preference A forms in the case in which he imagines himself to be in her precise situation must conform to her considered preference. A must then form the current preference (which is surrogate or resultant now-for-then preference), that were he forthwith to be in her precise situation, F_x be satisfied. However, if we consider A's preferences on their own, the current preference that F_x be satisfied, is <u>for him</u>, an imprudent preference. As I pointed out in section 1, rationality requires, all things told, that he prescribe that it be satisfied. We may ask: Should he not, <u>qua</u> prudent, reject it on grounds that he has no reason which he can square with rational self-interest to prescribe its satisfaction? This obviously runs contrary to R.

But, there is an assumption in this requirement of rationality which is well worth considering. This is that A has no other preferences which may serve as reasons presently to desire that F_x be frustrated. Suppose A believes that wives ought to obey their husbands because they are wives. Let the preference that his wife obey him be z. Suppose further that z corresponds to a long cherished personal moral ideal, and that A is so strongly attached to z that he is willing to prescribe its satisfaction,

even in the hypothetical situation in which he would be the person who would be adversely affected. Should z be allowed to count?

Now, Hare maintains that external preferences, i.e. preferences for states of affairs which are not currently within the experience of the preferrer, should be excluded.[95] The reason is that they are not compatible with the requirements of KR.[96] External preferences include personal moral ideals of the kind which is here being considered.[97] Hare could simply rule z and all other kinds of external preferences out of court for the reason given, and accept that there are limitations to what kinds of preferences his theory can accommodate.[98] But this move would leave Hare without a way of dealing with 'fanatical' preferences, which he wishes to avoid. Z must be allowed to count and be treated as just one preference among those which enter into moral thinking.[99] There are two points worth noting here. A has to treat his ideal as just one preference among others,[100] and he has to form the current preference that F_x be satisfied in ignorance of the facts of his own actual case.[101] The latter is a condition of <u>full</u> representation, and is independent of universalizability.

Now, if the moves which commit A to KR are all conceptual moves, he cannot refuse them. If he is committed to KR, he must prescribe, albeit only initially, that F_x be satisfied. He cannot then legitimately evade forming this preference. Indeed, he is expected to form this preference, not <u>qua</u> the person with his individual identity and name, but rather <u>qua</u> 'I', as one who identifies the person in the reversed-role situation as <u>himself</u>.

We may note the following: First, the crucial point about KR is that it is a conceptual truth which can be defended independent of universalizability. Hare says:

> 'I claimed that it is a conceptual truth that if I know that I would be prescribing something were I in exactly someone else's position with his preferences, I must now be prescribing the same thing with the same intensity.'[102]

Hare notes that in establishing the possibility of KR, we may appeal to universalizability, to a 'limited extent', to show that rational decisions in moral thinking depend on our knowledge of the relevant facts of a situation.[103] However, and this is the crucial point, we may dispense with this appeal if we accept the view

> '... that all prescriptions, universal and singular, have to be made in cognizance of the facts if they are to be rational.'[104]

Fulfilling the requirement of correct factual information about B's case is clearly a condition of rational assent to her prescriptions. I indicated above that ST defines the bottom-line of what is involved in this request for information: A is not in possession of this information unless he himself has a preference, were he in her precise situation with her preferences, that these preferences should be satisfied. This is to say, the bottom-line consists in himself being averse to suffering as B is suffering, and being concerned, just as B is, about the satisfaction of the preference to start a family now.

Second, I indicated above that Hare's concept of personal identity underlies KR. A cannot use the word 'I' of himself in the reversed-role situation without currently preferring what he would then prefer, given that he would then be situated just as B is in all relevant universal respects. And, in using the word 'I' of himself in that situation, he

acquires a concern - albeit only a hypothetical concern - for the satisfaction of the preferences he would then have. As I understand, the sense of his concern must be akin to the sense - appropriate in the sphere of prudence - in which a single person experiences concern for the satisfaction of his (future) preferences. However, we may ask: what constrains A to use the word 'I' of himself in the reversed-role situation? We may also ask, on the assumption that continuity obtains, what reason we have to believe that this (supposed) fact supplies the required motivating force at present? For, as I understand Hare, in the reversed-role situation the satisfaction of Fx must be, for him, a matter of self-interest, albeit only hypothetical self-interest. If this can be shown, he would be committed, qua rational in the prudential sense which Hare requires, to prescribing that, were he in her precise situation, his suffering should stop.

It is a crucial assumption of Hare's argument that the same person must be thought to occupy different positions, qua author of a prescription and qua recipient of it, albeit only at different times. The identity-requirements of Hare's theory demand that the similarity between actual and hypothetical situations be extended not only to the persons in these situations, but also to the preferences they actually have in these situations, and this means that the required identity does not obtain unless A has equal aversion to suffering as B is suffering. If Hare is right, A need - initially - form only a conditional preference in the case in which he imagines himself to be in B's precise situation. And he need - initially - only acquire a hypothetical concern for the satisfaction of this preference, as KR requires. He must then, because of universalizability, turn this hypothetical concern into an actual concern for the satisfaction of that preference.[105]

So, on the assumption that treating an ideal as just another preference presents no obstacle to satisfying KR, R will constrain A presently to desire that F_x be satisfied, albeit only initially, and strictly in proportion to its strength. But, allowing Z to count is unfair to B. How might Hare remove the counter-intuitiveness of allowing this external preference to count? The obvious answer is that B's ideals must also count as preferences. Should A then, in the case in which he imagines himself to be in B's precise situation, form the current preference that F_x be satisfied, or the current preference that wives be treated as their husbands' equals, or both? The development of Hare's position if he were to admit external preferences would seem to require both. But doing so could generate significant unfairness to B. So perhaps Hare ought to exclude all external preferences and find some other way of dealing with 'fanatical' preferences. I have raised the problem of external preferences because they present an obstacle to satisfying KR. There are, however, more serious obstacles. I shall have something to say about these towards the end of this section.

Now, Hare's case that KR is a conceptual truth which may be defended independent of universalizability rests on the presumed truth of his concept of personal identity. The thesis that 'I' is not wholly a descriptive word but in part prescriptive, says Hare,

> '... is the thesis to the extent that I know what it is like for a certain person to be prescribing or preferring something in his situation, and identify hypothetically with him (i.e. think that I might be he, which involves prescribing that if I were, his prescriptions, which would in that hypothetical case have become mine,

should be satisfied), I shall prescribe that those prescriptions should in that hypothetical case be satisfied.'[106]

This, says Hare, is a tautology. This thesis follows from the prescriptive element in the meaning of the word 'I', when combined with the requirement that hypothetical situations must be identical to actual ones, which identity must extend to the persons and the preferences they have in them. Does this mean that Hare defends the role-reversal procedure as a requirement of rationality and not as a requirement of universalizability?[107] But, though the thesis of the prescriptivity of 'I' is a tautology, it is not, thinks Hare, a trivial one. For it - combined with universalizability - explains how we advance from prudence to morality.[108] We advance from prudence to morality in virtue of being constrained to balance all considered competing preferences against each other on the basis of strength alone. The strategy Hare employs to achieve this requires that all these preferences be regarded as the preferences of a single person. However, it seems to me that in postulating this thesis, Hare assumes

(a) that the decision-procedure for the two-person case parallels the decision-procedure he advocates for the single-person case, i.e. that appeal to a form of rational prudence is the appropriate response in the two-person case.

In assuming (a), Hare also assumes a theory of personal identity which yields the judgements

(b) that A may have two sets of universal features, albeit in different positions and at different times, and

(c) that A has, in virtue of the (supposed) fact of continuity, a reason to be concerned about the welfare of the person he envisages becoming.

(b) and (c) are closely linked through Hare's concept of personal identity. I begin my discussion of these assumptions by asking: what constrains A to balance all considered preferences on the basis of preference-strength alone? The answer is that he is expected to think of the person in the reversed-role situation as himself, and that this involves, in virtue of the (supposed) fact of continuity, having the required concern for the welfare of that person. If I am right in this, Hare appeals to self-interest, albeit only hypothetical self-interest. Hare maintains that unless A accords equal weight, strength for strength, to B's considered preferences, he is either failing fully to represent what it would be like for himself to be suffering as B is suffering, or he is not really thinking of the person in the reversed-role situation as himself, i.e. he has not become fully knowledgeable and rational in the prudential sense which Hare requires.[109]

I shall now argue that applying a prudential rule in the two-person case demands more than what the supposedly parallel requirements of applying a prudential rule in the single-person case demand, and that the conceptual claims Hare makes regarding our knowledge of other people's preferences are false. I begin by noting two requirements of A's rationality.

1. Rationality requires that he presently prefer and prescribe, albeit only initially, that the preferences he will have, if his position were to change, be satisfied in proportion to their strength.
2. Rationality requires that he act on what he most prefers, all things told, in the case in which he has, in addition to his own preferences,

other preferences which he treats as 'his own' through an act of the imagination.[110]

I have mentioned Hare's view that universalizability enters his argument only once KR has been satisfied to constrain A to turn a merely hypothetical concern into an actual concern for the satisfaction of a preference he is expected to deem 'his own'. In this second stage of the argument, Hare's position rests on the possibility of A being able to see himself as one who has a set of incompatible preferences, and as rationally required to maximize the sum of preference-satisfaction. This, of course, assumes that he will treat other people's preferences no differently from his own, if he has no reason which satisfies the constraints of universalizability to treat them differently. Hare says:

> 'It follows from universalizability that if I now say that I ought to do a certain thing to a certain person, I am committed to the view that the very same thing ought to be done for me, were I in exactly his situation, including having the same personal characteristics and in particular the same motivational states.'[111]

I formulate two crucial requirements of universalizability as follows. Universalizability requires that A prescribe for the situation in hand

3. irrespective of the fact that <u>he</u> is the person whom he in fact is, because morality demands that moral judgements be accepted indifferently as regards <u>who</u> he is[112], and
4. irrespective of <u>the fact</u> that he prefers one thing for an actual situation and another thing for a hypothetical situation, because

morality demands that moral judgements be accepted indifferently as regards this fact.[113]

There are two points worth noting here. First, there appears to be some overlap between the requirements of the thesis of the prescriptivity of 'I' and the requirements of universalizability. Hare makes it clear that the former does not, by itself, entail the latter.[114] Universalizability enters the argument to compel A to treat merely hypothetical concerns as actual and 'his own'. But, if this is correct, we may ask: what prevents A, in the initial stage of Hare's theory, from refusing to use the word 'I' of himself in the reversed-role situation, and hence to reject F_x on grounds that it is an imprudent preference for <u>him</u>? The constraint Hare has in mind cannot be due to universalizability, for KR is independent of this thesis. The answer is that A must be thought to be constrained by his rationality, as 1 above formulates this. 1, in effect, is the requirement of correct factual information about B's case. 1, then, assumes that A can imagine himself to be B, and that the knowledge Hare thinks is available to him is knowledge he can have <u>in propria persona</u>, i.e. from the perspective of the first-person.[115] But 1 is not quite enough. For we may ask: what reason do we have to believe that the (supposed) fact of continuity has at present motivating force? It seems that we must here invoke the idea that a distinguishing feature of his present self, or of his present motivational set, supplies the reason for the concern he should have at present for the satisfaction of F_x. We have to suppose that Hare's argument for uitlitarianism begins with an appeal to rational prudence, and that this appeal translates to an appeal to hypothetical self-interest in the reversed-role situation. If this is right, KR and ST must then be thought to follow from the thesis of the prescriptivity of 'I' and the appeal to rational prudence, when combined with the requirement that hypothetical

situations must count for as much as actual ones. It is worth noting again that all this assumes (a) above, i.e. that appeal to a form of rational prudence is the appropriate response in the two-person case.

Second, if Hare is right, universalizability plays a role in constraining A to maximize preference-satisfaction. Universalizability constrains A, (a) to treat all (considered) preferences, irrespective of their content, as equally worthy of satisfaction,[116] and (b) to treat B's reasons for wanting to start a family now as 'his own' reasons. He cannot then fail to give positive weight to B's preferences in his final deliberation, for this constitutes a breach of universalizability.[117] But, we may ask: what is there in universalizability or in the thesis of the prescriptivity of 'I' to constrain A to accept (a)? The decision procedure for the two-person case is supposed to run parallel to the decision-procedure for the single-person case. But, we may ask: is there a rational requirement in the single-person case to the effect that all (considered) preferences be accepted indifferently as regards to their content? If Hare is right, A will be constrained, in virtue of universalizability, and qua rationally self-interested agent, to sum preferences and to balance the result on the basis of preference-strength alone, and the outcome of this, Hare believes, will maximize preference-satisfaction for all concerned.

Now, I have said that Hare's concept of personal identity underlies his defence of KR and its subsidiary claims, in particular ST, and that these claims underlie the role Hare assigns to the universalizability thesis. If this is correct, we may formulate the central claim Hare makes regarding the move to maximize preference-satisfaction as follows. The claim is

5. that A, by entertaining the thought that some person might be himself, which involves regarding some state of that person as his, acquires a concern for the satisfaction of that person's preferences, which then - because of universalizability - he treats as on equal par with his own (present) concerns.[118]

Now, the application of universalizability generates the following claim which has a bearing on the move to sum and balance. This is

6. that just as it is rational for A to act on what he, prudentially speaking, most prefers should be done in his own actual case, so it is rational for him to act on what he most prefers should be done in the case in which he has, in addition to his own preferences, other preferences which he treats 'as if they were his own' through an act of the imagination.

It is worth noting here that 6 is problematic. We may ask: what is there in universalizability which requires that a form of rational prudence be adopted in a two-person case of conflict? 2 is not independent of universalizability. The question arises: what constrains A to sum and balance on the basis of preference-strength alone, as 2 requires? In what follows I shall attempt to establish whether we are committed to 5 and 6, i.e. whether there is a rational requirement to the effect that A must maximize preference-satisfaction if he is to be rational.

Let us return to the 'army'-case. If Hare is right about personal identity, A cannot use the word 'I' to identify the person standing in his own future shoes without currently preferring that the preferences he will then have

be satisfied. Identifying that person as himself implies that he has at present the required concern. We might formulate this as follows:

7. If A knows that some state of some person will be a future of himself, he has a reason at present to be concerned about the satisfaction of the preferences he will have in that future state strictly in proportion to their strength.

7 may be read as a requirement of A's rationality, i.e. of the request for correct factual information about his own case. If Hare is right, there is a rational requirement to the effect that he must presently desire that future preferences be satisfied strictly in proportion to their strength. But, 7 does not follow from Hare's concept of personal identity alone. 7 assumes that the (supposed) fact of continuity has at present motivating force. So, to establish 7, we have to suppose that the required concern for the welfare of the person he will become obtains at present, i.e. we have to suppose that there is nothing about A's present self or his present motivational set which may serve as a reason presently to desire that his future preferences be frustrated, and that A is rational in the prudential sense. We need not suppose that an identity-related problem arises to see that the (supposed) fact of continuity may at present lack motivating force.

Now, 7 does not say <u>which</u> preference he should satisfy to secure his own maximal welfare. This is a matter for Hare's requirement of prudence. Hare thinks that if A acts prudently, he will do what he presently believes is likely to maximize the satisfaction over time of only his strongest rational preferences. Hare makes clear that whatever is rationally preferred at any time always depends on what the alternatives at that time

are, and that a prudent choice is always a choice between rational alternatives.[119] So, 7 should be reformulated to read thus:

8. If A knows that some state of some person will be a future state of himself, he has a reason presently to desire the satisfaction of the rational preferences he will have in that future state strictly in proportion to their strength.

But, 8 says nothing about the (psychological) causes or the moral legitimacy of the preferences he presently believes he will have in envisaged future states of himself. We may ask: is there a rational requirement to the effect that future preferences and prescriptions should be accepted indifferently as regards to <u>what</u> is preferred or prescribed? Hare's thesis of the prescriptivity of 'I' does not address this question. And, even if we allow that he is rational in the prudential sense, we still have no reason to accept that the content of future preferences should play no role in deciding which preferences to admit for consideration. One crucial question is whether there is a requirement in the single-person case that all considered preferences should be accepted for consideration indifferently of their content. It seems to me that whether one will have reason currently to desire that preferences one will have at other times be satisfied or frustrated, is largely a contingent matter. It is plausible to suggest that such reasons arise from a distinguishing feature of ones present self, or from a distinguishing feature of one's present motivational state, which at present (or at any given time which is present) one singles out to identify the person about whom one is concerned, and which is the reason for this concern.[120] Now, there seems no reason to accept that one is committed to prescribing that one's future preference be satisfied indifferently of their content if at present one

has reason not to be sympathetically concerned about the person one envisages becoming. But, I have assumed that he does not lack this concern. If then, A is seeking that course of action which is likely to maximize the satisfaction, in sum, of his rational preferences over time, he has a reason to weigh all considered preferences against each other on the basis of strength alone, i.e. he has a reason to satisfy the principle which requires that equal preferences should count equally, irrespective of their content. I am suggesting that in seeking to maximize the satisfaction, in sum, of his rational preferences overtime, A follows the following prudential rule.

9. Prudential rule: If A at present believes that he will have preferences in some future state of himself which promote or serve the distinguishing feature of his present motivational state, he has a reason presently to desire that they be satisfied strictly in proportion to their strength.

Now, matters are less straightforward in the 'contemptible-desire' case. We may ask: what reason does A, the non-conformist, have to accept 7 and 8? If I am right, there is nothing in 7 or 8 to compel him to accept that all his preferences are equally worthy of satisfaction. It is not implausible to suggest, as I have done, that if the person he envisages becoming does not resemble him in those respects he most values at present, he has a reason at present to disallow that the preferences of that person are worthy of satisfaction. If this is in order, the following rule serves to qualify 7 and 8.

10. If A at present believes that he will have preferences in some future state of himself which will generate reasons for action he cannot at

present accept, he has a reason presently to desire that those preferences be frustrated.

If I am right in claiming that Hare supplies no good reason to suppose that present aversions of the kind we are considering should count for nothing in making prudential decisions, then the following rule serves to qualify 9.

11. If A at present believes that he will have preferences in some future state of himself which promote or serve the distinguishing feature of his present motivational state, he has a reason presently to desire that they be satisfied in proportion to their strength and indifferently of their content - otherwise not.

The point is not that it would be irrational for him presently to desire that those preferences be satisfied, but rather that they do not at present generate reasons for action (i.e. at present they lack motivating force). It is worth noting again that 11 does not imply that A has no reason to satisfy the principle which demands that he grant equal weight to all (considered) preferences of equal strength. I suggested above that he satisfies this principle whatever he decides to do.

Let us now return to the two-person case. What reasons do we have to believe that A, the newly-wed, is committed to 7 and 8? I noted above that Hare's argument for act-utilitarianism begins with the appeal to rational prudence. This appeal works with the thesis of the prescriptivity of 'I' to constrain A to satisfy KR. According to Hare, A satisfies the continuity requirements of personal identity by imagining that he is the person in B's precise situation, with B's set of universal features. So let us suppose

that the actual and hypothetical situations (S_1 and S_2 respectively) are identical in all relevant universal respects, with this one irrelevant difference that in S_2 A performs the role of recipient. If this be granted, and supposing that all other things are equal, A cannot evade 7. If this is correct, he will have acquired a hypothetical concern for the satisfaction of preferences he will have in S_2. This is for the satisfaction of the preference that F_x. If appeal to the requirement of prudence is in order here, A can also not evade 8. This, however, assumes that the decision-procedure for the two-person case parallels the decision-procedure for the single-person case, i.e. putting himself in B's shoes is supposed to run parallel to putting himself in his own future shoes.

If Hare is right, it is a condition of <u>full</u> representation that he must form the current preference that F_x be satisfied in ignorance of any information about the facts of his own actual case. It is part of what Hare means by becoming fully knowledgeable and rational that he has to identify the person in the reversed-role situation as himself. But it is difficult to see how the requirement that he must form the preference that F_x be satisfied in ignorance of all other information about himself, counts as a requirement of rationality in the required sense in which a preference counts as rational if it is formed in full knowledge of all relevant information.[121] Hare holds that it is possible to form a fully informed preference without appealing to a set of current preferences.[122] Yet a problem arises here concerning the reason A has to satisfy 9. If it is possible to say that he can use the word 'I' to refer to the person in the reversed-role situation as himself, this cannot be because of some distinguishing feature of his present self, or of his present motivational state, which he singles out to identify the person about whom he is concerned, and which is the reason for this concern. It seems that we must

suppose that the word 'I' attaches to the person in the reversed-role situation indifferently of these things. But, if this is so, what reason do we have to believe that, in forming the preference for F_x, he has a motivating reason presently to desire that it be satisfied strictly in proportion to its strength?

Now, the notion that a distinguishing feature of A's present self, or of his present motivational state, supplies a reason for the required concern seems indispensable. If I am right, this supplies the motivational content of a reason for action. But it is creating problems. It blocks the view that there can be some overriding reason, like the (supposed) fact of continuity itself, for satisfying preferences if there is at present no motivating reason to satisfy them, and in particular the view that rationality requires that A accept that preferences be satisfied strictly in proportion to their strength and indifferently of their content.

It seems that the sort of rule Hare requires must be one that allows the motivational content of present reasons for action to be derived from motivational states which do not at present obtain. So, the following rule would seem to be in order:

12. If A at present believes that in some state of himself he will have preferences which promote or serve a distinguishing feature of himself, or a motivational state he will then have, he has a reason presently to desire that those preferences be satisfied strictly in proportion to their strength.

12 requires that preferences be accepted indifferently as regards to what is preferred and as regards to who is doing the preferring. If 12 were in

order, A has a reason presently to desire that all preferences be satisfied strictly in proportion to their strength, albeit only initially. But 12 gives rise to a problem of a different kind. We may ask: why should A accept that all relevant preferences are unproblematically his preferences? The constraints of Hare's method require a very strong interpretation of his concept of personal identity. If Hare is right, he satisfies the continuity requirements of personal identity by imagining that he is the person in the reversed-role situation, though in that situation he will have a different set of identifying features. But, what reason do we have to believe that the person he imagines becoming and the person he presently identifies as himself are one and the same person?[123]

It is worth noting that universalizability rules out as irrelevant any information A may have regarding which motivational state is actually his own. But a reference to himself is indispensable for our understanding of what constrains him to maximize the satisfaction of his rational preferences over time. This reason provides the basis for the application of a prudential rule in the single-person case. But will it do as a reason in the two-person case? We may ask: what criteria of personal identity yield the judgement that he may have two sets of identifying features, albeit in different positions and at different times? If this is in order, he has as yet no compelling reason to accept some set of preferences as the rational set, nor a reason to maximize the satisfaction of this set.

Now, how might A decide what the 'prudent' course of action will be? Preference-strength is obviously one relevant criterion for coming to a decision. But it is not the only criterion. I have argued above, in connection with the single-person case, that a prudential rule which works on the assumption that all preferences be admitted as equally worthy of

satisfaction, would be unsatisfactory if taken to rule out as irrational reasons one may have at present to desire that one or other member of one's preference-set be frustrated. There has to be a reason for A presently to desire that one or other member of his preference-set be frustrated. 2 above suggests that preference-strength is the final criterion. If I am right, Hare cannot defend at least one requirement of rationality - 2 above - without reference to universalizability. 2 is essential for Hare's argument as a whole, for it is the premiss which commits A to 6. But 6 is problematic. We may ask: what is there in universalizability which requires that a form of rational prudence be adopted in the two-person case of preference-conflict?

I think the only reason for A presently to desire that some of his preferences be frustrated, turns on a belief about which member of his present motivational set is or will be the distinguishing feature of himself. We might then say that there is a reason for A to maximize the satisfaction of his rational preferences over time, only if these preferences serve or promote that member of his present motivational set which he at present identifies as the distinguishing feature of himself. But this is a move which universalizability blocks. If Hare is right, he has to accept that all members of his present motivational set are 'his own', and then to make a final decision on the basis of preference-strength alone. But, this is to assume that in compelling A to think of the person in the reversed-role situation as himself, he has at present reasons not to discount preferences which serve or promote motivational states he cannot at present accept, not even initially. I think this suggests that the rational requirements of decision-making on the basis of a prudential rule in the two-person case demand more than the parallel requirements of the single-person case demand.

Is appeal to a form of rational prudence the appropriate response in the two-person case? Can a single principle of rational choice govern decision-making in two different kinds of conflict situation, viz, conflicts between distinct individuals and conflicts within particular individuals?[124]

I argued above in connection with the single-person case that a distinguishing feature of one's present self, or of one's present motivational set, is indispensable for our understanding of why one may have reasons presently to desire that fully represented future preferences be satisfied or frustrated. If I am right, there is always something about one's present self, or one's present motivational set, which supplies the reasons for a preference ordering or ranking, and the motivation presently to act in ways compatible with what one will have reason to do or to avoid at other times. We may grant Hare the claim that my knowledge of my own envisaged future suffering necessarily involves having present motivations that the suffering stop. Hare claims that this link is 'conceptual'. But, if I am right, it does not follow from this that I will necessarily have no reasons at present to reject some preferences as unworthy of satisfaction. I suggested that a prudential rule which rules out such reasons as irrational would be unsatisfactory in that it would fail to take account of what is distinctive about oneself.

Now, in the two-person case the conceptual link just referred to holds between my knowing what it is like for you to be suffering, and my having at present equal aversion to suffering for a similar hypothetical situation in which I am the sufferer, which universalizability then compels me to treat as actual and 'my own'. The strategy Hare employs to compel me to grant all considered preferences the same consideration as I grant my own considered preferences, which is to reduce the inter-personal case of

conflict to an intra-personal one, does not deny the simple fact that the preferences I treat 'as if they were my own' are the preferences of a different person. I shall grant Hare the claim that to the extent I am compelled to think of the person in the reversed role situation as myself, to that extent different persons are treated equally. This is Hare's way of ensuring that we will not refuse the principle of granting equal weight to the equal preferences of all parties concerned, and to balance all considered interests on the basis of preference-strength alone.

But, even if we grant Hare the reductionist move, it does not follow that it serves as a constraint to maximize. I argued above in connection with the single-person case, that the mere fact of continuity itself does not provide a blanket reason to accept that all one's preferences are equally worthy of satisfaction, for such a reason ignores what is distinctive about one's present self, or one's present motivational set. I also said that we need to assume in addition a strong degree of self-concern as a component of one's present motivational set. This, taken together with one's present concerns, yields a reason presently to desire the maximal satisfaction, in sum, of all considered preferences, for these preferences will serve or promote one or other motivational state one at present accepts. Regarding the two-person case, I suggested that one needs to treat one or other member of one's present motivational set as if it were or will be the distinguishing feature of one's present self. But universalizability blocks this move. So the question arises what grounds we have to believe that he has a reason to maximize? It is worth noting that the preferences he is seeking maximally to satisfy must serve one or other member of his present motivational set, for otherwise they cannot at present generate reasons for action. As I judge, Hare's theory lacks the required motivational structure. It is indeed difficult to see what there is in

universalizability which requires that a form of rational prudence be adopted. If there is a (motivating) reason to maximize, this must come from elsewhere, from outside Hare's theory itself.[125]

Earlier on I asked whether a single principle of rational choice can govern decision-making in two different kinds of conflict situation. I think Hare's attempt to do so is methodologically inappropriate.[126] The objection is that a prudential rule does not work in the two-person case, and this is largely due to universalizability. The idea of an 'I' or a 'me' with a set of concerns and a motivational set, which serves as the distinguishing feature of myself, loses all force in the reductionist move, and creates the motivational gap to which I have drawn attention.[127] This suggests that the analogy between the prudent individual and the prudent universalizer is a false analogy.

There is, however, no compelling reason why we should grant Hare the reductionist move. We should recognize that the reductionist move, which demands that all relevant preferences be viewed as the preferences of a single person, is a methodological move. But the constraints of this move require a very stringent interpretation of Hare's notion of what constitutes a person.[128] I satisfy the continuity requirements of personal identity by imagining that I am the person in the reversed-role situation, though in that situation I will have a different set of identifying features. But, the person I imagine becoming and the person I presently identify as myself are different persons, for it is obvious that we do not share the same set of identifying features. In Chapters 2 and 3 I have argued that it is logically impossible to suppose that I can imagine what it would be like for myself to be you, though not that I can imagine what it is like for you, if we assume that I have access to your experiences in

just the way I have access to my own. So, if the person I represent as having certain experiences is not myself, I cannot treat my representations of what it is like for him as representations of what it would be like for myself. If this is correct, the knowledge Hare supposes is available to me, is not knowledge I can have. If the 'felt'-disposition account of my knowledge of your preferences is the appropriate account, I cannot construe the feel of what it is like for you as knowledge of what it would be like for myself, because it makes no sense to say that I can imagine myself to be you. If knowledge of your experiences is not available to me in the sense in which this is required, then the crucial claim that my knowledge of your experiences entails that I prefer what you in fact prefer, may be rejected as false.

I said that the knowledge requirement and its subsidiary claims underlie the role Hare assigns to the universalizability thesis. If I am right, the points just raised remove all necessary grounds for the universalizability thesis to come into play. It is therefore difficult to see what constrains Hare's fanatic to set aside the view that his ideal matters simply because it is <u>his</u>. It is possible, in the intial stages of the argument advanced in MT, to evade satisfying the knowledge requirement, and to refuse the subsequent steps. If I am right, it is difficult to see how the role-reversal procedure can come into play, and hence what force the thesis of the prescriptivity of 'I' can have. In view of this I am inclined to think that the new thesis advanced in MT, which Hare attempts to defend independent of universalizability, does not place the argument of MT in a better position than the arguments advanced in FR and ETU.

Notes

1. Hare (1981): Moral thinking. Oxford University Press. New York. p.108. Hare says: 'Moral judgements are, I claim, universalizable in only one sense, namely that they entail identical judgements about all cases identical in their universal properties.'

2. Hare says: 'It follows from universalizability that if I now say that I ought to do a certain thing to a certain person, I am committed to the view that the very same thing ought to be done to me, were I in exactly his situation, including having the same personal characteristics and in particular the same motivational states.' Hare (1981) p. 108. However, this thesis assumes that I know the relevant facts of the situation I am judging, which facts I cannot know unless I accept prescriptions for hypothetical cases in which I am the affected party. Acceptance of these prescriptions is a requirement of my rationality.

3. Hare says: 'The prescriptivity of moral judgements can be explained formally as the property of entailing at least one imperative.' Hare expounds this as follows: 'We say something prescriptive if and only if, for some act A, some situation S and some person P, if P were to assent (orally) to what we say, and not, in S, do A, he logically must be assenting insincerely.' Hare (1981) p. 21.

4. Hare (1981) p. 107.

5. I am indebted for this point to F. Feldman. Feldman says: '... if a person issues or is committed to an imperative, "Let X occur!", then that person prefers that x occur (rather than not).' See Feldman (1984): Hare's proof. Philosophical studies, vol. 45, p. 273.

6. This way of putting the matter is due to Hare. See Hare (1989a): The structure of ethics and morals. Essays in ethical theory. Clarendon Press. Oxford. p. 183.

7. Hare (1981) p. 94 and p. 113.

8. Hare (1981) p. 91-94.

9. This principle is variously called 'The principle of hypothetical self-endorsement' (I. Persson. 1989: Universalizability and the summing of desires. Theoria, vol. LV, p. 161.), 'The conditional reflection principle' (A. Gibbard. 1988: Hare's analysis of 'ought' and its implications. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford. p. 58), 'The preference appropriation principle' (T.M. Reed. 1988: Review of Moral thinking. Philosophia, vol. 18. p. 274).

10. W. Rabinowicz (1989): Hare on prudence. Theoria, vol. LV, p. 148.

11. Hare (1989b): Reply to W. Rabinowicz. Theoria, vol. LV, p. 154. Hare says: 'I wish it to be noted that KPEP [Knowledge of preference entails preference] is also distinct from universalizability. The method of moral argument which I have suggested is founded on a combination of KPEP, argued for in MT ch. 5, and universalizability, which is joined to it in MT ch. 6. So we must not think that KPEP by itself entails universalizability.'

12. Hare (1981) p. 95.

13. Hare (1981) p. 96.

14. I assume that the 'felt' intensity account of my knowledge of the strength of your preferences is the appropriate account. Hare is not explicit about this. J. Griffin remarks that we '... cannot be after "strength" in the sense of felt intensity, because just how strongly we feel our desires is largely a matter of upbringing ...' See Griffin (1988): Well-being and its interpersonal comparability. Hare and critics, p. 79.

15. Hare (1981) p. 98.

16. Hare (1989b) Amoralism: Reply to Peter Sandøe, Theoria, vol. LV, p. 206.

17. Hare (1981) p. 92-92. A.F. McKay supports this point. McKay says: 'If you know what it is like for another person to be experiencing something, this implies that you yourself have a duplicate experience directed, not toward the actual situation, but to the possible situation that you envision when performing the manoeuvre. That is, in order for me to know what it is like for you to be intensely preferring that the bulldog quit snuffling around your leg, I must myself equally intensely prefer that, were I you in that situation, the bulldog quit snuffling around my leg.' He says further: 'This induces in us the occurrence of ... duplicate, proxy experiences.' See McKay (1986): Extended sympathy and interpersonal utility comparisons. The journal of philosophy, vol. LXXXIII, p. 310.

18. I. Persson (1988): Rationality and maximization of satisfaction. Noûs, vol. 22, p. 545.

19. The term 'suffering thesis' is due to Reed. See Reed (1988): p. 274. Hare says: '... having an aversion to my forthwith suffering like that is a condition of full representation.' Hare (1981) p.

99. See also p. 128. A Gibbard regards KR and ST as distinct requirements. He says 'The sympathy theorem is independent of the Principle of Conditional Reflection.' See Gibbard (1988): Hare's analysis of "ought" and its implications.' Hare and Critics, p. 67. As I read Hare, satisfying the latter involves satisfying the former.

20. Hare (1981) p. 99.

21. Hare charges G.F. Schueler (Some reasoning about preferences. Ethics, vol. 95, 1984), F. Feldman (Hare's proof. Philosophical studies, vol. 45, 1984) and I. Persson (Hare on universal prescriptivism and utilitarianism. Analysis, vol. 43, 1983) with failing to see that the so-called 'veto'-option is false. Hare says: 'All three discussants use arguments which would imperil my position only if I held that, if someone has a preference that something should happen to him, he has no other preferences that could outweigh this.' See Hare (1984): Some reasoning about preferences: a response to essays by Persson, Feldman, and Schueler. Ethics, vol. 95, p. 81.

22. Hare (1984) p. 82.

23. The example is borrowed from Feldman (1984) p. 276.

24. Hare (1984) p. 82.

25. Hare (1984) p. 83.

26. The clearest statement of this point occurs in Hare (1989a) p. 186.

27. Hare (1981) p. 222.

28. Hare (1981) p. 105-106.

29. Hare (1981) p. 105.

30. Hare's discussion of the points just raised occurs in Hare (1981) p. 101-106.

31. This point is suggested by Rabinowicz. See Rabinowicz (1989) p. 145. T.L. Carson supports this interpretation. Carson observes regarding certain '... border-line cases in which an ideal observer's balance of preferences concerning the choice between x and not-x is nearly equal ...' that his '... commitment to a particular set of ideals could affect his on balance preference. Given the requirement of prudence, cases of this sort could not arise, since this requirement makes it impossible that an ideal observer or archangel could be the type of person who allows his now-for-then preferences to override his then-for-then preferences in cases in which they conflict.' See Carson (1986): Hare's defence of utilitarianism. Philosophical studies, vol. 50, p. 103. See also notes 96 and 97.

32. See Rabinowicz (1989) p. 146.

33. Hare (1981) p. 100-101.

34. Hare (1981) p. 105.

35. Hare (1981) p. 105.

36. Hare's discussion of this point occurs in Hare (1981) p. 101-106.

37. Hare (1981) p. 102.

38. Hare (1981) p. 102-103.

39. Hare (1981) p. 103.

40. Hare (1981) p. 104.

41. Hare (1981) p. 103.

42. Rabinowicz seemingly disagrees. As he reads Hare 'Satisfying a non-simultaneous preference (for example, now-for-then) at the expense of a simultaneous one (for example, at the expense of an expected then-for-then preference) is a sign of imprudence, even if the former preference is stronger.' See Rabinowicz (1989) p. 145.

43. See Rabinowicz (1989) p. 145-146.

44. Hare (1981) p. 105-106.

45. Hare (1981) p. 106.

46. Hare (1981) p. 99-100.

47. Hare (1981) p. 99.

48. Hare (1981) p. 99-101.

49. Hare (1981) p. 99-101.

50. J.C. Harsanyi contests this point in Harsanyi (1988): Problems with act-utilitarianism and with malevolent preferences. Hare and critics. p. 89-99.

51. Hare (1981) p. 180.

52. The clearest statement of this occurs in Hare (1989a) p. 186. See also Hare (1981) p. 108 and p. 221.

53. We must say 'A prefers' and not 'A would prefer'. In this I follow Rabinowicz. Rabinowicz says: 'In KPEP, my present preference is not just conditional, since I know that the situation in which I have a preference for A has actually occurred or will occur.' See Rabinowicz (1989) p. 149. See also Hare (1981) p. 104 and p. 116.

54. Hare (1981) p. 94-96 and p. 113.

55. Hare (1981) p. 94-99.

56. Hare (1981) p. 88-89.

57. Hare (1981) p. 108.

58. Hare (1981) p. 127-128.

59. See M. McDermott (1983): Hare's argument for utilitarianism. The philosophical quarterly, vol. 33, p. 386.

60. This way of putting the point is due to McDermott. See McDermott (1983) p. 386. See also R.B. Brandt (1988): Act-utilitarianism and meta-ethics. Hare and critics, p. 27-28.

61. For Hare's discussion see Hare (1981) p. 90-99 and p. 108-109.

62. See R.B. Brandt (1988) p. 36. Brandt says: Hare's theory ... essentially appeals to a person's self-interest.' Hare prefers to regard this appeal as an appeal to 'hypothetical self-interest'. See Hare (1988): Comments on Brandt. Hare and critics, p. 218.

63. In MT Hare regards the 'equal chance' option as an 'alternative' to occupying my recipients' positions 'in random order'. See Hare (1981) p. 129. However, I agree with I. Persson that the 'equal chance' option is an essential assumption of Hare's argument. See Persson (1989): Universalizability and the summing of desires. Theoria, vol. LV, p. 163-169.

64. This point is due to Rabinowicz (1989) p. 149. Hare concurs with Rabinowicz. See Hare (1989b) p. 154.

65. Hare (1981) p. 110.

66. Hare (1981) p. 109-111.

67. Here I follow I. Persson. Persson (1989) p. 166.

68. This point is due to I. Persson. See again Persson (1989) p. 167.

69. Here again I follow I. Persson (1989) p. 166.

70. See again Hare (1989b) p. 154.

71. Hare (1981) p. 92.

72. Hare (1989b) p. 154.

73. This way of putting the matter assumes Hare's concept of personal identity. See Hare's discussion in Hare (1981) p. 96-99 and p. 220-223.

74. Hare (1981) p. 98.

75. Hare (1981) p. 221.

76. Hare (1981) p. 98-99.

77. Hare's discussion of this point occurs in Hare (1981) p. 90-92.

78. Hare (1981) p. 101 and p. 104-105.

79. Hare (1981) p. 96-99. See also Hare (1989b) p. 206.

80. Hare (1988) p. 216. Hare charges Brandt with failing to understand that KR is a conceptual claim and not an empirical one. See also Brandt (1988) p. 34-35. Here Brandt attempts to argue that KR must be construed as an empirical claim which is not true at all.

81. See again Hare (1989b) p. 156.

82. Hare (1981) p. 96-99 and p. 220-223. See also I. Persson (1988): Rationality and maximization of satisfaction. Noûs, vol. 22. p. 546.

83. Here I follow Persson (1988) p. 546.

84. This point is suggested by I. Persson. Persson says that Hare conceives of '... then-for-then preferences as immutable or unavoidable; he does not seem to recognize the possibility - indeed the necessity - of appraising and criticizing them in the light of current representations, thereby perhaps preventing them from arising.' See Persson (1983): Hare on universal prescriptivism and utilitarianism. Analysis, vol. 43, p. 46. Persson pursues this point again in Persson (1988) p. 549.

85. Hare (1981) p. 222-223.

86. This example has been suggested to me by my supervisor Professor I. Macdonald.

87. See R.B. Brandt (1988) p. 35 and A. Gibbard (1988) p. 61. Gibbard's example of Cheops' desire for a big funeral makes the same point, though his case is somewhat different. As Brandt puts the point: 'For the case of desires not so directly related to experiences, it seems clear that I shall not have a desire I know I shall have later ...' See Brandt (1988) p. 35.

88. I. Persson argues that '... matters of continuity do not constitute practical reasons for a rational person ...' See Persson (1988) p. 550.

89. Hare (1981) p. 216-218.

90. I. Persson advances a similar point. See Persson (1988) p. 549.

91. See again I. Persson (1988) p. 550.

92. Hare (1981) p. 105.

93. Hare (1981) p. 105..

94. Hare is not very clear on the reason for the exclusion of auto-fanatical preferences. I am here speculating about Hare's reasons.

95. Hare (1981) p. 104.

96. See Hare (1988): Comments on Gibbard. Hare and critics, p. 231. If Hare is right, Z counts as a now-for-then preference which the requirement of prudence disallows.

97. See P. Sandøe (1989): Review of Hare and critics. Theoria, vol. LV. p. 219. Hare concurs, as is evident in Hare (1988) p. 231. Hare says: 'The other, overlapping kind of preference which could be included in K [the balance of the preferences we may have after what Gibbard calls "basic preferences" and what he calls "sympathetic preferences" has been accounted for] is the preference of the fanatic that things should universally happen, independently of how they affect his or anybody else's experiences.'

98. Hare contemplates this possibility in Hare (1988) p. 232.

99. Hare (1988) p. 233-234.

100. Hare (1988) p. 233.

101. Hare (1988): Comments on Griffin. Hare and critics, p. 235-236.

102. Hare (1981) p. 222.

103. Hare (1981) p. 108.

104. Hare (1981) p. 108.

105. Hare (1981) p. 223.

106. Hare (1981) p. 222.

107. P. Singer argues for the latter. He says: 'I have been arguing that as long as we reject the idea that there can be objectively true moral ideals, universalizability does require that we put ourselves in the place of others and that this must then involve giving weight to their ideals in proportion to the strength with which they hold them.' See Singer (1988): Reasoning towards utilitarianism. Hare and critics, p. 152.

108. Hare (1981) p. 222-223.

109. Hare (1988) p. 234.

110. James W. McGray employs the notion of the 'extended person' in his discussion of Hare's rationality constraints. See McGray (1986/87): From universal prescriptivism to utilitarianism: a logical gap. Philosophy research archives, vol. XII, p. 79 and McGray (1990): Universal prescriptivism and practical scepticism. Philosophical papers, vol. XIX, p. 41.

111. Hare (1981) p. 108.

112. Hare (1981) p. 223.

113. Hare (1981) p. 115-116.

114. Hare (1989b) p. 154.

115. Hare (1989b): Universalizability and the summing of desires: reply to Ingmar Persson. Theoria, vol. LV, p. 176.

116. See again Singer (1988) p. 152. See also Hare (1981) p. 144. Hare says: 'We are committed by the formality of our method to a Benthamite answer to the basic question: equal preferences count equally, whatever their content. This is because the only question the method allows us to ask is, What shall we rationally universally prescribe, or from an impartial standpoint prefer, if we are fully informed and make no logical mistakes?'

117. Hare (1989b) p. 156.

118. Hare (1981) p. 220-223.

119. Hare (1981) p. 101-106.

120. This line of reasoning is suggested by I. Persson. See Persson (1988) p. 549.

121. This point derives from I. Persson (1989) p. 170.

122. Hare (1988): Comments on Griffin. Hare and critics p. 236. See also Hare (1981) p. 129.

123. I am here appealing to objections raised by C.C.W. Taylor to the argument presented in FR. See again C.C.W. Taylor (1965): Review of Freedom and reason. Mind, vol. LXXIV. Hare appeals to Z. Vendler for support. See Hare (1981) p. 119-121. As Hare interprets Vendler, it is legitimate to say that '... putting myself in somebody else's shoes does not involve supposing myself to have simultaneously two incompatible sets of properties; it involves merely supposing that I might lose one set and acquire the other.' See also Hare (1988): Comments on Vendler. Hare and critics p. 280-282. But Vendler claims that the move Hare needs to carry his argument is logically impossible. Vendler says: '... I can imagine being you, but not me being you, or you being me. For these "you's" and "me's" denote distinct individuals, which cannot be mixed or exchanged.' See Vendler (1988): Changing places? Hare and critics p. 181. Vendler, in effect, argues that the role-reversal procedure, as Hare conceives of it, which involves making judgements in propria persona in the reversed-role situation, is impossible and that Hare can dispense with this procedure. See Vendler (1988) p. 182-183. I agree with Vendler. S.L. Reynolds advances a similar point. Reynolds says: '... even if I would naturally claim to be imagining myself [to be Napoleon], surely I'm not really imagining myself if there's nothing about me in what is imagined.' See Reynolds (1989): Imagining oneself to be another. Nous, vol. 23, p. 624. G. Brown advances the same point of criticism. Brown says: '... if this "I" has all the preferences of N, and is in the precise circumstances of N ..., then this "I" is none other than N, and all I am doing is imagining N in N's situation.' See Brown (1985): On pleasing all of the people all the time. Theoria, vol. 51, p. 22.

124. This way of putting the matter is due to T. Nagel. See Nagel (1988): The foundations of impartiality. Hare and critics p. 110-111. Nagel says: 'I myself think there is something in the

view that compensation across lives cannot take the same form as compensation within lives, and the same principles should not govern the two kinds of conflict.'

125. T. Nagel comes to much the same conclusion. See Nagel (1988): p. 111-112.

126. I owe this point to L.A. Mulholland. See Mulholland (1986): Rights, utilitarianism, and the conflation of persons. The journal of philosophy, vol. LXXXIII p. 329.

127. N.O. Dahl argues that Hare's appeal to rational prudence is not 'morally neutral'. Dahl says: 'Unless Hare can provide additional grounds for his requirement of prudence, grounds that are themselves morally neutral, he is not entitled to appeal to it in his account of critical thinking.' The problem here, as Dahl sees it, is that '... Hare's principle of rational personal choice, his requirement of prudence, says about personal choice what utilitarianism says about choices involving others.' See Dahl (1987): A prognosis for universal prescriptivism. Philosophical studies, vol. 51, p. 411 and p. 413.

128. Hare remarks that his way of handling the problem of personal identity has this consequence that '... it makes personal identity no longer an all-or-nothing affair ...' Hare (1981) p. 97. But this is surely wrong. I am required to identify the person whose role I perform as myself i.e. I am required to accept that I may have different sets of identifying features.

LIST OF WORKS CONSULTED

1. Braithewaite R.B. (1954): Critical notice of The language of morals. Mind vol. 63 p. 249-262.
2. Brandt R.B. (1988): Act-utilitarianism and metaethics. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.
3. Brown G. (1985): On pleasing all of the people all the time. Theoria vol. 51 p. 16-23.
4. Carson T.L. (1986): Hare's defence of utilitarianism. Philosophical studies vol. 50 p. 97-115.
5. Caton C.E. (1963): In what sense and why 'ought'-judgements are universalizable. The philosophical quarterly vol. 13 p. 48-55.
6. Crowell E.F. (1974): Hare on desire and moral reasoning. Theoria vol. XL p. 152-162.
7. Dahl N.O. (1987): A prognosis for universal prescriptivism. Philosophical studies vol. 51 p. 383-424.
8. Feldman F. (1984): Hare's proof. Philosophical studies vol. 45 p. 269-283.
9. Gardiner P.L. (1955): On assenting to a moral principle. Proceedings of the Aristotelian society vol. 55 p. 23-44.
10. Gibbard A. (1988): Hare's analysis of 'ought' and its implications. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.
11. Gorr M. (1985): Reason, impartiality, utilitarianism. Morality and universality edited by N.T. Potter and M. Timmons. Reidel. Dordrecht.
12. Griffin J. (1988): Well-being and its interpersonal comparability. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.
13. Hare R.M. (1982): The language of morals. Oxford University Press. Oxford.

14. Hare R.M. (1973): Universalizability. Essays on the moral concepts. University of California Press. California.
15. Hare R.M. (1980): Freedom and reason. Oxford University Press. Oxford.
16. Hare R.M. (1989): Ethical theory and utilitarianism. Essays in ethical theory. Clarendon Press. Oxford.
17. Hare R.M. (1981): Moral thinking. Oxford University Press. New York.
18. Hare R.M. (1989): Relevance. Essays in ethical theory. Clarendon Press. Oxford.
19. Hare R.M. (1989): Supervenience. Essays in ethical theory. Clarendon Press. Oxford.
20. Hare R.M. (1989): The structure of ethics and morals. Essays in ethical theory. Clarendon Press. Oxford.
21. Hare R.M. (1984): Some reasoning about preferences: a response to essays by Persson, Feldman and Schueler. Ethics vol. 95 p. 81-85.
22. Hare R.M. (1988): Comments on Vendler. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.
23. Hare R.M. (1988): Comments on Brandt. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.
24. Hare R.M. (1988): Comments on Gibbard. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.
25. Hare R.M. (1988): Comments on Griffin. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.
26. Hare R.M. (1989): Reply to W. Rabinowicz. Theoria vol. LV p. 152-158.
27. Hare R.M. (1989): Amoralism: Reply to Peter Sandøe. Theoria vol. LV p. 205-210.
28. Hare R.M. (1989): Universalizability and the summing of desires: reply to Ingmar Persson. Theoria vol. LV p. 171-177.
29. Harsanyi J.C. (1988): Problems with act-utilitarianism and with malevolent preferences. Hare and critics edited by D. Seanor and N.

Fotion. Clarendon Press. Oxford.

30. Kim J. (1984): Concepts of supervenience. Philosophy and phenomenological research vol. XLV p. 153-176.
31. Lycan W.G. (1969): Hare, Singer and Gewirth on universalizability. The philosophical quarterly vol. 19 p. 135-144.
32. Locke D. (1968): The trivializability of universalizability. The philosophical review vol. LXXVII p. 25-44.
33. Locke D. (1981): The principle of equal interests. The philosophical review vol. XC p. 531-559.
34. McCloskey H.J. (1978/79): Universalized prescriptivism and utilitarianism: Hare's attempted forced marriage. The journal of value inquiry vol. 12-13 p. 63-76.
35. McDermott M. (1983): Hare's argument for utilitarianism. The philosophical quarterly vol. 33 p. 386-391.
36. McGray J.W. (1986/87): From universal prescriptivism to utilitarianism: a logical gap. Philosophy research archives vol. XII p. 79-86.
37. McGray J.W. (1990): Universal prescriptivism and practical scepticism. Philosophical papers vol. XIX p. 37-51.
38. McKay A.F. (1986): Extended sympathy and interpersonal utility comparisons. The journal of philosophy vol. LXXXIII p. 305-322.
39. Milkman K.A. (1982): Hare, universalizability, and the problem of relevant descriptions. Canadian journal of philosophy vol. XII p. 19-32.
40. Mulholland L.A. (1986): Rights, utilitarianism, and the conflation of persons. The journal of philosophy vol. LXXXIII p. 323-340.
41. Nagel T. (1988): The foundations of impartiality. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.
42. Norton D.L. (1980): On an internal disparity in universalizability criterion formulations. Review of metaphysics vol. XXIII p. 519-526.

43. Oldenquist A. (1968): Universalizability and the advantages of nondescriptivism. The journal of philosophy vol. LXV p. 57-79.

44. Persson I. (1983): Hare on universal prescriptivism and utilitarianism. Analysis vol. 43 p. 43-49.

45. Person I. (1988): Rationality and maximization of satisfaction. Noûs vol. 22 p. 537-554.

46. Persson I. (1989): Universalizability and the summing of desires. Theoria vol. LV p. 159-170.

47. Rabinowicz W. (1989): Hare on prudence. Theoria vol. LV p. 145-151.

48. Reed T.M. (1988): Review of Moral thinking. Philosophia vol. 18 p. 271-283.

49. Reynolds S.L. (1989): Imagining oneself to be another. Noûs vol. 23 p. 615-632.

50. Roxbee Cox J.W. (1986): From universal prescriptivism to utilitarianism. The philosophical quarterly vol. 36 p. 1-15.

51. Sandøe P. (1989): Review of Hare and critics. Theoria vol. LV p. 211-222.

52. Schueler G.F. (1984): Some reasoning about preferences. Ethics vol. 95 p. 78-80.

53. Singer P. (1988): Reasoning towards utilitarianism. Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.

54. Taylor C.C.W. (1965): Review of Freedom and reason. Mind vol. 74 p. 280-298.

55. Vendler Z. (1988): Changing places? Hare and critics edited by D. Seanor and N. Fotion. Clarendon Press. Oxford.

56. Versényi L. (1972): Prescription and universalizability. The journal of value inquiry vol. 6 p. 22-36.

57. White H.J. (1969): An analysis of Hare's application of the thesis of universalizability in his moral arguments. Australasian journal of philosophy vol. 47 p. 174-183.