HARE AND "MORAL THINKING"

by

TIMOTHY BENJAMIN LEBON

Degree: M.Phil. (Philosophy)

College: Bedford College,
London

CONTENTS

Abstract of Thesis

This thesis is a critical study of R.M. Hare's recent work "Moral Thinking" and an attempt to develop some of the ideas in it. Hare's book contains illuminating discussions on many ethical and metaethical issues, but two central themes can be isolated. One is his suggestion that there are two levels of moral thinking, the critical and the intuitive; the other is his detailed argument that preference-utilitarianism operates at the most fundamental level of moral thinking, the critical level, to ultimately determine what we ought to do.

Hare's argument for utilitarianism will be studied in depth. In particular his appeal to linguistic intuitions, his assertion that 'ought' is prescriptive and the argument of chapters 5 - 6 of "Moral Thinking" will all be assessed, as will his claim that prudence rules out amoralism.

Hare admits that his book leaves some important issues less than fully discussed: I attempt to see what implications these have for Hare's theory. Most of these relate to the concept of 'preference'; in particular to the question of whether 'irrational' preference should be excluded.

I also try to see what connection there is between Hare's preference-utilitarianism and happiness based theories. These issues are of independent interest as they affect the feasibility of preference-utilitarianism in general, not just Hare's version of it.

Finally, Hare's suggestion that there are two levels of moral thinking will be assessed, as well as his proposal that this allows both for a defence of utilitarianism and a reconciliation between it and deontological theories.

Chapter I Introduction

"Moral Thinking" is perhaps first and foremost a continuation of the suggestions made in "Freedom and Reason" concerning the constraints universalizability and prescriptiveness place on the moral thinker. Hare's mature view is that these constraints are so powerful that if we are rational we will exercise our freedom to reason in only one way, that being in line with preference-utilitarianism. Hare's work is consequently to be seen as one of the most detailed and eloquent proofs of utilitarianism. But alongside this argument for utilitarianism there runs a significant meta-ethical contribution. Hare suggests that there are two levels of moral thinking, which are complementary to each other. Hare calls the most fundamental level the "critical" level of moral thinking. Critical moral thinking requires that the thinker ignores his normative views and makes moral judgments in line with the logical properties of moral words and the facts of the situation. It is at the critical level that preference-utilitarianism operates, and the major part of Hare's book is devoted to proving this. But critical thinking presupposes degrees of knowledge, free time and impartiality which human beings seldom possess. Paradoxically by its own reasoning, too frequent appeals to critical thinking would be sub-optimal. It is far better (from the point of view of critical thinking) that we are disposed to adhere to intuitive principles of the form: "Do not lie": critical thinking is appropriate only when framing these intuitive principles or when they conflict in a certain situation. Hence much of our everyday moral discourse is at the intuitive level. Hare argues that the distinction between levels allows for a reconciliation between deontological themes and utilitarianism. According to Hare, they are not conflicting theories, since they operate at different levels of moral thinking. Utilitarians are right in that utility is the ultimate determinant of what is right, yet deontologists are quite correct in urging us not to make too facile appeals to utility. "Moral Thinking" is consequently a most ambitious work: not only does it aim to show that the rational moral thinker must be utilitarian, it also traces the source of much argument among moral philosophers.

I propose to concentrate on three main themes:

(1) Critical thinking and its derivation, providing Hare's answer to "What ought I to do?"

(2) Complications which may arise concerning the concept of preference and preference-utilitarianism

(3) The two levels of moral thinking and the consequent reconcilitation between utilitarianism and deontological theories.

Summary of the main issues discussed in each section

It will be helpful to have set out in front of us from the outset the argument Hare uses to show that the rational moral thinker will be utilitarian. Since Hare thinks that it is prudent to do what you ought to do, critical thinking also provides the rational answer to the question "What shall I do?"

Imagine that I am contemplating what to do in a given situation. The following is my interpretation of Hare's argument which is meant to lead me to the conclusion that I should act in accordance with preference-utilitarianism (which may involve a direct calculation or, more probably, the adherence to an intuitive principle derived from preference-utilitarianism).

H1: If I am prudent I shall do what I ought to do; i.e. amoralism is ruled out (but on prudential rather than logical grounds). (Moral Thinking [MT] X1).

H2: I must frame a moral principle for the situation. There are, however, two levels of moral thinking, and I must decide what level I am operating on. If the situation is new to me, or if I find that my intuitive principles conflict, then I must do some critical thinking. In any case, critical thinking is essential at some stage in each individual's life in order to criticise the intuitive principles he finds himself with (MT 2, 3).

H3: In order to answer the question: "What ought I to do?" rationally I must understand the words contained in the question. In particular, I must be aware of the logical properties of the word 'ought'. (MT 1).

H4: An investigation into the logical properties of the moral 'ought' reveals that it is prescriptive, universalizable and, at the critical level, overriding. These features, in conjuction with others (H5 - H10) will constrain moral thinkers in that there will be very many 'ought' statements which in any given situation they cannot rationally adhere to. (Language of Morals, Freedom and Reason, Mt 1).

H5 Rationality requires of us that we know all the facts relating to the situation. These include facts about other people's preferences (in practice we will not be able to know all the facts, but we must attempt to discover all those which are relevant, i.e. affect the outcome of critical moral thinking). (MT V).

H6 If I know that a person whom I may affect prefers with intensity L that action A be done (or not be done) then I must know that were I, hypothetically to be put in his exact position (including having his preferences), I too would prefer that A be done with intensity L. (MT V).

H7: I cannot know that if I were in that position I would prefer that A be done with intensity L (H6) unless I now prefer with intensity L that if, hypothetically, I were in his position A shall be done; i.e. H5 and H6 mean that I must now have a preference relating to what would happen to me were hypothetically I to be put in the other person's shoes (MT V).

H8: By universalization I must make my hypothetical concern for other people actual, i.e. I must convert my hypothetical preference that "were I in his shoes, A should happen" into an actual preference that A be done (even though as a matter of fact I am not in his position). (MT VI).

H9: From H7 and H8 it follows that I must have a preference that A be done (or not be done) corresponding to the preference of all those affected by A (including myself). (MT VI).

H10: By prescriptivity and overridingness (H4) I can say "A ought to be done" sincerely only if I prefer on balance that A be done. From H9 it is clear that the only action which I prefer to be done on balance will be that prescribed by preference-utilitarianism. (MT VI). Therefore my answer to "What shall I do?" will be governed by utilitarianism.

H11: But an appeal to preference-utilitarianism reveals that, paradoxically, performing a cost-benefit analysis each time we were contemplating action would be self-defeating in that it would not lead to maximum preference satisfaction. Consequently we must work out for outselves a set of simple principles to guide our lives, which in general will maximise preference satisfaction. These principles are our intuitive moral principles: only when they conflict or when we need a new situation should we actually undertake critical thinking (MT III).

Section I comprises an appraisal of the above argument H1 - H11.

Chapter II questions H1: is it really prudent to be moral?

Chapter III questions H3: is Hare's method valid?

Chapter IV looks into H4 and in particular prescriptivity and moral weakness.

Chapter V examines the rest of the argument in H5 to H10.

Section II: Complications arising from the application of critical thinking.

It is by no means obvious what preference-utilitarianism is. Hare alludes to but does not enter into a detailed discussion of some of the following complications:

(1) What is a preference? - the ambiguity between its being a purely behavioural disposition to act and a rational evaluation of what is of value.

(2) The relationship between preference - and happiness - versions of utilitarianism.

(3) "Irrational" preferences - i.e. preferences which the rational agent would not have or would override: should these be excluded?

Section III: The two levels of moral thinking.

Chapters VII and VIII of the present work examine the claim that there are two levels of moral thinking and the suggestion that utilitarianism and deontological theories are compatible.

Chapter II

Prudence and Morality

Like many moral philosophers, Hare wants his writings actually to influence the way people live. In steps H2 - H11 (in my presentation of Hare's argument, given in Chapter I) Hare demonstrates that if people wish to say things like "You ought to do X", and wish to use language correctly and be rational, then they must do so in accordance with what Hare calls critical thinking. But before we embark upon an examination of this argument, it is worth considering what reasons Hare gives for anyone to commit himself to making moral judgments. For Hare's position (quite clearly stated in MT 10) is that the logic of the moral concepts only get a hold if an agent chooses to use them: thus we are quite free either to refrain from using the moral words or to use them only to make judgments of moral indifference (e.g. "it is not the case that you ought to do X"). If Hare left it at that his enterprise would not be totally unsuccessful: for assuming that the argument H2 - H11 could be upheld, then he would at leave have given a correct account of how to use 'ought' (and consequently of how to act since 'ought' is prescriptive) which would be binding on anyone who wished to be moral. But the question arises: what can be said to the rational egoist, the man who wishes to maximise his own preference-satisfaction? To avoid circularity, whatever reason we suggest to him for his taking part in moral discourse must be of a non-moral kind: for the choice facing the rational egoist is either to opt out of morality altogether or to make judgments in accordance with critical thinking. In MT 11 Hare contends that on grounds of prudence the egoist should reject amoralism.

On the face of it this is a rather surprising claim. Of course it is generally agreed that morality pays humanity as a whole, i.e. it is in everyone's interest that everyone is moral rather than not. But because morality entails each individual making some sacrifices - i.e. on occasions undertaking the action which is not in his self interest - it is even more preferable, from each individual's point of view, for all other people to be moral but for him to act in his own self-interest; that way he gets the best of both worlds. In order of rpeference the three possibilities are:

(1) everyone else moral and my always acting selfishly
(2) everyone - including me - being moral
(3) no-one being moral

Consequently so long as my not being moral does not lead to other people following likewise [i.e. because of interdependence of choices (1) leading to (3)] it is plausible to argue that morality does not pay.

This conclusion would also seem to follow from an application of Hare's own theory. Morality - i.e. critical thinking - means weighing the individual's own preference with everyone else's and then acting upon that judgment which maximises preference-satisfaction overall. Prudence can be characterised as involving many different things, but a plausible definition which would certainly be favoured by the preference-utilitarian would be that it requires maximising the agent's own preference-satisfaction. The suggestion that the morally correct act and the prudential act always coincide is quite ludicrous. Obviously there are many occasions when my preferences do not outwigh the preferences of other people affected by my action, and in those situations to be moral I must make some sacrifice. Hare could accept this, and his argument for prudence and morality coinciding does not rest on the wildly implausible suggestion that the right act will always be the one which is in our self-interest. But before we go on to consider Hare's argument given in MT 11, it is interesting to look briefly at a few ways in which the results of critical moral thinking and critical prudential thinking (i.e. maximising one's own preference satisfaction) may be closer than might be thought.

Most importantly, society provides devices which artifically induce one to take other people's preferences into account. Thus intrinsically I may have no desire whatsoever that Jones's property should not be stolen. Consequently were there no legal system, there could be circumstances where I would act on my preference to have his property - even though his preference to keep his property might be stronger. This is a very simple illustration of the way in which prudence and morality may conflict. But now suppose that a law is introduced whereby I am severely punished should I steal Jones's property. It may well be that my preference to avoid punishment is stronger than mine for his property, and so I do not steal it. In this case the law has led to my doing the morally correct action: but not because I have taken other people's preferences into account, but rather because institutions exist which ensure that my own (selfish) preferences are adversely affected should I undertake actions which lead to other people's dissatisfaction. Note that it is not only the law that has this effect: praise, blame and social acceptance or disgrace widen the network of this process. But still it must be rather limited. The law may punish my taking Jones's property but it does not touch me if I hurt his feelings in some other way. Selfish behaviour is an infinitely wider class of behaviour than criminal behaviour: neither is it the case that all selfishness is socially disapproved of (if discovered). Neither would it be true to say that the laws of this land or established social mores always provide

prudential reasons for following Hare's particular moral system. For instance, it may well be that critical thinking requires of us that we do not keep those who are suffering alive if it is against their wishes (Hare himself cites this case in MT 10). But in this example not only is it not true that society provides reasons for acting morally, on the contrary it actually imposes additional reasons for acting immorally (if Hare's morality is the correct one). So to the extent that critical thinking is at odds with conventional morality the effect described above actually works in reverse. We are undoubtedly very far from the position we would be in were we to have certain knowledge that we were to be punished in the afterlife in exact proportion to the moral crimes we commit in this life.

A second way in which prudence and morality may converge arises if one considers different types of preferences (or desires). Many of our preferences - especially those for pleasurable experiences - are essentially self-regarding in that they centre upon our own experiences. Others. such as those for the well-being of other members of one's family, are other-regarding desires. It is notoriously difficult to characterise the exact difference between these two categories of desires (or preferences). Intuitively it is, I take it, uncontroversial to say that some desires are self-regarding whereas others are not: indeed we use such terms as "self-centred" to describe people who only care about what happens to themselves. But the obvious definition - i.e. that self-regarding desires are ones which if satisfied increase our own well-being - fails because in a sense we get something out of all our preferences being satisfied; e.g. although my preference that a friend of mine should get the job he wants is other-regarding, on the proposed definition it would count as self-regarding, since if it is satisfied one of my preferences is satisfied and I will probably feel good about his getting the job (i.e. it will increase my level of preference satisfaction). Brandt (A Theory of the Good and the Right) suggests that self-regarding preferences are those which cannot be fulfilled unless we exist. On this basis my desire for my friend to get a job clearly counts as other-regarding, which it should. Whilst this definition may have some counter-examples it will serve us in our discussion.

The relevance of this distinction may not be immediately obvious. My point is simply that the existence of other-regarding desires might lessen the divergence between morality and prudence. Imagine two worlds: one in which people have only self-regarding desires (WS) and another where they have both types (WB). Consider the example previously alluded to of my contemplating stealing Jones's property.

In WS I would never have an intrinsic perference for Jones to keep his property (since this is an other-regarding desire): if I refrained from theft it would be because my preference to steal is outweighed by the adverse side-effects. In WB though there is the possibility of my refraining from theft even if there are no adverse side effects and I have a preference to steal his property, I can have the other-regarding preference that Jones should keep his property. I could have come to have this preference in a number of ways: Jones could be a friend of mine, in which case I could have the preference that he specifically should retain what is his. Alternatively, I could have what Hare would call the universal desire that everyone in this sort of position should keep their property. Quite clearly then there are a variety of different types of other-regarding desires, ranging from moral desires to desires for the well-being of our family and friends. We can even imagine a third world (Wo) where everyone has overriding universal preferences, i.e. in every situation what everyone cares about most is what they feel about what shall be done in the situation in general, regardless of the role they have. Were our world like this then there would be a complete convergence of morality and prudence: for critical thinking requires the maximal satisfaction of preferences and in Wo each individual's preferences already mirror the outcome of critical thinking.

Three considerations limit the practical impact of other-regarding desires on the question of whether it is prudent to be moral. Most obviously, the world is not like Wl: indeed it resembles WS more than Wo. Many of our strongest desires are for our own experiences, and these may well conflict with other people's preferences. Secondly, there is the not remote possibility that our other-regarding desires do not exactly match other people's self-regarding desires. In my example my desire that Jones should not have his property stolen precisely mirrored his own desire. But in the world as it is there is no reason why it should be exactly the same strength: more seriously there is the distinct possibility that it may even go against his own desire. This is quite common in the parent-child situation. Here the parent cares about his child but it certainly is not always the case that he wishes him to do exactly what the child himself wants. There is also the possibility of malevolent other-regarding desires. So although the existence of other regarding desires introduces the possibility of prudence and morality converging, in reality it may have the opposite effect. More seriously still, it might be suggested that it is prudent to maximise the satisfaction only of self-regarding desires, i.e. preferences for others' well-being should be excluded from the calculations of what is in our own interests. I think our intuitions run two ways here. Certainly with moral desires

e.g. a man's desire that a certain war should end - it would be odd to suggest that it was prudent for him to take measures which would increase the probability of its ending. (Though in a sense he would benefit since a preference of his would be satisfied). Yet with some other-regarding desires - e.g. a desire that one's family should flourish - it would be equally odd to say that one's happiness would not increase by the desire being satisfied. Clearly there is much room for debate over where the line should be drawn exactly, but it would at any rate seem fairly safe to conclude that the satisfaction of some, but by no means all, other-regarding desires would normally be thought of as contributing to a person's happiness (or self-interest).

A third route by which morality and prudence can coincide is when my taking what appears to be the self-interested course of action leads to others doing likewise, the result being that I would in fact have been better off being moral. Recall that my brief argument for the divergence between morality and prudence at the start of this chapter rested on the contrary assumption: i.e. my taking option (1); me self-interested, others moral, did not lead to option (3): everyone being self-interested. Whilst in general this assumption may be valid - for instance whether or not I pay my taxes to the full does not affect whether the other fifty million citizens of this country do - in certain instances a degree of interdependence can exist. The natural example that comes to mind is of truth-telling: in a particular situation it may appear in my interest to lie, but it may be that the person to whom I lie may discover my deceit and do likewise to me in a similar situation (note that this example differs from the more complex circumstance of there being no chance of discovery, when it is still arguable that in some way it is prudent to tell the truth, because of the subtle effects on the intitution of truth-telling). Another exmaple might arise in a mutually profitable relationship between a journalist and a politician. It may appear to be in the journalist's interest to publish an exclusive account of the scandal the politician has told him about "off the record". However, this course of action would breach the trust between the two and would result in the politician withholding such information in the future. Consequently it is in the journalist's enlightened self-interest to keep his bargain with the politician. This sort of situation is probably quite common: a breach of an understanding is in both parties' interests if the other does not find out, but as he will, both parties sacrifice their immediate interests to get the second-best option of keeping the agreement rather than the worst (least preferred) option where both breach the agreement. But there must also be plenty of cases where detection is not likely and equally as many

where the agreement has precious little to do with morality, i.e. whilst it certainly can be the case that option (2) is the morally preferred option, there is no necessity that it is (the agreement could be between thieves or it could be in the public interest that the scandal the politician wants hushed up is instead exposed).

Despite the qualifications required, though, these three factors do at least show that enlightened self-interest may be closer to morality than might appear to be the case. To recap on the positions, it may help to consider a particular example. Imagine I am considering breaking a promise. The common-sense appraisal of the situation might be that this would be prudent but not moral: for we are to imagine that there are considerable advantages which accrue to me should I break the promise. But consider the five ways in which even here morality and prudence could coincide.

(1) Other people's preferences could be in line with mine, i.e. no-one minds if I break the promise. Since critical thinking says that the morally required action is that which leads to maximum preference-satisfaction, obviously both morality and prudence require that I break the promise.

(2) Others might object to my breaking the promise, but my preference could be stronger than theirs. Again, critical thinking would recommend doing the self-interested action.

(1) and (2) are ways in which the moral act may turn out to be the self-interested one too. (3) - (5) are the three routes by which the reverse happens, i.e. the morally correct act is also in my enlightened self-interest.

(3) Others object to my breaking the promise, and because of this my other preferences will be adversely affected if I break the promise, e.g. the law may be used against me or I may be shunned socially. It may well be that the net effect of breaking the promise is that my total preference-satisfaction would be reduced compared to keepting it.

(4) Although I want to break the promise, I also care about the welfare of the person it will adversely affect, or possibly about the upholding of promise-keeping in general. It is quite possible that my other-regarding preferences outweigh my self-regarding ones to break the promise.

(5) My breaking the promise would affect the way the person affected treated me: it might turn a mutually profitable relationship. It is consequently in my enlightened self-interest to keep the promise.

It would be ludicrous to argue that the moral act and the prudent act are never the same. But it would be equally ludicrous to suggest that they are always the same; (1) - (5) do not apply to all situations by any means. But if this is the case, why cannot the rational egoist pick and choose when to be moral, i.e. to breach morality whenever it was in his interests to do so?

Hare's answer rests on an assumption concerning human capabilities. He admits that an omniscient, omnipotent arch-demon would do exactly this. But he suggests that we humans cannot do a cost-benefit analysis on each situation, and if we did we would very likely get it wrong. We must each of us act on whatever dispositions we have: if we are moral we will tend to do the moral act and if not then we will do the self-interested act. So we have the choice of whether to be either generally moral or generally self-interested: Hare argues that, paradoxically, the former is the prudent course of action.

Hare sets out his argument for this conclusion on pages 194 - 198 of MT. His first argument rests on the suggestion that "crime does not pay" and so resembles my argument (3). As has been argued, this is of limited use to Hare, since there are plenty of immoral actions which are neither illegal nor subject to the disapproval of society. Hare's main argument is put by means of considering what principles (i.e. dispostions of character) we would make a child have if we were solely interested in his well-being. According to Hare we would inculcate in him the prima facie moral principles recommended by critical thinking. Unfortunately he does not give very many reasons for this conclusion, apart from his suggestion that crime does not pay and that "those who do not love their fellow men are less successful in living happily amongst them." Also he suggests that it is impossible to be selective about when to be moral if one cannot have a disposition to follow the principles only to the extent that they are required for the approval of society.

Of course it is an empirical question how good Hare's arguments are. But it seems to me that they are not very plausible. To be sure, if we were bringing up a child with only his interests in mind we would not mould him into being a criminal; but the gap between criminality and immorality is a large one. It is my guess that we would, as Hare suggests, try to make him obey the law and not do things society disapproves of in general. We might also teach him to be loyal to friends and family as otherwise, as Hare would argue, he would miss out on any decent relationships. But this is all a very long way from him having the moral principles suggested by critical thinking. It seems to me that human nature is such that we are very prone to being very "moral" to some members of our country, or race, or community, whilst the reverse

to outsiders. Our model child would probably be best served doing likewise; he would be kind to his family and friends, but if we really wanted to maximise his well-being we would not instil in him the disposition to risk his life to help a stranger.

It has to be admitted that my objections to Hare cannot be regarded as conclusive since they rest on unverified opinions on empirical matters. But on the face of it, it would be rather surprising if the principles recommended by critical thinking were identical with those of critical prudential thinking. This is particularly the case when one remembers that utilitarianism is in two separate ways quite a demanding moral code. The first reason has already been alluded to: namely that obeying it will bring social disapproval since our conventional morality is certainly not completely utilitarian. The other reason is that utilitarianism, more I think than any other moral codes, requires an unusual degree of self-sacrifice. Utilitarianism demands at once that one must sacrifice any amount of one's onw happiness - perhaps even one's life - if by so doing one adds to the total amount of happiness in the world. Other ethical theories differ in that many would distinguish between its being morally praiseworthy and obligatory to make such a sacrifice, and would also suggest that not all people have the right to make claims on you.

Hare deals with similar points towards the end of MT 11 when he argues that acts of supererogation would not be required by critical thinking. But his argument here is not particularly persuasive. He suggests that only those with a "vocation" for doing saintly deeds should undertake them. The idea is that if we set the standards of morality too high, too many people would opt out of morality. This is perhaps plausible if one is going to hand over a set of intuitive principles to someone on an all-or-nothing basis, i.e. the receiver of the principles is what Hare terms a "prole" who can do no critical thinking and who either practise all of morality or give it up altogether. But in the case of someone who can do his own critical thinking such a danger is not present. If, for example, giving 50% of my money to charity is the morally correct act, how is it allowable that I should not do it? It seems but a small step from someone saying that he has no "vocation" for such deeds, even if he should do it, to saying that he has no "vocation" for basic moral acts like truth-telling.

This, however, is not the place for detailed argument concerning the amount of self-sacrifice required by utilitarianism: it is obvious that it requires more than a Nozick-type ethical system ascribing negative rights to others. The question is, then, why not inculcate your child with a less demanding moral code that utilitarianism, preferably one which is in line with conventional morality? I do not see that Hare had

any arguments which will show than on prudential grounds utilitarianism is preferable.

So far I have taken it for granted that Hare is arguing for it actually being in someone's interests to be a utilitarian rather than immoral, amoral, or follow an alternative moral code. Such is certainly the impression one obtains from MI 11. But in the context of the main argument, H1 - H11, Hare need not make such a bold claim. For all he has to do is to show it is better than amoralism: the other two alternatives are ruled out on logical grounds. That is, granted that the rest of the argument - H2 - H11 - is correct, one cannot adopt another moral code yet still use words like 'ought', since the logical properties of 'ought' constrain one to make utilitarian prescriptions. Nor can one be self-confessedly immoral yet maintain that one is using the moral words properly. It is interesting to see here a possible conflict in the dictates of rationality. Hare maintains (MT p.2) that if we are to be rational we must understand the words we use, when we ask "What ought I to do?" and will (MT p.7) "if we are rational, exercise our freedom (to reason) in only one way". In other words once we pose the question "What ought I to do?" it is rationality which requires us to do the utilitarian act. Yet at the same time it is generally accepted that it is rational to act prudently. But in view of the preceding argument it may well be prudent to be immoral. A good example is that it may well be prudent not to pay in full one's income tax, i.e. to understate one's earnings. Particularly amongst the self-employed this is general practice and it is not subject to social condemnation: moreover, the risk of detection is slight and the benefit quite large.

The tax dodger will not - if H2 - H11 is correct - be rational, in the sense that he will not be doing what he morally ought to do; if he claims he is doing what is morally right then Hare can throw rationality via following the logic of the moral concepts at him. Yet what he does provides a correct answer to the different question: "What prudentially ought I rationally to do?" indeed it is arguable that the moral act is irrational from a prudential point of view. Rationality provides different answers because the questions being posed are different: the answer to "What rationally ought I prudentially to do?" Hare cannot say, then that rationality is on his side in his debate with the rational egoist who is contemplating whether or not to be moral.

In view of considerations such as those, Hare never claims that rationality or logic alone compel anyone to be moral. Only if the agent wishes to make a moral judgment does rationality ensure that he issues utilitarian prescriptions. But even if someone does want to be moral it is arguable

that this bears the price of acting irrationally in the sense that it requires imprudence. For as we have seen, the self-interested agent will ensure that he has the dispositions required by critical prudential thinking rather than critical moral thinking. It seems likely that critical prudential thinking will not lead one to be totally immoral or criminal (the routes which we saw led to some convergence of morality and prudence ensure this) but will adhere to conventional morality, be non-self-centred enough to strike up proper relationships, but not make the sacrifices required by Utilitarianism.

Chapter III Hare's Method: Appeal to Linguistic Intuitions

If the argument of the previous chapter is correct then Hare cannot claim that on grounds of prudence alone a man should act as prescribed by critical thinking: though the more limited proposal that logic and prudence in tandem make it rational to do so is possibly more defensible. In any case we have as yet done nothing to dampen what would be for Hare a more important assertion, namely that critical thinking provides a rational way of determining what one (morally) ought to do, assuming that one wants to do the morally correct act. For Hare could quite rightly point out not only that the majority of people do at least claim that they are doing the right thing and besides that it is the main task of the moral philospher to analyse our moral concepts and that it is only a secondary - though highly desirable purpose - to actually have some effect on the way people live. For the remainder of Section I we will therefore consider whether Hare succeeds in this task, i.e. whether H1 - H11 succeed in establishing that critical thinking is the rational way to answer moral questions.

The most fundamental premise of the whole argument is that the way to make progress in moral philosophy is to analyse the moral concepts and that discovering their logical properties will enable one to lay down restrictions on the way moral words can be used. The key assumption of MT - i.e. that the logical properties of the moral concepts are, in harness, so powerful that they constrain all moral thinkers to be unanimous in their approval of a situation: unanimous if they are fully informed, rational, and using the moral words correctly. The great attraction of Hare's system is that moral conclusions are arrived at which are in a sense both objective and rational without anything being appealed to but the logic of the moral concepts. In particular, at no stage does Hare feed into the argument his own moral beliefs. Nowhere does he say anything like "It is self-evident that preferences are morally relevant", nor that anyone brought up properly could not fail to see that justice is less important than preference-satisfaction. It is quite possible that Hare has these moral intuitions but his point would be that they should be regarded precisely as these - i.e. they belong to the intuitive level of moral thinking, and are therefore in need of justification via critical thinking. It may well be that preference-satisfaction is morally relevant, but Hare's position is that this should be the output rather than the input of a moral system. Only formal consideraitons (i.e. the logic of the moral concepts) should be relied upon: substantial moral judgments are ruled out for purposes of justifying a moral system.

Hare's method holds an attraction proportionate to the defects of its competitors. Relying on the plausibility of moral judgments is unsatisfactory, partly because, by the nature of the subject, they cannot be proved or even tested in any conclusive way. Ultimate moral principles, by virtue of their being ultimate, admit of no proof and appear to be simply a matter of irrational choice. To be sure one can test out our moral principles for consistency with each other and can in some Rawlsian sense arrive at a "reflective equilibrium" after rejecting those which can be rejected at least cost to our initial moral system; but there is no guarantee that these principles will not conflict with that arrived at by another person. Such an approach encourages subjectivism as a meta-ethical theory or even relativism, since it is arguable that different cultures produce different sets of principles about which final arbitration is impossible. The real problem with relying on moral intuitions is not just that there is no accepted decision procedure for determining the correct intuitions, but it is also difficult to see what such a decision procedure could be. Any moral philosopher who, like Rawls, relies on the reader agreeing with his moral intuitions runs the risk of him not doing so, in which case the whole work becomes futile. It is not so much that a writer like Rawls has not good enough arguments to convince someone that his moral intuitions are wrong: it is rather that rational argument about such things is not really possible.

In contrast Hare would claim that his method has all the merits of any scientific empirical study. For his method involves investigating the logical properties of the moral concepts, a task concerning which rational agreement certainly should be possible. Hare is not alone in rejecting an appeal to moral intuitions: indeed his position has marked affinities with descriptivism. Descriptivists are those who, like Hare, think that a study of the moral concepts is essential but who, unlike Hare, think that a straightforward account of their meaning in terms of what facts they state can be given. Hare cites Mrs. Foot as a contemporary descriptivist. Mill would have been had Moore's account of him been correct and Moore himself would be if we take it that he thought 'good' referred to some non-natural indefinable property. The typical descriptivist approach (Moore as a non-naturalist would be an exception) when studying the moral concept 'X' is to give an account of 'X' in terms of some other concepts 'Y' and 'Z' and the claim that to say something is 'X' is, let us say, to state that it has properties 'Y' and 'Z'. Consequently the way forward in moral philosophy is to give definitions of the moral words (where possible in terms of non-moral words); then it will be relatively straightforward to answer questions containing that moral concept. For instance a descriptivist might suggest that 'good' simply means

"productive of pleasure", in which case it will be comparatively an easy task to determine whether something is good.

Hare's quarrel with descriptivists is not in their reliance on the analysis of language, which he goes along with, but rather in their over-simple account of the meaning of the moral words. Hare does not think that moral judgments state facts at all, rather they prescribe courses of action: hence any attempt to say that 'good' means 'has property X' is foredoomed. In MT4 Hare provides a strong argument for rejecting descriptivism. He admits that with some moral words - which he calls "secondary evaluative terms" - descriptivism is fairly plausible. This is because evaluative terms contain both a descriptive and a prescriptive element, and with some words - like rude, cruel and courageous - the descriptive element is relatively well-established. It is plausible to maintain both that if one says that an action is cruel one is evaluating it adversely and also that certain actions necessarily and uncontroversially qualify as being cruel. From these two facts it is tempting to conclude that, with these words at least, the descriptivists succeed in reaching an evaluative conclusion from purely non-evaluative statements of fact, i.e. from the fact that action X 3as such-and-such, it indisputably follows that it was cruel, which implies that it ought not to be done. Descriptivists claim that having derived an 'ought' from an 'is' they have succeeded in their intention of using the meaning of words to determine what is morally correct.

Hare's objection is that although 'seconday evaluative terms' which combine a definitive type of action or trait of character with approval or disapproval of it do exist, if one does not agree with the normative overtones of the word there is always the option of not using it. That is, it may be that if someone does X he must be courageous (or industrious, lazy, just, etc.) and that if I call him courageous I am necessarily commending him. But, quite simply, if I don't want to commend him all I have to do is to refrain from using the word "courageous" in these circumstances. I have the option of using evaluatively neutral language instead.

Hare states the objection neatly when he points out (MT p.19) that with any of these secondary evaluative words we can always ask "Granted that it would be X (i.e. cruel, lazy, etc.) would it be wrong?" Now although Hare does not distinguish them it is most important to see that this objection succeeds for two reasons. One reason is that these words are somewhat specific, i.e. one may be appraising only one aspect of the situation. For example, even if I admit that an act was courageous, I can still answer "No" to the question "Was it the right action?" if there are other features of it which would (in my view) merit condemnation.

In other words the pro-attitude I express implicitly carries with it a ceteris-paribus clause. If an act is both courageous and treacherous I am not committed to either of the views that it is right or wrong. The upshot of all this should be obvious: the descriptivist's attempt to derive an 'ought' from an 'is' fails because the evaluation is not as general as an 'ought' statement is.

The second objection to the descriptivist is more fundamental. This is the possibility of the moral reformer; the man who agrees that the act satisfies the descriptive criteria for being (e.g.) courageous, but does not concur with even the view that it ought to be done ceteris paribus: he may think its being courageous a matter for moral disapprobation, or may think that it is not morally relevant at all. This man - the moral reformer - is, according to Hare, not committed to evaluating the action favourably merely by the existence in the English language of the word courageous. He is quite at liberty to refrain from saying "He is courageous" (which would imply "he ought (ceteris paribus) to have done it") and instead say something like "He faced danger willingly but it is not the case that he ought to have done it") (even ceteris paribus). As moral reform takes place in society as a whole, this is reflected by the falling into obsolescence of some of these evaluative words. For example, the terms "chaste" or "chivalrous" becoming rare owes its cause not only to the condition or acts they describe becoming correspondingly rare (though this is also a contributory factor): even when they do occur people don't use these words because they do not want to express their approval. The consequence is that the descriptivists' efforts are wasted, since even if they could show that a very general moral word like 'ought' had some definite equivalent (like 'productive of most happiness') it would always be possible for the moral reformer not to use 'ought' but to turn to some alternative world of moral approval.

All this may appear to be in Hare's favour. We have so far argued against the two alternative methods: an appeal to moral intuitions, and the descriptivists' attemps to state the meaning of the moral words. But then the suspicion grows that what Hare is trying to do is not all that different from the descriptivists, so why should he be immune from the arguments he uses against them? After all, both he and the descriptivists take it that language is somehow binding in that discovering its meaning (or in Hare's case its logical properties) constrains us in what evaluations we can make. Hare says that his method is more sophisticated, in that he both allows the amoralist escape route (whereby one is free to say "It is not the case that I ought to do .....") and uses the combined logical properties of the moral words to constrain the agent

rather than some suggested synonym. But it is interesting to see how Hare fares when we turn his own objections to descriptivism against his own methods.

Obviously the first objection, that words like 'courageous' are not sufficiently general and so generate only ceteris-paribus 'ought' statements is irrelevant when we consider 'ought' statements themselves. But the possibility of the moral reformer seems to pose as great a difficulty for Hare as it does for the descriptivist. If we are free not to use the word 'courageous' if we do not want to give our moral approval to people who face danger willingly, why exactly cannot we do the same with 'ought' if we do not think that preference-satisfaction is the fundamental source of value? Even were it the case that 'ought' was almost unanimously acknowledged to be universalisable and prescriptive and to consequently commit the user to preference-utilitarianism, would not the moral reformer refrain from using 'ought' and use alternative moral words? (and perhaps use 'ought' in the inverted commas sense that whilst an action maximises preference-satisfaction he does not endorse it.)

In Freedom and Reason (pp 200-201) Hare counters the suggestion by arguing that whilst the man who disapproves of courage can flee to the more general word 'ought', the man described in the last paragraph has no such option. But once we have separated out the two reasons why 'ought' does not follow from 'is' we can see that this argument is not conclusive. On the basis of his own argument, Hare needs to allow for the possibility of the moral reformer. Now it may be that a particular society's language had no alternative general moral word other than 'ought'. But then why should the agent not invent a new word, or use some expression like "the morally right act"?

But perhaps Hare could reply that for a particular society the logical properties or meaning of the central moral word - i.e. 'ought' - defines what is to count as being moral, i.e. the word 'moral' itself only gets its meaning from the moral words. If this claim could be sustained then Hare could say that there would be no room for someone to disagree with an 'ought' statement based on critical thinking and yet still to be able to claim he was being moral. The trouble with this claim is that it seems arbitrarily to tie the meaning of "morality" down to the use of a moral concept one particular society happens to have. Even if (as we are granting for the sake of argument) 'ought' as used in our society is universalisable and prescriptive, we have no right to suppose that all societies will have identical words with identical logical properties. Hare's method might give normative conclusions he likes in our society, but he might not be quite so happy with those

that would result from an analysis of other possible societies' moral concepts.

In order to see this the following "1984" type fable might be of some assistance: imagine that a totalitarian regime is established at some future time in Britain, which sets out to control its subjects' lives. Naturally it is frequently making pronouncements like "You ought to do X" and "You ought to do Y", where 'X' and 'Y' reflect the State's values. Furthermore, suppose that the regime is entirely successful; through its propaganda machine it eventually has all its citizens believing that the state is the sole purveyor of wisdom and should be the ultimate authority on how the people should conduct their lives.

What would Hare's method tell us about what we ought to do in such a society? Admittedly Hare would eschew direct appeal to moral intuitions - which would have told us that we ought to do X and Y, where these are the acts ordained by the State. Hare's method would be to test the various hypothesis about the use of the word 'ought'. For example, to test whether a suggested word or phrase was a synonym for 'ought' one would see whether asserting that it was the case that it was not Z (where Z is the suggested synonym for 'ought') leads to self-contradiction in the way in which it is self-contradiction to say "There are a dozen eggs in this basket, but there are not twelve". Now it may be that in our imagined society "You ought to do it, though it is forbidden by the state" meets with the same reaction - i.e. the citizens linguistic intuitions might well tell them that this statement amounts to self-contradiction. If this were so Hare will have no option but to conclude that 'ought' meant 'recommended by the state'.

What would have happened in this society would be that the belief in the State's moral authority is so well entrenched that its language has come to embody this belief. Language is to some extent fluid: people have associated "X being right" with "X being recommended by the state" so frequently that the latter in time becomes the descriptive meaning of " being right".

It is not difficult to see what the effect of claiming that the central moral word of a society constitutes its morality would be in this society. Since 'ought' means 'prescribed by the state' anyone who disagreed with the state would be labelled immoral (or amoral). The effect would consequently be merely to endorse the moral practices of that society and rule out the possibility of the moral reformer. I am quite sure that Hare would not want to admit that linguistic analysis would endorse the conventional morality in the "1984" morality: but I find it difficult to give any convincing reasons as to why it should

do anything else. After all it is meant to be a virtue of Hare's method that it is a scientific enquiry into the logical properties of words and takes no account of our moral prejudices. But if this society existed it would be our moral prejudices rather than empirical research which would inform us that we ought not to do what the state told us to do. Perhaps Hare could claim that in the "1984" society the moral language had become perverted. On page 15 (MT) he does indeed say that he is "convinced that our ordinary moral concepts are serviceable". But what criteria can one use in assessing the moral concepts one already had? One wants to say something like "It would be best to use 'ought' in our sense", but assuming that this is a moral 'best' and that there is an intimate link between 'ought' and 'best', then this appraisal would appear somewhat circular. We would be appraising our own concepts in terms of themselves: compare those in the "1984" society saying that by their own concepts they ought to use their sense of 'ought' rather than ours.

Of course 'serviceable' can mean something other than "the concepts we morally ought to have". There are a number of ways in which moral concepts could be criticised. They could become so vague or confused that no-one knew precisely what was meant by them (as McIntyre argues that our present concepts actually are in "After Virtue"). Or they may rest on false factual beliefs. Or it could be that some concept a society has is so alien from what we understand by a moral concept that it cannot be called moral. But there are very many different sets of moral concepts which could survive all these tests: very likely those of the "1984" society could. So to say that our present concepts are serviceable hardly provides conclusive reasons for accepting them. But then again it is difficult to see what sort of reasons could be given for accepting (or rejecting) them, other than those just given. One suspects that the reason Hare thinks our concepts (as he describes them) are serviceable is that he finds himself in broad agreement with the normative results of linguistic analysis on them. But this is nothing but a moral intuition, which is of course inadmissible. If Hare has to say that linguistic analysis is justifiable in our society but not in the "1984" society, because in the "1984" society the moral concepts they have are not conducive to human flourishing, then the whole operation would become pointless. For then his system would ultimately rest on the moral intuition that human flourishing was what matters - and the whole point of linguistic analysis is to avoid having to base a system on moral intuitions.

Hare does give another defence for his method in MT page 18. He says "We come into moral philosophy asking certain moral questions, and the questions are posed in certain concepts. If we go on trying to answer those questions we are stuck with those concepts."

My objection is that had Hare been brought up in the "1984" society he would, on this line of reasoning, have to be satisfied with his investigations concluded that 'ought' meant 'prescribed by the state'. It is also arguable whether or not "What I ought to do?" is the only question of moral importance. There are also the questions "What rights have I?" and "What is the just action?" But on his own terms Hare has a good defence against this argument, namely that these concepts apply only at the intuitive level or moral thinking and are to be determined by critical thinking (which asks the question "What ought I to do?").

The possibility of the moral reformer seems to pose a serious threat to Hare's project. Even if 'ought' is both universalisable and prescriptive and the rest of Hare's argument is sound, there remains the further question of why anyone should accept the concept as he finds it, any more than one has to use the word 'courageous' just because it exists in English. Any temptation to suggest that moral reform is possible with 'courageous' but not 'ought' should be resisted if my "1984" example is taken into account, for in this scenario moral reform would seem, to us, highly desirable. But given that there are a number of different, alternative moral concepts we could use, just how do we determine which is best? Once we have filtered out the inconsistent and incoherent, only our moral intuitions can tell us what to use, e.g. whether to use the language of rights or the language of 'ought' or the language of some 'ought' in an imaginary society. Yet appeal to moral intuitions is inadmissible: consequently it would seem that there is no valid method of choosing between different sets of concepts. So it would seem that in the final analysis appeal to linguistic intuitions gets us no further and appeal to moral intuitions.

But perhaps there is one last defence Hare can make. The assumption has been made that there exist (or could exist) many conflicting sets of moral concepts each of which will, after linguistic analysis, generate different normative conclusions. To what extent different cultures actually do produce different moral concepts (all of which are still somehow recognisably moral) is an empirical question which I am not able to answer. But in our own society we have enough moral concepts which, if they do not actually conflict, at the very least stress different mspects of 'ought', rights, the virtues, justice. On the face of it McIntyre's argument that we have a vast array of concepts

with different origins is very plausible. But it may well be that Hare would dispute this claim; certainly he would argue that 'ought' is in some way the most fundamental of our moral concepts. By so doing he would evade the previous argument: if he could maintain that any moral concept had to be (or to he derived from) universalisability and prescriptivity then it would follow that no competing concept would count as moral. There is evidence that Hare holds this view. In MT (p.187) he says:-

> "I can dispute with (Homer and Aristotle) and with Nietzche if they are prepared to offer some affirmative universal prescription (as no doubt they would be)".

i.e. he thinks that even the above thinkers, who certainly used different concepts from our own, would, in making moral judgments, make universal prescriptions. More directly in Freedom and Reason (p.201), he says:

> "If a man wants to flee from my concepts, where is he going to fleet to? To singular prescriptions expressing selfish desires?"

But is the view that universalisability and prescriptivity are both unanimously agreed upon necessary and sufficient conditions for a judgment being moral very plausible? Although argument about such a matter is not easy to decide, it is my guess that Hare's claim - if this is his view - is wrong. To begin with, note that Hare contrasts universalisable prescriptions with selfish ones. But a non-universalisable prescription need not be selfish; anyone might be prepared to make a sacrifice oneself, but not necessarily say that everyone in the same position as you ought to. Equally, although we would normally say that one should have some inclination towards doing what we think is right, prescriptivity is hardly uncontroversial in claiming that we must actually, ceteris paribus, do it given the opportunity: much more will be said on this in the next chapter. My point is simply that whether or not 'ought' is universalisable and prescriptive this is hardly obvious and uncontroversial claim it would surely be were this defence to succeed. It would seem that Hare would be legislating were he to insist that the term 'moral' be confined to universal prescriptions. It would be more in line with common parlance to call a judgment moral if it is non-self-interested and seriously affects interest or rights of others. Hare would be mistaken on empirical grounds if he claimed that all concepts which have been thought at some time by some people to be moral have been prescriptive and universalisable, e.g. moral terms as employed by existentialists.

It would be a little unfair to suggest that these considerations count only against Hare. If I am right then appeal to neither linguistic nor moral intuitions has any final authority. Consequently it is difficult to see how any moral philosopher can make the foundations of his system solid. Perhaps Hare's most honest position would be to say that he has set out the constraints imposed by one set of moral concepts and then to claim that by a moral intuition these moral concepts are preferable to other possible ones.

Chapter IV Prescriptivity and Overridingness

The main conclusion to be drawn from the preceding chapters is that Hare's justification for his system is inadeuqate. In both chapters I have assumed that his account of the logical properties of the moral concepts and their implications is correct and have queried the sanction of those logical properties from two different standpoints. In Chapter II it was suggested that Hare gives no telling reason why a rational agent should use moral concepts at all; in Chapter III, I argued that even if someone does want to be moral, it is not obvious that he should be committed to using a moral concept like 'ought' which we happen to have in our language. But even if these objections are correct, Hare's project may not be irreparably damaged. He can claim that at the very least he has given an accurate account of the logical implications of the concepts we have, and that if we wish to carry on using them we should be aware of them, i.e. aware that rationality and the logical properties of the moral concepts constrain us to issuing utilitarian prescriptions.

We must now turn to examining whether this last claim can be vindicated. In this chapter I want to examine Hare's claim that 'ought' is prescriptive. Hare defines prescriptivity on page 21 of MT:

> "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."

Prescriptivity is closely connected to, but should not be confused with, overridingness. Hare defines overridingness on p.56 of MT:
"To treat a principle as overriding is to let it always override other principles when they conflict with it and in the same way let it override other prescriptions, including non-universalisable ones."

Hare says that at the critical level moral principles are both prescriptive and overriding. Prescriptivity implies that if one holds that "In situation S, one ought to do X" then we must usually do X if one gets into situation S: it therefore might be thought that one must be treating the principle as overriding. In other words, since prescriptivity entails action (under the right circumstances) it is easy to conclude that it entails overridingness, and that overridingness is a necessary condition of prescriptivity. However, although the definition on MT p.21 does not make this clear, elsewhere Hare says that the two requirements are distinct, since moral principles at the intuitive level must be prescriptive but need not be overriding. That this has to be the case is obvious, when we recall that moral principles can conflict with each other. Therefore it would be nonsense to say that they could both be

overriding. Hare's position is that at the intuitive level some 'ought' judgments are overridable - but only by other moral 'ought' judgments. So the statement "I ought to keep this promise" is still prescriptive if on an occasion it is overridden by the principle "I ought to help a friend": but not if it is overridden by the non-universalisable prescription "Let me enjoy myself". Thus prescriptivity can be seen as saying that moral principles are overriding over non-moral principles. At the critical level of moral thinking, where principles can be framed in such detail that they do not conflict, moral principles are held by Hare to be both prescriptive and overriding.

It seems to me that prescriptivity as Hare defines it is a pretty extreme requirement. Prescriptivity is plausible insofar as it asserts the connection between making a moral judgment and action. The opposing doctrine - descriptivism - states that the meaning of a moral judgment is fully exhausted by statements relating to what properties an action has; e.g. "X is right" means "X is approved of by most people" or "X produces most happiness". Hare's view is that moral words are partially descriptive but also have a prescriptive element - in fact their main purpose is to guide conduct. Some words are obviously purely descriptive e.g. "horse". To call something a horse is merely to say that it conforms to having certain properties; it is not in itself a way of commending it. In contrast a prescriptive statement need not state a fact at all - e.g. the imperative "Do it!" Hare's suggestion is that moral statements have such an affinity with imperatives that they cannot be considered purely descriptive. Compare for instance "X is a black horse" and "X is the best horse". The former statement would only actually be a way of commending it (e.g. if you know that the hearer liked black horses): in itself the word "black" is purely descriptive. Conversely "best" is prescriptive in that to say that it is the best horse is not so much to say that it has certain properties (although it may well be doing this as well) as that it is the one that you would pick, given that you were going to pick a horse, ceteris paribus. Hare would say that a moral "best" shared with this non-moral "best" the property of being prescriptive.

Now it seems quite right to say that moral statements do in general aim to guide conduct and do in general commit the speaker to have some inclination to see that the prescribed state of affairs is brought about, given that he is in a position to do so.

But Hare goes much further than this: he is not merely saying that if we say "You ought to do X" we have some motivation towards doing X; he is going so far as to say that if we can do it, and no other moral judgments override it, we will do it. Two doctrines can be distinguished, which for obvious reasons I will label Ps and Pw.

Ps: If an agent says he ought to do X, then given that he is in a position to do so he must do it (or withdraw his 'ought' statement) unless he is also committed to another 'ought' statement which overrides it.

Pw: If an agent says he ought to do X, he must not treat the doing of X with total indifference: he must in certain circumstances actually be motivated towards doing X: though in certain circumstances he may let an 'ought' statement be overridden by a non-moral prescription.

Ps affirms that an agent will never let an 'ought' statement be overridden by a non-moral prescription whereas Pw allows for this possibility. The trouble with Ps is that it denies the existence of a very well-known phenomenon, i.e. backsliding. Backsliding or moral weakness occurs precisely when an agent does X despite his believing that he ought not to do X, when he is not somehow forced into doing X. Common sense informs us that moral weakness is a feature of everyday life: certainly most people would not claim that they did everything that they thought they ought to do. Before we go into the prescriptivist's possible explanation of backsliding we should perhaps look into the reason why Hare holds Ps and not merely Pw (according to Pw, of course, moral weakness can occur). In order to do this we need to jump ahead of ourselves a little in our examination of the argument H1 - H11. For the reason why Pw will not suffice for Hare is to be found in the transition from H9 - H10. By means of steps H5 - H 9 Hare hopes to have established that at the end of critical thinking I must have a preference of the form "Let X be done" or "Let X not be done" corresponding to the preferences of all those people affected by the action X (including myself). But without Ps it would not be obvious what bearing this fact would have on what I say ought to be done. Why must I say that X ought to be done merely because I prefer on balance that it should be done? Ps provides the answer: I can say "X ought to be done" sincerely only if I am going to actually do X in the circumstances, which for Hare means that I must prefer on balance that X be done. The argument works like this: H5 - H7 ensure that the agent prefers on balance that the utilitarian action is done, and prescriptivity ensures that only that prescription which one prefers on balance be done can be made in an 'ought' statement. Prescriptivity rules out any non-utilitarian 'ought' statements, e.g. suppose that someone started off with the non-utilitarian prescription that he ought to exact revenge on an enemy. Steps H5 - H9 (if correct) would mean he would prefer that the utilitarian outcome would prevail (i.e. that revenge be not exacted); so prescriptivity would entail that he withdraw his original 'ought' statement. We can now see why Pw will not do for Hare; for if this were true the seeker of revenge could still hold his statement that he ought to exact revenge, since he

could in some way still be instructed to do so. He does have a preference to exact revenge; but his overall preference is to refrain from doing so. Consequently Pw would not be sufficient for Hare's argument for utilitarianism to go through.

So Hare is committed to defending Ps, which means that he must somehow explain away backsliding. I will first consider three possible accounts of backsliding which would save Ps, and then give both empirical and theoretical reasons why I think these fail.

First we must consider the account of backsliding given by Hare in MT. In MT III he suggests that cases of moral weakness can be compared with those of moral conflict. The latter phenomenon may at first sight appear to be an equally effective refutation of his theory. For since at the intuitive level moral principles conflict they cannot all be acted upon: so how can they be prescriptive? Hare's answer is that whilst they cannot all be overriding they can retain their prescriptivity. He says that "if applied they (the principles) would require a certain action, but we just do not apply them in a certain case". He then makes this relevant to moral weakness by suggesting that much the same happens here: we do not apply a (still supposedly prescriptive) moral principle in favour of a non-moral prescription.

All of this is somewhat curious. What, if anything, does it mean to say that "we hold a principle, but do not apply it" (MT p.59). For it appeared to be Hare's previous view that to hold a principle one must, in the appropriate circumstances, apply it. The whole point of prescriptivity is that it rules out my saying that I ought to give away all my income to Oxfam if I have not the slightest intention of doing so. Yet on the revised account I could hold this principle but simply not apply it. Perhaps what Hare is getting at is that to hold a moral principle one need not act on it every time one has the opportunity: it might be that one can sincerely say that one ought to do X in S if, say 80% of the time in S we do X. This is perhaps a more realistic view of what actually happens but it won't do for Hare, since in the first instance it contradicts his definition of prescriptivity (if it is consistent with Pw but not Ps) and secondly it won't plug the gap between H9 and H10: on this view I would say "I ought to take revenge" yet in this situation not act on it, saying it was one of the 20% of occasions when I do not 'apply' the principle.

It seems to me that this solution works for neither moral conflicts nor weakness and that whilst the first can be solved in a different way

the latter problem cannot. The problem with moral conflicts is caused solely by the principles' generality. For example, someone might hold both: (1) One ought not lie

(2) One ought not to hurt peoples' feelings.

One might easily find onself in a situation where one cannot avoid breaking one of these moral principles. This does of course imply that they cannot both be overriding. But if they are still to be Ps they must be overriding over non-moral princples. As I have argued, Hare's suggestion that one does not 'apply' one of them is unsatisfactory since it allows in cases where the applied principle is not a moral one. But all one has to do to preserve the prescriptivity of principles (1) and (2) is to suggest that they are subject to the tacit qualification: "provided that there is no feature of the situation which critical thinking informs me brings a competing and overriding moral principle into play". If this is done then clearly whichever is applied neither (1) nor (2) is breached.

However such a solution is not possible in the case of moral weakness. For then the qualification would have to be something like "provided there is no feature of the situation which means it is greatly in my interests not to apply it"; which is plainly non-univeralisable. The fact of the matter is that cases of moral weakness are precisely those states where a non-universalisable prescription overrides a supposedly prescriptive universalisable one: here any solution must, for it to work for Hare, deny that moral weakness proper ever occurs.

One such solution would be to claim that in all cases of supposed moral weakness the agent doesn't really accept the principle. No doubt there are cases where this is a suitable characterisation of the situation: cases where we are using an inverted commas sense of 'ought' and really mean something like "it is generally held that one ought to keep promises" and not saying that it is a principle we ourselves hold. Again it could be that the agent is insincere - he is pretending to us that he holds a moral principle himself but knows he doesn't really - or perhaps is even deceiving himself - has even convinced himself he believes that he ought to do X, but somehow in reality he does not. All these are quite possible, but it seems rather unlikely that they can be made to cover every instance of moral weakness. A problem here is that whatever counterexample one brings forth, it is always open for the prescriptivist to counterclaim that for some reason this isn't a bona fide case of an agent fully believing that he ought to do X yet not doing it in a situation where he can. As has been allowed, this can be a legitimate ploy, but if carried to extremes it can make the prescriptivists's

argument circular. No doubt to him it isn't a proper use of 'ought' since for it to be so an agent must actually do the action when he is able to. But this is the very assertion that is being tested, so it can hardly be assumed. Rather we must have an open mind about whether prescriptivism is true, and use our linguistic intuitions to see if ordinary language allows that this is a proper use of 'ought' (my argument of the previous chapter has not denied that linguistic intuitions cannot be used to see whether a word - as it is used - has certain properties: I suggested that we need further reasons as to why we should use the word in this sense).

So any argument is bound to be inconclusive, since the prescriptivist's linguistic intuitions are bound to tell them that in all cases of moral weakness the agent doesn't fully believe he ought to do X (or else that there is some other reason why it is not moral weakness proper). However, I will pretend this problem doesn't exist and hope that the majority of people will agree with me that in the following example the agent does believe that he ought to do what he refrains from doing, and that this is in no way an improper or weak sense of 'ought' The example I have in mind is as follows:

A bank clerk considers that it is wrong to steal money. He has never stolen any before and whenever he reads in the newspapers of burglaries he is filled with genuine moral disapprobation. Like most people, he believes that stealing is wrong, and proof of this is provided by both what he says and what he does. To make this last point more persuasive, letus suppose that for very many years he has known a foolproof way of stealing a large quantity of money from his bank without anyone ever knowing he was responsible. Let us assume that such is the ingenuity of this scheme that the probability of his being found out is negligible. Yet he does not steal the money, not because it is against his interest to steal, but because he believes it would be wrong. We have here, I suggest, as bona fide a case as any of someone believing it is wrong to steal. Now supposing that through some unsuccessful business enterprise he loses his life savings and, filled with the prospect of having to sell his house and car, he decides to undertake his scheme which involves robbing the bank and doing what he thinks is wrong. We can even suppose that at the very moment he is undertaking the robbery eh is feeling considerable moral guilt, and afterwards, though he is never suspected, he regrets that he had to do the act (though would still do it again if the same situation arose). I submit that in these circumstances if freed from theoretical preconceptions most people would say that the man did what he thought he ought not to do. Of course the prescriptivist could try to maintain that this was not the case; he could say (not

very plausibly) that the man had changed his moral principles. But to repeat, on Hare's own admission, one must take people's linguistic intuitions about how to use the word 'ought' as binding: so if my claim that most people would think this was a valid use of 'ought' is correct then this defence of prescriptivism fails.

At no point does Hare himself place much emphasis on the above defence: I mention it only because it rules quite a number of alleged cases of moral weakness. Hare himself, in Freedom and Reason, argues that if the agent accepts a moral principle but fails to act on it then it must be the case that he was psychologically unable to follow it. The suggestion is lent credence by analogy to cases of physical compulsion; i.e. we would not say a man believed he ought to lie if he only did so when someone was pointing a gun at his head. The prescriptivist is saying that, if he could, the agent would adhere to his moral principles: but psychological forces deprive him of his free will. Clearly there are examples where this happens, e.g. kleptomania. But is it really plausible to suggest that this happens in every case of moral weakness? Take our bank clerk again. The prescriptivist could claim that he was forced by his impending ruin to act as he did and that psychologically he could not cope with the shame inherent in losing his possessions. The trouble with this sort of suggestion is that in a sense it is empty, since it is impossible to prove one way or the other. As Taylor argues in his review of "Freedom and Reason", Hare needs to give a separate account of "psychologically unable". It is no good at all simply to assume that in cases where moral weakness occurs the agent was psychologically unable to do otherwise; this would be circular and would make prescriptivism true whatever the world was like. If Hare can tell us what it is for someone to be psychologically unable then we can investigate as to whether those prone to moral weakness actually are unable to do otherwise. But in Freedom and Reason he provides no such account. Furthermore I suspect that were he to do so this view would not only be falsifiable, it would also be falsified. A reasonable first attempt at a definition of "psychologically unable" might be that "the agent could not have done otherwise, whatever he had decided". Of course in a sense we are always "overcome by desire" in that it is analytic that the strongest desire wins out. But in the majority of cases this does not preclude our making a rational and deliberate choice to do the immoral act. Not all cases of moral weakness can be said to be cases where the agent could not have done otherwise, whatever he had chosen.

As I have been at pains to stress, the above arguments must have an air of inconclusiveness about them.

Not only do they rest on hypothetical examples, which it could be claimed are suspect in that they are atypical, but also they rest on my linguistic intuitions which, again, are in no way authoritative. Moreover, there is always an alternative position for Hare. He could say that there are grey areas in our language and that 'ought' is used as I have suggested so that it allows moral weakness. But he could suggest that his theory is true of an ideal language, where 'ought' was used in the way he suggests. No one could suggest that a language could not exist where 'ought' was Ps: I have only said that our 'ought' is not. Of course this revised Harean position would entail some alterations in the argument: he could no longer claim that he was merely analysing the concepts as he found them. But if the argument of the previous chapter was correct, the fact that we use 'ought' in a certain way gives it no sanction over us anyway. So it would not weaken his position.

I now wish to present an argument that will both dispel any feelings of unease that I am relying too much on anecdotes and also show that appeal to an ideal language will not do. This argument owes much to Singer's "The Triviality of the debate over the 'is-ought' and the definition of 'moral" (APQ 1973). Prescriptivism appears meritorious in that it ties an agent's moral principles to his action. According to prescriptivism, if I think I ought to do X, I will do X if I can. But in reality different people are motivated by different sorts of things, and in consequence they treat different considerations as being overriding. If we knew enough about someone we could make a list in order of priority of all the things he considered important. Now it seems a fairly indisputable fact that for some people moral concerns would be at the top of the list, for others selfish non-univeralisable ones would be and for yet others aesthetic considerations would be. It follows that in a given situation some will do the right act, some the self-interested one, and some that which is most aesthetically pleasing. I am not claiming that people can realistically be said to always do either the moral, self-centred or aesthetic act; rather that different people tend to give priorityto each. The self-interested man may usually do the selfish act, but where the self-sacrifice is not very great he may do the moral act. Now so long as 'moral' is taken to mean "what would commonly be regarded as moral", this in itself does not provide a refutation of prescriptivism. For Hare can reply that whatever sort of concern each man treats as overriding - that, for him, is his moral principle. Hare himself admits that were a man to treat aesthetic concerns as overriding then they would thereby become a moral principle. So at the cost of

making no constraint on content, prescriptivism can be saved. But it will mean that there are some fairly odd 'moral' principles. For instance the self-interested man would say that it was quite all right for him to harm others so long as it did not adversely affect himself in the long run.

If it is accepted that some people do not consider their moral principles as overriding then it follows that moral weakness occurs. But if we re-define their moral principles as being their overriding principles, then moral weakness will not occur. But then we cannot make any other constraint on moral principles other than they be overriding - unless all principles that are treated as being overriding already share this property. For instance we cannot say that all moral principles must give weight to the interests of others - because some overriding principles do not. This holds whether it is the form or content of moral principles we are putting a constraint upon. If Hare holds that moral principles are always overriding (at the critical level) he cannot also maintain that they are always universalisable, because some overriding principles are not universalisable. Take our self-interested man who holds it is right to treat others badly. He does not hold this in the form "It is all right for everyone to treat everyone else badly". He thinks that it is all right for him to do so, but that everyone else should be moral in their dealings with him. His overriding prescription "Let me treat others badly, and them treat me well" is non-universalisable.

The problem for prescriptivism is that the world is not such that people always do acts which can, in any normal sense of the word, be described as being moral. Neither is this always through some failure in rationality on their part: it might be that in their list of values moral principles do not come very high. Given this state of affairs, prescriptivism will be false if it insists that moral principles resemble what we now call moral principles, e.g. Hare's insistence that they are universalisable (which of course he needs for the argument H1 - H11 to go through). The alternative, to place no constraints on their form or content but to say that moral principles are those that are treated by the agent as overriding, is not very palatable either because it then becomes an empty doctrine, true by definition and one which is misleading in that its moral principles may be what we would give very different names.

Chapter V The Derivation of Utilitarianism

A central claim of "Moral Thinking" is that a rational agent will make moral judgments in line with utilitarianism. To the reader familiar with Hare's earlier works this might be a little surprising. In "Freedom and Reason" (FR) Hare gives a very similar account of the logical properties of the moral concepts - i.e. that they are universalisable and prescriptive - yet comes to a somewhat different conclusion concerning their impact on moral argument. It is worth quoting fairly fully what Hare had to say on this subject in FR p.97:

"..... it is of the utmost importance to stress that the fact that two people express the same thing by 'ought' does not entail that they share the same moral opinions. For the formal, logical properties of 'ought' .... are only one of the four factors (logic, the facts, people's inclinations and their imagination) whose combination governs a man's opinion on a given matter. Thus, ethics, the study of the logical properties of the moral words, remains morally neutral ... its bearing upon moral questions lies in this, that it makes logically impossible certain combinations of moral and other prescriptions. Two people who are using the word 'ought' in the same way may yet disagree about what ought to be done in a certain situation, either because they differ about the facts, or because one or other of them lacks imagination, or because their different inclinations make one reject some singular prescription which the other can accept."

The Hare of FR held that whilst the logic of the moral concepts entailed some constraints on what moral judgments an agent could make, there was still substantial scope for deciding upon one's own moral principles. The fanatic might have been awkward for Hare in that it meant he had no conclusive arguments against say, a Nazi who was willing to say that Jews should be exterminated even if he himself was a Jew; but it did mean that his non-descriptivism remained intact. Although people without strong ideals who agreed on the facts and had similar capacity for imagination would tend to agree on what was right, these qualifications meant that one could not derive an 'ought' from an 'is'. Two questions spring to mind: given that Hare's account of the logical properties is similar to that given in FR, how come he now thinks that utilitarianism is entailed when he didn't then and even if his new argument is correct, how can he claim that he is still a non-descriptivist?

It is not immediately obvious how prescriptivity (P) and universalisability (U) lead to even the constraints suggested in FR, i.e. that one had to consider the interests of those once affected, even if in the end one's ideals allow one to harm them. To see how this argument works it is hard to improve on Mackie's account of Hare's argument given in Chapter 4 of

of his book "Ethics: Inventing Right and Wrong". Mackie suggests that there are three stages of universalisation, each of which imposes further constraints on what the agent can prescribe. The first and weakest sense of universalisation rules out moral principles which mention names, times or places; i.e. one cannot say "I ought to be let off this crime". Strictly speaking a moral principle is allowable if it contains non-universal terms as long as the agent is willing to translate it into sentences which contain only universal terms, i.e. the above sentence would be all right as long as the agent was willing to assent to the sentence "All those in a relevantly similar position should be let off this crime". Interpreted in this way universalisation rules out the more flagrant egoist principles such as "I ought to be let off this crime but others should be punished". As Mackie points out, though, as it stands it is a relatively weak constraint: egoism is not ruled out as long as one can assent to the universal principle "Everyone should promote only their own happiness" and you can frame moral principles which are unfair in that they discriminate against people with different tastes, ideals, or in a different social position, e.g. if you are very wealthy you have no problem translating "Let me pay only 5% income tax" into "Let everyone only pay 5% income tax".

Hare's position in Freedom and Reason actually involves a stronger sense of universalisation than this (or more accurately it involves a combination of universalisation and prescriptivism). Not only has one got to exclude non-universal terms but also one has to put oneself in the place of those whom one might affect. Thus the wealthy man alluded to above may not be able sincerely to assent to the proposition "Let me receive the benefits consequent on everyone paying only 5% tax if I am poor and sick". There is a problem here though over the meaning of putting oneself in the other person's position. The idea is that the moral agent still has the freedom to choose what moral principles he likes, but he must be willing to have those principles carried out even if he was in the position of those whom they harmed. So the rich man has to suppose that he could be the poor man; in an example used by Hare the jazz fan has to imagine that he is in the position of someone who dislikes jazz music. It may be that a fanatical Nazi would assent to his being exterminated were it discovered that he was a Jew. But plainly even if one imagines one has the other person's social standing, and tastes, one would not fully be putting oneself in his position. In particular one won't be imagining oneself with his ideals. This then is the third stage of universalisation: to put oneself so thoroughly in the shoes of the other person that one has his ideals, tastes and social circumstances. But now the enterprise becomes somewhat incoherent: one is supposed to be judging - from the point of view of

one's own ideals - what one would say if one were in the position of someone else - with different ideals. I suggest that as it stands this test makes no sense.

As I have said, stage 2 universalisation matches Hare's position in Freedom and Reason. He argues (FR p.113) that stage 2 can in fact be derived from prescriptivity and stage 1 universalisation. Hare argues that "we would not be prepared to prescribe universally that people's likes and dislikes should be disregarded by other people, because this entails prescribing that other people should disregard our likes and dislikes". In other words we prefer the sort of world where others take note of differences in taste to one where people assume that everyone has the same likes as they. Even this is debatable in certain instances - e.g. a devotee of classical music may say that even were the jazz fan who detested classical music he would want classical music to be played. But then the man's preference for classical music would have turned from being a mere liking into being an ideal.

Stage 2 universalisation (which I will henceforth refer to as U2) provides a powerful basis for moral argument. Of course it is subject to all the criticisms levelled in Chapters 1 - 3 of the present work, but if these are waived U2 would settle many moral disputes. In making a moral statement the agent would be asked to imagine himself in the position of those whom it harms, with their likes and dislikes, and to decide whether he still wants the action to be done. He will be able to decide only if his desire for the act to be done is greater than any aversion he has in another's positon. One can therefore begin to see that U2 has some affinities with a desire - or preference-based utilitarianism. But plainly it does not come close to being identical with utilitarianism. Consider a bilateral case - e.g. Hare's example of the man contemplating putting another in jail, for non-payment of a debt. U2 says I must put myself in his place - with my ideals - and still assent to "Let me be put in prison." I will be able to do this only if my ideal that justice be done is stronger than my preference not to be put in prison. But preference-utilitarianism would balance all the preferences involved. In this case one can distinguish four preferences: my ideal (or moral preference) that justice be done; my non-moral preference that this man who has deceived me is harmed; his non-moral preference not to be put in prison and his moral preference (which could be anything). U2 does not weigh these four preferences: when I put myself in my position only my preferences are considered and when I put myself in his position only my moral and his non-moral preferences are taken into account. At no stage does U2 take his moral preference into account. Only in the special case where my overriding ideal was preference-utilitarianism

would all the preferences be taken into account - even though U2 would not be equivalent to utilitarianism, since his non-moral preference would be double-counted. So even in the bilateral case U2 is not equivalent to utilitarianism - a fact which Hare acknowledges by admitting that U2 cannot touch a fanatic who is prepared for even his interests to be harmed in order to allow his ideal to be fulfilled.

In multilateral cases there are additional reasons why U2 may conflict with utilitarianism. U2 only considers two preferences at any one time: my moral preference and someone else's non-moral preference. It is a good test in that it ensures that my moral preference is not cooked to favour my own non-moral preferences; but it cannot be claimed that it succeeds in taking into account everyone else's non-moral preferences. To be more precise, U2 as it stands is far more restrictive on what counts as a moral principle than preference-utilitarianism is. In the latter case I can say that I ought to do X if, on balance, the total preference-satisfaction of all concerned is maximised by my doing X with U2, though, I have to go round all the parties individually and in every case my moral preference to do X has to be greater than their aversion for X. To make this clearer, consider the example mentioned in Fr of the judge and the prisoner. Utilitarianism obviously sanctions a judge to send a convicted criminal to prison (assuming there is overwhelming evidence for his guilt, etc.) Under U2, however, the situation is by no means so clear. If the judge whole-heartedly puts himself in the position of the prisoner it may well be that he cannot sincerely assent to "Let me be put into prison" if his desire for freedom is stronger than his desire for justice. In this case it would not be that U2 endorsed the prisoner's view that he ought not to be put in prison; for when the judge puts himself in his own shoes - or in those of all those members of society who feel threatened by the liberty of the criminal - he cannot prescribe this either. So in the multilateral cases U2 seems much too strong; there will be many situations where no moral principle will pass the test.

Hare admits this but suggests that in the multilateral cases the strict procedure of U2 should be amended. Rather than putting oneself in each of those affected's positions one only has to apply U2 to the balance of everyone's preferences. So the judge doesn't have to prefer that were he the criminal he should be sentenced: rather he has to prefer on balance, taking everyone's preferences into account equally, that the criminal be jailed. One way of doing this would be to imagine that one had an equal chance of being any of those affected; e.g. suppose there are a million people in the country, including the criminal: the criminal prefers to be freed one hundred times as much as each of the others prefer that he be jailed. The judge will balance out the hundred "votes"

against jail with the 999,999 votes for; so he will be able to prescribe universally that he be sent to jail. Whilst this method clearly allows moral judgments to be made in a way in which U2 did not, it is by no means obvious how this move can be justified. In FR Hare sees it justified by the impartiality required by universalisability: one must treat everyone equally unless there is some relevant difference. From this it could be argued that everyone's preferences must be considered equally. But universalisation has nothing to do with preferences (it is prescriptivity that makes these relevant): it only says we should, in a certain sense, be impartial. To say it means we must be impartial amongst preferences is to assume that preferences are the only relevant moral factor. The position is as follows: In the case of U2 preferences are relevant because universalisability and prescriptivity entail that for me to say "I ought to do X" I must be prepared to have it done (i.e. prefer it) even when X harms myself. In multilateral cases U2 is too strong as it rules out even cases where just one party is seriously harmed: on the other hand, I do not see any justification for Hare's assumption that we must be prepared to prescribe according to the balance of preferences.

Another difficulty with Hare's account of multilateral cases in FR is that there is an indeterminacy concerning which preferences should be counted. Recall that the agent must act on the balance of preferences. But this could lead to any of three possibilities:

(1) The agent balances his moral preferences with the balance of everyone's non-moral preferences.

(2) The agent simply balances everyone's non-moral preferences.

(3) The agent balances out everyone's preferences, including their ideals.

Possibility (1) is most akin to U2 whilst (2) and (3) are obviously versions of utilitarianism. Version (1) still allows for the existence of the fanatic. All three possibilities would require a more satisfactory justification than is given in FR.

We can now see that FR left Hare's theory in a rather unsatisfactory state. U2 certainly follows from universalisability and prescriptivity but does not seem to work in multilateral cases. The account given of multilateral cases approximates to utilitarianism (though is not identical in that Hare allows for the existence of the non-utilitarian fanatic) but is hardly justified properly in terms of universalisability and prescriptivity. In MT Hare attempts to remedy this situation by, in effect, abandoning U2 for a utilitarian position (equivalent to version(3) above). To arrive at utilitarianism Hare adds arguments H5 - H10, which for the sake of convenience I reproduce below:

H5: Rationality requires of us that we know all the facts relating to the situation. These include facts about other people's preferences (in reality we will not be able to know all the facts, but we must attempt to discover all those which are relevant - i.e. affect the outcome of critical thinking).

H6: If I know that a person whom I may affect prefers with intensity L that action A be done (or not be done) then I must know that were I, hypothetically, to be put in his exact position (including having his preferences) I too would prefer that X be done with intensity L.

H7: I cannot know that if I were in that position I would prefer that X be done with intensity L (H6) unless I now prefer with intensity L that if, hypothetically, I were in his position, X should be done. i.e. H5 and H6 mean that I must now have a preference relating to what would happen to me were, hypothetically, I to be put in the other person's shoes.

H8: By universalisation I must make my hypothetical concern for other people actual, i.e. I must convert my hypothetical preference that, were I in his shoes, X should be done, into an actual preference that X be done (even though as a matter of fact I am not in his position).

H9: From H7 to H8 it follows that I now have a preference that X be done/not done, corresponding to the preferences of all those affected by X (including myself).

H10: By prescriptivity and overridingness (H4) I can say "X ought to be done" sincerely only if I prefer on balance that it be done. From H9 it is clear that the only action which I prefer to be done on balance will be that generated by preference utilitarianism.

Therefore what ought to be done is answered by appeal to preference utilitarianism.

In the rest of this chapter I wish to examine this argument in detail. H5 is important in that Hare needs to say that we must know what other peoples' preferences are: for according to H7, this means that I must actually have a corresponding hypothetical preference. At face value H5 seems unobjectionable; few would deny that we need to know all the relevant facts. What could be doubted is whether facts relating to preferences are the only relevant. Hare is not guilty of pre-supposing that they are relevant: rather he is saying "Imagine that they might be, see if given the logical properties of the moral words, this knowledge makes any difference, and if it does then it is relevant". Thus he could (to take a frivolous case) have said "Let's take as a candidate knowledge of peoples' names". He would then have found that this knowledge had

no bearing whatsoever on what we can prescribe universally; so it could be rejected as being relevant. Where Hare can perhaps be criticised is that, having discovered that knowledge about preferences is relevant, he does not then go on to question whether or not other facts are relevant, e.g. facts about the existence of God, or facts about what is objectively right.

H6, if it is true at all, is a tautology. The common sense response might be to suggest that just because Jones has, say a preference for junk food over homemade food doesn't mean that I would have if I were in his position. But of course if it is stipulated that I am in exactly his position then it follows that I must have his preferences. What can be doubted is whether the imagined person would really be me. If I am in exactly Jones's position then I must have exactly his (or her) likes and dislikes, ideals, sex, personality, social standing; in other words I am stripped of everything that makes me what I am. Hare's reply is that whilst this is true, the word "I" is not just a descriptive word, i.e. it doesn't just refer to the bundle of properties I happen to have, but also refers to me the prescriber, over and above the properties. In making this claim Hare may have common opinion on his side. We do tend to think that there is some essential "I" which exists independently of whatever embodiment we have, and that this "I" could exist even if any characteristics were different (it would still somehow be me having the experiences). Thus we accept that in Kafka's Metamorphosis the subject of the story could turn into an insect yet still be him; equally many people think they could exist in an afterlife (presumably in a very different form). However, this view has been challenged by Parfit (Philosophical Review 1971), who argues that instead of identity we should talk in terms of survival. Survival is a matter of degree and the extent to which we survive, according to Parfit, depends on the degree of psychological and physical continuity and similarity. If Parfit's view is correct, then it is not clear how someone with whom I have nothing in common can be me. However, as the questions of personal identity are notoriously difficult this can hardly be taken as anything approaching a refutation of Hare.

More serious criticism can be levelled against H7. By H5 and H6 Hare establishes that I must know that were I in Jones's shoes I would prefer X with strength L. But H7 goes further in claiming that this entails that I must now prefer with strength L that X should happen, were I in Jones's position. It is claimed that knowledge about preferences must affect my set of preferences. The first difficulty arises concerning whether or not my preference with strength L that X should happen is properly termed a preference. At best it can be called a hypothetical

preference in that - unlike most preferences - it does not relate to my actual circumstances, but only to hypothetical ones, i.e. normal preferences influence my action whereas hypothetical ones only do so if my circumstances change. My preference that were I Jones X should happen is on a par with my preference that were I say, a Russian, the Americans should suffer setbacks in their foreign policy. In both the latter cases the preference need have no influence whatsoever on what I do: my objection is that Hare may be misleading in saying that it is now amongst my set of preferences.

Leaving this difficulty to one side, it is dubious whether the move made in H7 is valid anyway. It is correct only if there is a necessity for me to identify with this hypothetical future self. But if one thinks of cases, where my hypothetical preferences are radically opposed to my present ones, this loses some of its plausibility. For example, suppose a liberal is told that he is going to be brainwashed into holding extreme right-wing views - so that his future me will prefer that racist policies are affected. In this case there seems no contradiction in saying that I should not, now, prefer that the preferences of the future me are satisfied. Hare considers this objection (MT p.96) and counters it by suggesting that "I" is partially prescriptive. By this he means that by identifying myself with some person hypothetically I automatically identify with his preferences. Phillips Griffiths (Philosophy 1983) argues against this that there is no necessity in "I" being prescriptive. At most all that is true is that normally we do tend to express considerably greater concern for ourselves (including our future selves) than for others. But this need not preclude a special sense of "I", which does not share this feature. In the case of the brainwashed liberal it would make perfect sense to say that that person was still him in that there was physical continuity, but still not identity with his preferences. It is arguable that the fact that we generally do want out own future preferences to be satisfied owes something to the arbitrariness of some preferences. By this I mean that with very many preferences (e.g. a preference for tea over coffee) we don't very much mind which way it goes, but once we have it we want it to be satisfied. That is, with some preferences we are not much concerned with what they are for; possibly this is because they don't figure prominently in our conception of a good life. We dont't very much mind whether we are the sort of person who likes tea or one who likes coffee. With these preferences it is obvious why we are ready to prescribe that they are satisfied in the future: since their object is irrelevant it doesn't matter if it changes. With other preferences though - political, moral and aesthetic ones - we are as much concerned with the object of the desire as with the fact that it is us desiring them. The difference is that if I were told that in ten years

time I would change my preference from tea to coffee, I would have no reason not to identify with this preference, whereas with preferences for things I think are intrinsically good this need not be the case. In fact I might wish that my preferences are satisfied whatever they are - if I think my views will be more mature and that they are developed autonomously. But as the brainwashing case shows we should have the right to withdraw our concern that our future preferences are satisfied in exceptional cases. So in the case where it is some other person we are meant to be identifying with, where the preferences are moral, or maybe they are irrational, it is doubtful whether I have to admit that I now have the preference that, were I in his situation, that preference should be satisfied.

So it is by no means certain that the argument of MT V works in establishing its intended conclusion, i.e. that if I have full knowledge of another person's preferences I shall myself have acquired preferences equal to his regarding what should be done to me were I in his situation. But even if this is granted, why does this, together with prescriptivity and universalisability, entail utilitarianism? The situation becomes clearer if it is compared with that in FR. As we saw U2 was too strong in multilateral cases, so Hare wanted to say that one had to balance out everyone's preferences. As I say, this is what he wanted, but it was by no means clear how he could justify such a step. The preceding argument appears to do just that. H5 - H7 ensures that I have a hypothetical concern for each person I affect. H8 claims that universalisability means that I must make this hypothetical concern actual: since I must balance out all the preferences of those whom I affect. According to Hare, the situation is akin to when our own preferences conflict and prudence requires that we act on the stronger preference. In his own words (MT p.110):-

> "I See no reason for not adopting the same solution here as we do in cases where our own preferences conflict with one another ..... Suppose that it is my own bicycle and that it is moderately inconvenient not to be able to park my car. I shall then naturally move the bicycle, thinking that that is what, prudentially speaking, I ought to do."

He continues by saying that in the bilateral cases if I have full information I will have acquired preferences equal to the other person's which are now conflicting with my original preferences. Since both the conflicting preferences are mine, I can deal with it precisely as I deal with cases where the preferences are mine actually: I prescribe what, on balance, I prefer.

However, McDermot (PQ 1983) has a persuasive argument suggesting that

the cases are not parallel. Consider the case where both the bicycle and the car are mine. I have a preference that the bicycle should stay where it is and a stronger preference that it should be moved to enable me to park my car. Note that these preferences both relate to the same case: I am in the same position (i.e. a person who owns both a bike and a car) with respect to both preferences. Consequently the rational procedure is to balance them out and act on the stronger preference. But now compare this with the situation when someone else is the bike owner. Hare says, quite correctly (from H5 - H7) that I must have the preference that, were I the bike owner, the vike should stay where it is. According to the U2-type argument this would mean that I would have to withdraw my statement that the bike ought to be moved, since I cannot prescribe this action for the hypothetical situation when I occupy the other person's shoes. But Hare suggests that I have to balance this preference with my original preference that the bike be moved. It is correct to say that I will have the following preferences:

(1) I prefer that the bike should be moved;

(2) I prefer that, were I in the other person's position, the bike should not be moved.

Hare's position is that since (1) and (2) are both my preferences they can be balanced against one another. But McDermot points out that (1) and (2) are not conflicting in that they are not preferences about the same case. For in (1) I am in my own shoes, whereas in (2) I am in the other person's. Thus when considering my own predicament (2) would not be taken into account at all, since it is only a hypothetical preference. Equally when considering it from his viewpoint (1) will not be included in the calculation either.

McDermot's objection in effect echoes the earlier objection made about the multilateral version of universalisation given in FR. However, to be fair to Hare, it should be pointed out that McDermot may have missed something in Hare's argument. For as he states, universalisation has not entered into the argument at all. It is universalisation that means I must make my hypothetical concern actual. Hare says (MT p.223)

> "Morality requires us to argue; since, if I were going forthwith to have the preferences which he actual has, I must now prescribe that they be satisfied, and since morality admits no relevant difference between "I" and "he" I am bound, unless I become an amoralist to prescribe that they be satisfied. This prescription will have to compete with others, but it is enough to have secured it a place in the competition."

This is suggestive of the following move. Preference (2) "I now prefer that if I were the bike owner it would not be moved" is converted into (2a) "I now prefer that, even though I am not the bike owner, it should

not be moved". If (2a) were a necessary result of critical thinking then McDermot's objection would fail. Hare seems to think that it does follow; the question is, does universalisability entail this, i.e. that I prefer what I would prefer if I did not know what role I played in the situation. As I said earlier, it isn't obvious that it does. Hare says (MT p.108) that universalisability means we must make identical judgments about all cases identical in their universal properties.

Perhaps Hare has something like the following in mind. A moral judgment, being a universal prescription, must fulfill at least two criteria. We must be prepared to act upon it (from prescriptivity) and we must be prepared to do so whatever role we are in (by universalisability). In complex situations no principle is likely to pass such a stringent test. Since most courses of action are against someone's interests, a strict application of U2 would lead to no moral principles passing the test, i.e. to amoralism. Consequently somehow a compromise must be reached. We must issue a prescription which, even if we wouldn't be prepared to issue it from all viewpoints, does at least take all of these viewpoints into equal account. Since we know what we would want done were we in other peoples' positions, we must balance out these preferences and act upon the conclusion.

So runs what I take to be a possible interpretation of Hare's position in MT. But whilst it is certainly a valid way for an individual to go about moral decision making, it is hard to see how it can be claimed that it is the only valid method. I do not see how it is required by universalisability as Hare defines it. Morality may require impartiality, but it is a value judgment in itself to say it requires impartiality of interests or preference satisfaction. This is the step that I think Hare misses in MT. As was argued earlier, it is quite a viable position to argue that it isn't just preference-satisfaction that matters; for some preferences at least what is the object of the preference matters too. Only if the agent thinks that the object of the preference does not matter (or thinks that everyone's choice as to what matters should be respectived equally) does the above compromise work. In effect this is the problem of the fanatic - or as he is better called, idealist - returning again. The idealist will not accept that only preference-satisfaction matters. For instance the man who believes in the sanctity of life - and believes that this view is objectively right - will not accept the above compromise. Suppose he is a doctor, deciding whether to allow a mongoloid baby to die. Hare cannot use the U2 type argument on him. "Would you accept this if you were the baby?" Since U2 can rule out even utilitarian prescriptions. All Hare can do is to tell him that he must be impartial, i.e. it must not be his own interests which influence

his decision. But it isn't his interests but his ideals which determine the doctor's moral decision. He knows that were he the baby he would want to die. He might even admit that the baby's preference to die is stronger than his own preference for it to go on living. But if he thinks that the object of the preference matters as well as its strength Hare cannot say that he is making a selfish or non-moral decision. With his present beliefs - including the one that life shall be prolonged where possible - the doctor can quite sincerely prescribe that, whatever role he finds himself in, the life should be saved.

It follows that I do not regard the argument of MT as an improvement on FR. If anything, the argument in FR was better, since it recognised the last-mentioned problem. Both the arguments in MT and FR are susceptible to the criticisms made in Chapters 2 - 6. Prudence does not require us being moral, the reliance on linguistic intuitions is dubious and the claim that 'ought' is prescriptive in the strong sense required by Hare is false. In this chapter I have argued that the argument of MT V and VII does not show that the critical thinker will be utilitarian, even granting the rest of the argument. In the next two sections we shall investigate to what extent this wrecks Hare's purpose.

Section II

Chapter VI Preference-utilitarianism, "irrational" preferences and happiness

Hare's official position in MT is that he does not need to justify his claim that preferences are morally relevant other than by showing that critical thinking (by arguments H1 - H11) requires this. But if my arguments in Section I are correct, then this view requires amendment. It might well be questioned whether Hare or anyone else would want to defend preference-utilitarianism (PU) if the argument H1 - H11 fails. It is undoubtedly the case that the failure of H1 - H11 would force Hare to modify his claim that PU was sanctioned by our use of language. But, as we shall see in Section III, Hare's distinction between the two levels of moral thinking makes PU a much more defensible doctrine than would otherwise be the case. Moreover PU could be justified in a way which might not be so very alien to Hare's way of thinking. It could be claimed that it was prudent - and consequently rational - to maximise one's one preference-satisfaction and that morality requires the treatment of others as we would like to be treated; hence we must maximise preference-satisfaction for everybody. This last conclusion could be reached either by appeal to universalisability or else by some impartial ideal observer theory. On this revised view Hare would no longer rely in any way on prescriptivity, nor on the dubious argument of MTv and VI, nor indeed on a claim about the sanction of linguistic intuitions. He would, however, have to give a strong argument showing that preference satisfaction was intrinsically desirable. Even if Hare were to try to defend this position in MT he would have to give a much fuller account of what a preference is in order for critical thinking to be workable. In this section my minimum objective is to consider some of the ambiguities relating to PU; at most I hope to show its links with happiness-based theories and reasons for thinking preference-satisfaction is morally relevant.

Before entering into a discussion of PU, it may be in order to say a few words about happiness-utilitarianism (HU) and why PU has been more popular of late. HU, as advanced by Mill so that it does not relate exclusively to physical sensations of pleasure, has a great initial attraction. Conceptually there is a difficulty in saying one does not want happiness: it is arguable that happiness is something that a rational agent must want. But of late HU has fallen out of favour for two very good reasons. One is the difficulty of measuring happiness. HU requires that we add up the total amount of happiness created by an action, yet any value given to happiness necessary for interpersonal comparisons appears arbitrary. More serious still, it is by no means clear what we are supposed to be measuring. No utilitarian has been able to give a

clear account of happiness. Lloyd Thomas (Philosophical Quarterly 1968) argues that HU has seemed plausible only because utilitarians have traded on an ambiguity between two different meanings of happiness. In fact Lloyd Thomas distinguishes four uses of "happy" but only two of these are said to be very significant. The first is roughly equivalent to feeling happy at specific moments and the other involves an appraisal of one's life over a relatively long period. It is this last use which has by far the greatest appeal as a criterion to evaluate consequences by: it is far more plausible to suggest that we should make people have happy lives than to say we should give them feelings of happiness (which seem to be morally trivial). Lloyd-Thomas suggests that a person is happy in the second (ethically relevant) sense if the circumstances of life in which he is actually placed live up to his standards for a good life. Lloyd-Thomas argues that this sense of happiness gives HU its plausibility, but it is happiness in the first sense which makes the diction "pursue happiness" more than empty advice. One is told to live up to one's standards for the good life, but one is not told what standard to set. So HU falls between two stools: either it tells us to maximise feelings of happiness - which is implausible - or else it tells to try to live up to whatever standard we set outselves, which is hardly very useful advice.

Preference utilitarianism (PU) seems attractive because it avoids some of the defects of HU. In particular whereas happiness is a somewhat vague and ghostly concept a preference is directly cashable in action - as Hare says (MT p.104). "We shall in any case do what the balance of our present preferences requires", thereby implying that preferences are a behavioural concept. Whereas it is perhaps an impossible task to be certain which option would make someone most happy in a given situation it is, so the argument goes, much simpler to discover what he prefers most - it is simply what he does. Unfortunately if "preference" is a purely behavioural concept PU lacks some of the attraction HU has. As has been said, happiness seems self-evidently morally relevant; whereas PU somehow lacks this quality. After all, PU amounts to letting people do whatever they want to do, not what is in their interests. It is a commonplace for people to do things against their interests, either through some failing in perception or reason. If a preference is by definition something which disposes someone to act in a certain way then there seems no room for the preference-utilitarian to act in people(s interests where this conflicts with what they happen to want. But the behaviouristic definition is not the only possible one. We could equally say that someone only has a preference to do X (or a preference adjusted for failures in logic and knowledge) if X is what he would consider to be

best if he was reasoning properly and was in full command of the facts. We have therefore distinguished two senses of preference:

(1) PB, one has a preference to do X if one is disposed to X: if one does X then one has an overriding preference to do X.

(2) PI, one has a preference to do X if X is what one would PB if one were in full possession of the facts and reasoning properly.

PI still says that only preference satisfaction matters. Moreover it still says that only the agent's own feelings count: we cannot say that X is in his interests unless he himself would say this, were he in a better position to judge. PI has the substantial advantage over PB in that there is a much stronger argument for it being rational to maximise PI-satisfaction than it is PB-satisfaction. Moreover maximising PI-satisfaction has close affinities with making someone happy - if Lloyd-Thomas's second, ethically significant sense is being used. These advantages are gained at the price of a preference no longer being at all easy to discover. I propose to start off with Hare's apparent position that we should maximise PB-satisfaction, and see what amendments to this the rational agent would make (either in the sense of a rational, self-interested or moral agent).

The most obvious flaw with PB's is that on occasions they reflect incomplete or even mistaken information. At this point, it is helpful to distinguish between intrinsic and instrumental preferences. An intrinsic preference is for something regarded by the agent as an end in itself, whereas a preference is instrumental if it is held only because it leads to something else which one wants. Historically many utilitarians have been reductionists in that they have wanted to say that all preferences are intrinsic preferences - for pleasure, or happiness, or enjoyment. But this view is not very plausible, unless one makes these terms so vague and all-embracing that the claim that all preferences are for them is virtually empty. Preference utilitarians are by no means compelled to take this reductionist view though.

This need not mean that preferences are above criticism. If I understand him correctly, Hare held the fully consistent position that intrinsic preferences can be for whatever the agent chooses them to be, yet all preferences should be amended to withstand exposure to logic and the facts. In the case of instrumental preferences it is easy to see how they can be mistaken. For example, suppose I wish to go from London to Oxford and wish to get on a train purely because of this wish (i.e. my only reason for wanting to go on the train is that I wish to go to Oxford). Suppose, however, that unknown to me this train is the express non-stop to Swansea. Should my preference be counted in a utilitarian calculation? For instance, supposing a friend is standing by the platform, fully aware

of my situation. If he were a PB-utilitarian then he would not try to stop me getting on the train: as revealed by my behaviour, I have an overriding preference to get on this train. Yet if I were fully possessed of the facts I would no longer have this preference: indeed I would have a strong preference not to get on this train. Thus my friend, if he is a PI-utilitarian, will take measures to prevent me boarding the train and this, I take it, would be in line with most people's intuitions regarding the case. Most people would think it was a bad moral theory which said I should stay on the wrong train.

With instrumental preferences criticm is possible because the act might not lead to what we think it will. By definition, though, intrinsic preferences cannot fail such criticism, since we do not value them for anything they produce (if we did they would become instrumental). One position would be to say that instrumental preferences can be criticised but intrinsic ones cannot. But this would lead to rather implausible prescriptions in some cases. For instance, suppose I have an intrinsic preference for all Rembrandt paintings. The preference is intrinsic because I don't want the paintings for their value, or as a status symbol: I just like and admire them. Now suppose I see that a Rembrandt is coming up for auction next week, and I form an intrinsic preference to have this. But let us imagine that this painting is in fact a fake: should my preference now be discounted? (assuming I would not want the painting if I knew it was a fake). There seems no more reason for saying that it should count than in the case of instrumental preferences. The reason why preferences of both types can be criticised is that even intrinsic preferences may be desired for a reason, even if they are not desired because they lead to anything. Thus my preference for the painting is contingent upon its falling under a certain description (i.e. being painted by Rembrandt) though it would be tortuous to suggest that my preference for the painting was in fact an instrumental preference for a Rembrandt. It is accepted that intrinsic preferences can depend on reasons yet still be intrinsic, then clearly they are as much open to criticism as the reason is. To make the types of criticism misinformation can lead to clear, consider the following example:

Suppose someone prefers an Agatha Christie play to Shakespeare. Now a non-utilitarian may hold that on aesthetic grounds this preference should be discounted. A preference-utilitarian cannot discount it and still be regarded as a preference-utilitarian, unless he says that the preference is not what the agent would have if he were fully exposed to logic and the facts. If the preference is instrumental, e.g. the man wants to see that play which will gain him greater regard from other

people, then it can perhaps be shown that the preference is mistaken. But even if it is intrinsic the preference is not beyond criticism. This could be the case, not only if this play were not in fact written by Agatha Christie (which parallels the Rembrandt case). It could also be that the theatre-goer believes that this play has been performed on more occasions than the Shakespeare, and that this is a deciding factor for him preferring it. Were this claim false, then it appears that the reason for holding the intrinsic preference would be false, and hence the preference could be criticised.

Preferences can also be regarded as being open to amendment if they do not take future preferences into full consideration. One of the most common forms of irrational behaviour is to prefer the near to the distant simply because of its proximity in time. It is plausible to suggest that a prudent man will maximise his total preference-satisfaction and will not fall victim to favouring present over future preferences. If this claim is accepted then the man whose present preferences do not adequately reflect future preferences should have his preferences amended. Otherwise the results will be strongly counterintuitive. For example consider the man considering having another drink despite his knowing that it will lead to him having a dreadful hangover in the morning. Assume further that, if he were asked which was better - avoiding the hangover or having this drink - he would reply that the former was preferable. Yet irrationality may lead to him nevertheless having the drink. Let us once more imagine ourselves as an ideal observer. Do we prescribe that the man has another drink? Clearly we would not - yet the PB-utilitarian would have no option but to do so, since the man's behaviour would indicate that this is what he wants. A PI-utilitarian will therefore take into account, not an agent's current set of preferences, but both his now-for-now and now-for-then preferences, with the latter fully representing his later now-for-now preferences.

The first and second phenomena - i.e. misinformation and inadequate planning - often combine. Very frequently we will be uncertain or mistaken about what our future preferences will be. This is an inescapable fact of human existence, and in practice much misery is caused by it. The preference-utilitarian cannot help the situation, in that he is as prone to this uncertainty as much as anyone else (though he can at least point it out and consequently warn against arrangements which rely on preferences being stable e.g. arguably, marriage). But there will be situations where the moral thinker is in a position to know someone's future preferences better than the agent himself. The classic example of this is with children. PB-U would be disastrous with children - partially because

they are guilty of preferring the present, but mainly because they are unaware of what they will want in the future. A PB-utilitarian would have to give the child's preference not to go to school due consideration; it may well be that he would prescribe that they should not go. Yet the PI-utilitarian has no difficulty with this case; the children themselves would (in most cases) prescribe that they go to school were they aware of their future preferences.

There is a third way in which preferences could be amended. It is plausible to suggest that what we ought to aim for, as preference-utilitarians, is to get as many people as possible close to those circumstances they would consider best were they fully informed and planned correctly to take future preferences into account. But even if the two amendments suggested are made there would still be a difference between a person's preference and what they could consider is best. What people do is determined by their desires (in a sense of desire this is analytic). But we have attitudes to our desires, and we do not always regard those desires which are strongest (i.e. those that win out) as those which are most important, in the sense that they are what should determine our behaviour. In fact we can positively disapprove of certain of our desires. A dramatic example of this is kleptomania. The kleptomaniac does steal - hence he must have a preference to steal - yet he would not say that this is what he thought he (prudentially) should be doing: it is not what he desires to desire. This sort of thing does not only happen with abnormal people though. Suppose I am working late at night and am considering whether or not to read another chapter. I might decide that this is what I should do, that this is the rational course of action. Yet my weariness might lead me to say to myself "Never mind the consequences, I am going to sleep". This example shows that this phenomenon tends to merge with the second type, i.e. improper weighting of preferences. Nevertheless the two are distinct. Improper balancing is just one source of irrational behaviour. It may be that there are no future considerations involved.

For instance someone could find themselves doing things which would indicate that they valued money highly (they work long hours and are careful with their money) yet could honestly say that they did not really think money very important. Our attitudes to desires we find ourselves with may in some cases lead to those desires changing, but this need not happen.

People will do what they consider to be best only to the extent that their occurrent desires are in tune with these rational evaluations of value. It could be argued that we should take into account what people really think is best rather than what, through a failure of reason to overcome some drive, they end up in doing. This is less clear cut than

in the previous two criticisms, since it is also arguable that the best way of discovering what people value is to look at their behaviour. In any case there are real problems in determining what someone's evaluation of worth is. Apart from the practical difficulties, there may be no determinate answer to the question, since evaluations vary with moods. Obviously when someone is angry, or tired, or hungry, or irritated, his evaluation may not be very reliable. So it might be thought that we take his rational evaluation to be that when he is in a normal frame of mind. The problem though is that for many people there is no "normal" frame of mine; we can imagine them as typically being in one of several moods and their rational evaluation may vary with the mood. For instance when they are in an artistic frame of mind they might think that appreciating fine paintings is the most worth while pursuit; when hungry they might sincerely think that there is no greater delight than eating good food, when contemplative they might believe that the quest for knowledge is the most important thing for man. Many desires - those for food, sleep, sex - are cyclical. It seems arbitrary to say that one part of the cycle leads to a more normal frame of mind than any other.

These difficulties with the concept of something being considered to be of worth by the agent make a form of PU which insisted that this is what we need to discover problematic.

Yet it is equally unpalatable to take the alternative position, which involves counting the kleptomaniac's preference to steal. Perhaps some compromise can be found whereby we rule out abnormal desires which are out of tune with the normal character of the man. Undoubtedly this too would entail embracing difficulties; it is best left open whether or not this type of irrational preference is amended.

Our revised preference-utilitarianism (PU2) would perhaps be better labelled interest-utilitarianism. Nevertheless it retains the link with PU in that it relates to what people would prefer after exposure to logic and the facts. Moreover it copes with many of the difficulties of PUB. Brandt (A Theory of the Good and the Right, pp.2 and 8) suggests that happiness is a better guide to welfare than preference-satisfaction because we care about securing other people's happiness rather than getting what they want. Indeed it might be suggested that parents, who want to ensure the well-being of their children, see their prime task to be stopping their children doing what they want to do. If preference-satisfaction was an intrinsic good, the objector might continue, this would not be the case. We can now see why this sort of intervention occurs.

Parents are more interested in maximising their children's PI-satisfaction than PB-satisfaction. In general the benevolent do want to see others' preferences satisfied, but only those which they would retain after exposure to logic and the facts.

Perhaps it is enough that the revised preference-utilitarian will maximally benefit people's interests. But we might also try to link PU2 with happiness-utilitarianism. It is fairly obvious that if someone's preferences are satisfied they will at least end to be happy. Moreover if we insist that it is the agent's rational evaluation that counts the two may be closer; it is a commonplace for people to get what they wanted yet for them to be unhappy, simply because they ahve found that what is really important in life is not what they have aimed for.

It might seem that the last version of preference-utilitarianism - i.e. where we take into account what people normally take to be of value - is identical to a happiness-version. For it is very close to what Lloyd-Thomas thinks is the ethically significant sense of being happy, i.e. living up to one's standards for the good life. Clearly one's standards for the good life are conceptually close to one's evaluation of what is of worth. But important differences will exist between PUI and HU. Most obviously, there will be some preferences which are so insignificant that they cannot be said to contribute to long-term happiness. But this does not matter much as these very same preferences will not get a very high weighting on PUI. Equally happiness depends on an anticipation that one's preferences will be satisfied in the future as much as on their actually being satisfied. Finally we would not normally regard someone as being happy unless they evaluated their life as being so. Yet it is conceivable that someone might have all their preferences satisfied yet not be willing to say that he was happy. The concepts of 'happiness' and 'preference''satisfaction' differ in that the former has a subjective element lacking in the latter.

Consequently it would be going too far to claim that our revised form of preference-utilitarianism is identical to happines versions. Nevertheless we have in the end described a form of utilitarianism (PUI) which has several merits. Happiness-utilitarianism fails because we do not know what happiness is. Preference-utilitarianism fails if a preference is define behaviouristically because it is mot implausible to suggest that the rational agent would want to maximise preference satisfaction in this sense. PUI represents a compromise between the two. A preference is still something essentially behaviouristic since it is cashable in terms of action, yet by allowing for the two (or possibly three) types

of criticisms of preference we have made it a more plausible criterion to evaluate consequences. Thus I hope to have partially filled a gap left by Hare in MT, where he fails to enter into the complexities of preference-utilitarianism.

Section III

Chapter VII Moral Conflicts and the Two Levels of Moral Thinking

"Moral Thinking" represents as advance on Hare's previous work in two main ways. As we have already seen he develops his theory concerning the constraints the logic of the moral concepts places on moral argument into utilitarianism. But possibly more fundamental in his claim that there are two levels of moral thinking, the critical and the intuitive. In this chapter I intend to examine Hare's reasons for thinking that there are two levels: in the next, the uses he makes of the distinction.

Hare suggests that the distinction between levels is not entirely original. Indeed he suggests that the seeds of it can be seen in Plato's distinction between knowledge and right opinion. More striking still is the similarity between Hare's system and Mill's: his "secondary principles" play the same role as Hare's intuitive principles. There is no doubt that historically other philosophers have pointed out the need for both a set of moral principles which everyone can have and an intellectual procedure for producing and justifying them. Despite this, talk of two levels of moral thinking has not passed into our philosophical, let alone general vocabulary, so Hare needs to provide arguments to support his view. He does this in MT2, where he argues that a consideration of moral conflicts necessitates the separation of levels.

Moral conflicts are those occasions when we think that we ought to do each of two things, but can in fact only do one. For exmaple I might be able to keep a promise or keep an engagement, but not both. Hare thinks that this type of case supports his claim in two separate ways. Most obviously these cases highlight the need for some decision procedure to determine which principle would be acted upon: critical thinking, Hare argues, does precisely this. Equally importantly, the existence of two levels provides an explanation for an otherwise puzzling feature of some cases of conflict. We want to say that 'ought' implies 'can' yet in some cases we also want to say that we ought to do both fo the acts. Clearly these contradice each other since in these circumstances we cannot do both acts. Hare's theory allows for the possibility that 'ought' is ambiguous between the two levels. If both 'oughts' are retained at the intuitive level, but only one at the critical level, then our reaction to cases of moral conflict is accounted for well.

Bernard Williams ("Ethical Consistency", PASS, 1965) thinks he can use the existence of moral conflicts to argue against ethical theories that suppose one of the 'ought' statements is totally discarded. In many cases Williams's claim would be incorrect; for instance, if I can either save a life or keep a promise I would hardly still think that

I ought to keep the promise if I saved the life. I might feel some non-moral regret at letting someone down, but would hardly have any doubts that I did what I ought (and that had I kept the promise I would have done what I ought not to do). But in some cases where significant moral weight is attached to both alternatives the situation is rather different. Consider Sartre's example of his student, who did not know whether to fight for the Free French or look after his mother who might otherwise perish. In this sort of tragic case the real moral conflict arises. Whichever choice he makes the student will feel serious moral regret. This regret stems from his feeling that he has a duty to both his family and his country. In this case Williams claims that in a sense he ought to do both things is persuasive.

But as Williams recognises this claim leads to severe logical difficulties. For we are saying both that

(1) The student ought to fight for the Free French and

(2) The student ought to stay at home.

But (1) and (2) taken together imply:

(3) The student ought both to fight and stay at home.

Yet he cannot do both, and from 'ought' implies 'can' we obtain:

(4) It is not the case that the student ought both to fight and stay at home.

Williams avoids the hopeless position of accepting both (3) and (4) only by rejecting the 'agglomeration principle' which allows one to pass from "I ought to do A" and "I ought to do B" to "I ought to do A and B". Unfortunately though, not only does Williams provide no argument for rejecting it (other than it is the only way he can see out of the difficulties) but even if it is rejected Williams's position is difficult. It seems that even without the agglomeration principle a contradiction can be reached by the following route.

(1') I ought to fight

(2') I ought to stay at home

(3') If I stay at home I cannot fight

(4') I ought not to stay at home [from (1') and (3')]

Clearly (2') and (4') contradict each other.

If Williams is right concerning our unwillingness to discard either principle then an alternative explanation must be sought. Hare's distinction between the two levels of moral thinking fits the bill admirably. Hare can say that both (1) and (2) above are principles held at the intuitive level. Even if one is overridden it is still held - hence our reluctance to reject it. Our feeling of moral regret is also explained. Our moral teachers ensure that we have dispositions to obey our intuitive principles, so when we break them we do so with the greatest

reluctance. The distinction between levels allows us to say that one of the acts is the right one to do even if we do not in general want to contradict the overridden principle. We can say that at the critical level there is only one thing that we ought to do - but at the intuitive level both principles are valid. So Williams's puzzle is solved without tinkering with 'ought' implying 'can' or the agglomeration principle.

But even if one thinks that Williams's puzzle is illusory, the distinction between levels is still necessary. Suppose one takes the position that moral conflicts are unproblematic in the sense that only one of the 'ought' statements is kept. Suppose it is argued that the moral regret is just an irrational feeling, and that the "tragic" cases, the root cause is indecision about what is right (so it is not the case that we think we ought to do both: we just are not sure which is right). Even then it would be admitted that some decision procedure was essential to determine which principle should be overridden. Principles have to be simple in order to be learnable, so it is hard to see how they can be weighed within the confines of a one-level theory. In addition to the principles we need to compare their importance. This in turn means we must know what reasons we have for following them and how strong these reasons are. For instance to know whether to tell a lie or hurt someone's feelings we must have some idea both of the reasons why both of these are wrong and the moral weight attached to them. It is no use saying that we can appeal to another principle "Tell the truth except where this involves hurting people's feelings" because this would inevitably lead to principles becoming unmanageably long.

Some might doubt whether these considerations call for two levels of moral thinking. They might say that intuitive principles are just rules of thumb and should not be confused with proper moral principles. A utilitarian might well hold this view, so too might an Aristotelian intuitionist who claimed he "just saw" the right thing to do. But the dispute would be largely verbal. All would agree that in general people are guided by simple moral rules. This simplicity entails the possibility of conflict of principles, which in turn leads to the need for some means of determining what principle should override which. In fact the critical level has two roles: not only does it determine which is right, it also decides why it is right. If it is accepted that most of our moral discourse takes place at the intuitive level, it follows that the critical level is necessary to settle inevitable disputes between intuitive principles and to justify this (and justify morality as a whole).

However, one can grant the need for two levels without assenting to the precise form Hare thinks they have. In particular there is no reason why utilitarianism should determine what is right at the critical level, nor that linguistic intuitions are the only things one can appeal to there. With regard to the first point a Rawlsian could argue that appeal to what rational agents would decide to do under a veil of ignorance should determine intuitive principles. Equally a utilitarian could say that utilitarianism does hold at the critical level but that it is not justified by Hare's argument. This is basically what I take Mill's position to be. Whilst of course he did not refer to two levels of moral thinking he certainly thought that whilst utilitarianism in a sense determined what was right, appeals to the ultimate principle should be made not too often. For instance he says:

> "It is a strange notion that the acknowledgment of a first principle is inconsistent with the admission of secondary ones". And
> "We must remember that only in those cases of conflict between secondary principles is it requisite that first principles be appealed to" (Utilitarianism).

Mill's position is remarkably similar to Hare's. Both think that utilitarianism is the ultimate determinant of what is right (and Mill's happiness is closer to preference-utilitarianism than some would think). Both think that we need secondary principles of the form "Do not lie" as appeals to the ultimate principle itself would be counterproductive. Two major differences remain. Hare argues for the existence of two levels independently of utilitarianism existing at the critical level. Conversely in effect Mill argues that secondary principles are necessary if we assume utilitarianism is true. By the principle of utility itself it is better to stick to simple principles than to risk favouring oneself and wasting time by doing a cost-benefit analysis each time one is contemplating action.

The other difference is more fundamental. Hare thinks that the logic of the moral concepts ensures that utilitarianism operates at the critical level. Mill's "proof" was, as is well known, rather different; but in view of his belief that proof of ultimate principles was impossible the value of happiness would in the end have to be an unprovable intuition. I would suggest that in view of the criticism I have made of Hare's method in Section I, a Mill type position is preferable. Utilitarians must give reasons why utilitariansim operates at the critical level; but as we shall see in the next chapter, utilitarianism becomes a much more palatable theory once the distinction between levels has been made.

Chapter VIII Utilitarianism and Justice

Three types of criticism have prevented utilitarianism from being accepted more widely. Least serious is the failure of attempts to prove it. If my argument of Section I is correct then Hare's attempt fails, as much as others. But I say this is the least serious criticism because no competing ethical theory has been proved to the satisfaction of many either, and it may well be that ethical theories are not susceptible to proof in the normal sense of the word. A more penetrating criticism relates to failures by utilitarians to explicate what precisely they mean by happiness, or pleasure, or preference-satisfaction or whatever else they think should be used to evaluate consequences. In Section II, I examined this problem with regard to the satisfaction of preferences, and concluded that if it is to be at all plausible one must revise a simple behaviouristic concept of preference to one which "Corrects" preferences after the agent is made aware of logic and the facts. But traditionally perhaps a third type of criticism of utilitarianism has been most effective. This is to give circumstances where utilitarianism would recommend that we commit some atrocity. The argument is that no viable moral theory would entail committing such actions: hence utilitarianism must be discarded. One of the strongest points of "Moral Thinking" is that by using the distinction between the levels of moral thinking Hare effectively refutes this argument. In this final chapter I intend to examine Hare's defence of utilitarianism, and also look at the attempted reconciliation between utilitarianism and justice.

One example of the type of case devised by anti-utilitarians should show its superficial attraction and its weakness. Suppose three patients in a hospital are on the verge of dying. They suffer from a defective heart, liver and kidney respectively. Imagine that a transplant operation is technically feasible but sadly no organs are available. At that moment a perfectly healthy man walks into the hospital to visit a mildly ill friend. The doctors, if they are committed to act-utilitarianism, are supposed to kill the visitor and use his organs to save the three dying men. Three lives have been gained at the expense of just one. But of course everyone knows that this action would be an awful thing to even contemplate: therefore utilitarianism must be wrong. The critic might even go on to suggest that utilitarianism fails because it does not recognise the separateness of people, all of whom must be treated justly regardless of the consequences.

Hare's answer begins by asking at what level the objection is made. If it is at the critical level then the case can be as fantastic as the critic likes. But if it is made at the critical level then appeals to moral

intuitions are ruled out. These intuitions serve us well in normal cases (if they are good intuitions) but they are the product of critical thinking and cannot be used to decide the outcome of critical thinking. Since utilitarianism operates at the critical level clearly the outcome cannot go against utilitarianism. This may seem to make Hare unassailable by fiat, but his case is strengthened by the fact that if the cases are spelt out in detail it usually becomes clear that the utilitarian prescription is not so appalling. Our repugnance at doing the act is explained by its being wrong in all circumstances we are at alikely to encounter. Our intuitive principles are designed for the world as it is, so cannot be expected to cope with outlandish examples. In these fantasies serious moral thought reveals that the utilitarian act is not so obviously wrong. Consider the example of the three dying men. It is exceedingly doubtful whether utilitarianism would endorse the killing. All sorts of questions need to be asked. Is there no alternative source of organs? (unlikely). Is it certain that the three men will then survive? (again unlikely). And will the quality of their life be the same as the visitor's? (exceedingly improbable). What about the effects on the visitor's friends and the general public's feeling of insecurity? But the critic could say that in this case all these questions could be answered his way. The visitor had no friends, three men were certain to live long and healthy lives, the operation would be kept secret. But if this really is the case (and now the case is so fantastic that we really must forget our preconceived moral principles), would killing the man be the wrong act? The example has been cooked so that only one calculation is relevant: three lives against one. Imagine you had a 25% chance of being any of the four people. Do you prescribe that one lives of three? The anti-utilitarians make their criticism plausible only because they do not fill in the details of the cases; if they do, it becomes clear either that the act is not prescribed by utilitarianism or that it may be the better of two acts.

But the objector has an alternative to resorting to critical thinking. He can confine his objection to the intuitive level in which case he can appeal to his and his audience's moral intuitions. But then he is not allowed to use "cooked" fantastic cases. Hare says he has yet to see one such case. Moreover, the objector would have to provide a case where we do not have conflicting intuitions - for if we did critical thinking would have to be done anyway.

To be consistent Hare has to hold that in normal situations our intuitive principles - that is our feelings about what rights people have, what is just, what is fair, etc., are not incompatible with utilitarianism.

But in fact he advances a stronger claim: namely that these can all be derived from utilitarianism. In this way utilitarianism and deontological theories are in a sense reconciled; the former applies to critical thinking and the latter to intuitive thinking. Hare does not argue for utilitarianism being used to derive these principles separately: if his argument examined in Section I is correct, then he does not need to. But he does think that an appeal to critical thinking would justify many of the rights, rules of justice, etc. that we think we have. Thus consider the question of whether or not courts should punish innocent men. The effect of such actions would in all probability be to undermine confidence in the entire legal system; so utility dictates that there be a principle of justice such that only the guilty are punished. Similarly, it is plausible to argue that an appeal to what people prefer ensures such rights as the right to freedom of speech, and to fair treatment. But critical thinking would not endorse all rules of justice; it would almost certainly rule out some principles of justice in favour of others (e.g. a principle of the type "an eye for an eye" would be rejected). Another consequence of Hare's position is that what rights and rules of justice we ought to have is contingent upon our circumstances - in particular upon human nature. For example, in considering the distribution of money received from the sale of a product, facts, concerning human nature appear relevant. Suppose for the sake of argument that circumstances are such that equality and efficiency must conflict - i.e. if all labour is paid the same wage the total national income is reduced. Now consider two alternative possible worlds. In the first people tend to be envious and there is a high diminishing marginal utility of money. Here it seems a principle like "the benefits ought to be shared equally" is called for. Contrast this with a world where there is no envy and constant marginal utility of money. Clearly in these circumstances the case for differential rates of pay is great. Critical thinking might well suggest a right for equal pay in the first world but not the second. It seems very plausible to suggest that differences in the facts shall affect differences in principles in this way. However, it seems to me that Hare's case for thinking that the distinction between levels provides for a degree of utilitarianism is stronger than its use in reconciling the two types of theory.

Even if it is the case that in practice Hare's theory means people will respect rights and rules of justice this would hardly satisfy the deontologist. He would want to claim that utilitarianism was not even theoretically valid. Moreover it would hardly satisfy the holder of some

principle which critical thinking did not recommend. Suppose that workers are terribly exploited in terms of the differential between the value of their wages and what they produce. Many would argue that their wages should be increased, whatever the consequences. A believer in abstract notions of justice might well hold this view. Yet if the workers are quite happy with their situation, and if an increase in wages would reduce efficiency, utilitarianism would suggest the opposite. At best then Hare's theory only allows for partial compatibility.

There is also a quite different way in which Hare's theory might be said to go against our non-utilitarian principles. It could well be that critical thinking will favour those who are selfish as opposed to those who are moral; which is not only patently unfair, but also will in itself provide an incentive for not doing critical thinking or being moral. Imagine a man can visit either but not both of two sickly relatives in hospital on a particular day. Objectively there is little to choose between them. They both very much would like to see him and he could say the same comforting words. However, one is a strongly assertive, selfish person, who can see no further than his own immediate wishes, the other a considerate type. In view of this, the former character may be represented as having a stronger preference than the latter (who will understand that his younger relative has other obligations), Here utilitarianism requires him visiting the selfish man. Yet this seems to be grossly unfair; why should someone benefit by virtue of their moral defects? This example contains elements of Nozick's utility monster and the argument from evil desires, but is more realistic than either. In general it seems that everyone who adjusts his preferences in accordance with the general good will have weaker selfish preferences and will tend to have his wishes frustrated when another moral thinker applies critical thinking.

Aside from the theoretical difficulties, there may also be practical difficulties in the reconciliation. Hare wants to suggest that both principles of justice and utilitarianism are sound. Whilst this position is quite coherent (the former being general and amendable by the latter in some particular cases) it is difficult to see how it would work in practice. Hare says that omniscient, impartial archangels would never resort to the intuitive level, but humans should rarely depart from it. He thinks that psychologically we will have great difficulty in breaking our intuitive principles if we are well brought up. He also distinguishes between good men and the right action. A good man might not always do the right action, since the intuitive principles are so well-entrenched in him that even in the rare cases where he ought to go against them

(like our doctors case before) he will not be able to bring himself to do so. But this makes his position rather close to Dworkin's view (Taking Rights Seriously) that rights are trumps played by the individual against appeals to the general good. So if intuitive principles are seldom departed from a Harean might not be very utilitarian in practice. On the other hand, Hare also holds that one should do some critical thinking whenever intuitive principles conflict with each other. But in almost all cases where we are faced with a serious moral dilemma our moral intuitions conflict: indeed it is this feature that makes for a dilemma. Yet if we are doing critical thinking in these cases Hare can hardly claim that rules of justice or rights have the sanctity intuitionists think they have. This is no small problem for Hare. By hedging on this issue Hare is trying to have his cake and eat it. Utilitarians and their opponents differ on the issue of how often appeals to utility or the general good should override individuals' rights (if at all). The distinction between levels cannot in itself settle this question. In practice each moral debate concerns whether or not an admittedly good moral principle shall be overridden in some particular set of circumstances. Hare can go either of two ways. He can either say that whenever moral conflicts occur critic1 thinking must take place, in which case he will act like a utilitarian - or else that because of our insufficient knowledge, partiality, etc. we should hardly ever do critical thinking (except to frame principles) - in which case we would act like Dworkinites. Hare cannot please both camps at the same time since there is a substantial measure of disagreement. To illustrate the dilemma consider the following:

Suppose the government has to decide what to do with a village where radioactive material has been released with a massive risk of lethal contamination to the rest of the country. Now it might be that critical thinking would conclude that the village should be obliterated to remove the risk of a much greater disaster occurring. The question is, if the government have read Hare, what will they do? If they stick to the intuitive principles they could not possibly destroy the village and its inhabitants; yet critical thinking might well prescribe this course of action. In practice then it is by no means clear that even a two level moral theory can please both utilitarians and non-utilitarians alike.

So it would seem that the distinction between levels, powerful as it is, cannot do all that Hare intends for it. Its most notable achievement is to formalise Mill's suggestion that by its own application the ultimate principle should not be appealed to directly too frequently. Utility (or preference-satisfaction) itself requires "secondary" or "intuitive"

principles. This allows for a degree of reconciliation between utilitarianism and opposing theories. A two-tier utilitarianism does not deny the need for principles, nor the need for people to treat these principles as sacrosanct. But disagreement remains. Hare gives utilitarianism a theoretical primacy deontologists would agree with. Furthermore, in practice, it is difficult to see how both sets of protagonists can be pleased. But whilst Hare's distinction between levels does not end disagreement between utilitarians and their opponents, it does make utilitarianism a much more plausible theory.

Chapter IX Conclusion

As I remarked at the outset, "Moral Thinking" is a most ambitious work. It contains two important claims. First, that there are two levels of moral thinking and second, that the logic of the moral concepts constrains the moral agent into issuing utilitarian prescriptions. Moreover, these two claims, were they correct, would enhance each other. There being two levels of moral thinking makes utilitarianism all the more plausible and the argument for utilitarianism provides the decision procedure to operate at the critical level.

In Section I, I provided several arguments against Hare's argument for utilitarianism. In view of its defects - and in particular the falseness of prescriptivity and the inefficiency of appeals to linguistic intuitions - I do not see much prospect of this argument being patched up. But in Section III we saw that a two-tier utilitarianism will be more attractive than a one-level theory. In Section II a concept of preference was sketched, such that it is plausible to argue that the prudent agent will seek to maximise his own preference-satisfaction over time.

What is needed is an argument giving reasons why this revised preference-utilitarianism operates at the critical level. As I have said, I do not see how such a proposition could be proved. But perhaps the following is a feasible approach: It could be suggested that the rational, self-interested agent will atempt to maximise his own preference-satisfaction. The step from prudence to morality requires that we impartially seek to maximise everyone's preference-satisfaction.

If this argument is to work, at least three major problems have to be faced. First, much fuller argument has to be given as to why it is rational for the self-interested man to maximise his own preference-satisfaction. Secondly, conclusive arguments must be made against the idealist who thinks that the object as well as the strength of preferences matters. Finally, much has to be said of how the two-tier utilitarianism will work in practice. I suggest that all these questions merit investigation.

Bibliography

Brandt, R.B. (1979): A Theory of the Good and the Right (OUP)

Hare, R.M. (1952): The Language of Morals (OUP)

(1965): Freedom and Reason (OUP)

(1981): Moral Thinking: Its Levels, Method and Point (OUP)

Lloyd-Thomas, D. (1968): Happiness; Philosophical Quarterly, 1968

McDermott, M. (1983): Hare's Argument for Utilitarianism; Philosophical Quarterly, No. 133, 1983

Mackie, J.L. (1977): Ethics: Inventing Right and Wrong (Penguin)

Mill, J.S. (1861): Utilitarianism

Parfit, D. (1971): Personal Identity: Philosophical Review, 80

Phillips Griffiths (1983): The Triviality of the Debate over the 'Is/Ought' and the Definition of 'Moral': Americal Philosophical Quarterly, 1973

Williams, B.: Ethical Consistency (reprinted in Problem of the Self)

Finally, thanks are due to my supervisor, David Lloyd-Thomas, for his helpful criticism and suggestions on the first draft of this work.