A COMMUNITY OF

(IMPERFECT)

BENEVOLENT ARCHANGELS.

A philosophical approach to moral education

and

an educational approach to moral philosophy.

Jeffrey William Wardle

Thesis for PhD

Institute of Education, University of London

Preface

This thesis is about moral philosophy, moral education, and the relationship which one has to the other. I argue for a particular moral philosophy and derive from that a view of moral education. But I also argue that the relationship between the two is of a special nature and differs from the relationship which might exist between philosophy and education in general or between, say, the philosophy of mathematics and education in mathematics.

The moral theory I offer incorporates a view of moral thinking which is, in many respects, similar to that given by Hare. However, the thesis includes an extended criticism of Hare's form of utilitarianism and, especially, of his rationalist justification for the form of moral thinking which he recommends. The criticism of Hare's theory, and of his approach, forms the background against which I recommend a fundamental modification of utilitarian moral theory. Although the theory offered yields a utilitarian view of right action, it is a non-consequentialist theory which is based upon a notion of an ideal agent. The theory is founded upon a notion of the benevolent archangel as universal ideal.

The moral theory is offered as a perspective upon those moral views which we share. That perspective is recommended as one which can elucidate, underpin and inspire those moral views. The form of moral education which is derived from that theory focusses centrally upon the development of the virtues of benevolence, non-malevolence, understanding and humility.

CONTENTS

INTRODUCTION.

Philosophy of education and aims for education.

Peters (1972 p.vii) states that philosophy of education is to be conceived of "as drawing on established branches of philosophy and bringing them together in ways which are relevant to educational issues". When tackling such issues the philosopher of education will "draw on and develop" work done by philosophers in epistemology, ontology, metaphysics, ethics, aesthetics, philosophy of mind, philosophy of mathematics, and so on. In some cases the relevant philosophical work will not have been done and philosophers of education may then have to undertake the philosophical work themselves before proceeding to draw on and develop the results of that work in an educational context.

For example, Peters (1972) says, philosophical work on 'rights', 'punishment', and 'authority' may be drawn on and developed when tackling issues in education to do with the rights of parents and children, punishment in schools, and the authority of the teacher; and philosophers of education may themselves work on concepts of 'education', 'teaching', 'learning', and 'indoctrination' in order (I take it) to draw on that work when tackling issues to do with aims and methods.

The emphasis here is upon analysis of concepts and is in tune with the analytic approach which at that time pervaded philosophy itself. Those who are disenchanted with conceptual analysis (especially in an educational context) may point out that 'analysis' of a concept such as 'punishment' might help to avoid confusion in a discussion of, say, the role of punishment in schools but

it cannot begin to settle substantive questions about the desirability, legitimacy, and effectiveness of the use of punishment. So too, analysis, clarification, revision, or stipulation of the use of such terms as 'education' and 'indoctrination' may permit us to justify a remark such as "That isn't education, it's indoctrination", but it does not take us one bit closer to settling whether the activity referred to is to be condemned or applauded.

Furthermore, some may feel that philosophy of education need not merely consist of the application, to issues in education, of knowledge and understanding derived from established branches of philosophy. Some may wish to cast off such constraints in order, say, to elaborate and advocate a view of the broad aims of education - ie. to develop **'a philosophy of education'**. Such an enterprise may require the special skills of a philosopher (whatever they may be) and may, in part, draw upon knowledge and understanding derived from established branches of philosophy; but those who engage in it may feel that it is unnecessary (and, indeed, not possible) to found their philosophy of education upon specific philosophical theories - the link to established branches of philosophy may be partial and piecemeal.

Recognition of the sterility of much "conceptual jousting" - a phrase used by J.White (1982) - and a reluctance to be constrained by a 'narrower' (and less ambitious) view of the philosophy of education, may lead the philosopher of education to wish to get on with the "main business" of justifying, prioritising, comparing, and discovering the relations between, possible general aims for education. Surely, J.White (1982 p.x) says, one would expect "that general discussions of educational aims would be just what it [philosophy of education] would engage in". He goes on to remark that some may object that if one leaves off analysis and begins to put forward views as to what aims **ought** to be then one is no

longer doing philosophy. But, he declares, he is not too concerned with whether the advocation of aims is referred to as 'philosophy', 'casuistry', 'moralising', or even 'mush'; it remains true to say that discussion of what the aims of education ought to be is an important task which ought not to be neglected.

It is easy to agree with this last point - discussion of what the aims of education ought to be is extremely important. Indeed we may well feel that it is sufficiently important for us to wish to encourage all those involved in education to engage in it. But we may also feel that those involved in planning and pursuing **any** large-scale activity ought to spend some time considering the central aims and purposes which that activity might be designed to achieve. If there is disagreement over, or conflict between, or difficulty in achieving, those aims then we may also feel that it is desirable that there be some attempt to elaborate and advocate alternative sets of aims and priorities. But is there any reason for supposing that this task should especially fall to the philosopher, or for supposing that this is just what one would (or should) expect the philosopher working in education (or any other large-scale activity) to engage in?

The philosopher's approach to the consideration of broad aims for education may involve an attempt to subsume diverse aims under some more general idea; to ferret out and remove contradictions and incompatibilities; to achieve clarity, simplicity and coherence; and (most importantly) to provide convincing reasons for adopting the proposed system of aims and for rejecting alternatives. Such an approach is not confined to philosophers. The psychologist, historian, or sociologist may well adopt a similar approach when considering broad aims for education - the difference is perhaps likely to lie in the nature of the rationale

which is offered in defence of the proposed aims. But, leaving that aside, we may ask whether an approach from within the philosophy of education, **which does not draw heavily on knowledge and understanding derived from a range of disciplines other than philosophy**, can achieve the goal of advocating and elaborating one particular set of aims. Is it possible for the philosopher of education (as such) to achieve that goal? Furthermore, is the achievement of that goal necessary and would it (thus conceived) be sufficient to the task of settling rational debate over aims?

Achieving clarity and greater coherence is certainly possible for any set of aims for education (or for any educational doctrine); and, as Passmore (1980 p.9) points out, many philosophers of education, wishing to avoid "mere preaching" or "amateurish psychology or sociology", have seen clarification as their primary task. But to simply clarify existing educational doctrines or, alternatively, to expose educational 'theories' as pseudo-theories is, according to Passmore, a humble task; and there is "something more than a little unsatisfactory in this conception of the philosopher of education as an odd-job gardener" whose business is to tidy up the careless work of the educational theorist.

It is, of course, possible for rational and informed debate to go much further than clarification. For example, beginning with a favoured wider aim (such as a stable democracy, a successful economy, a contented population, a maximisation of artistic and scientific achievement) it may be possible to determine the aims for education which are most likely to lead to the achievement of that wider aim (or aims). But would an approach which did not draw heavily on knowledge and understanding derived from a range of disciplines other than philosophy be sufficient to **that** task?

Perhaps some philosophers of education would anyway reject this alternative goal and would seek to advocate a set of aims for education without reference to any wider aims. Yet it is clear that the educational aims which we pursue will have a significant impact upon the nature of our society, the nature of individuals within that society, and the lives which those individuals lead. Education may be seen, by some, as an end in itself; but it is also (at least in part) a means to other ends. Thus it is likely that the aims we adopt for education (and, perhaps, for some other large-scale activities) will reflect the views we have about the desirability of this or that form of life, or form of society. The philosopher may see this as precisely the point at which it is possible to make the most significant contribution: through discovering rationales for ways of life, or forms of society, and then deriving implications for educational aims.

For example, J.White seeks to advocate a particular view of the good of the individual, and of the relationship between that good and the good of society (where the latter centrally concerns the moral obligations which members of that society have to one another). The well-being of the individual, he says (J.White 1982 p.58 and p.95), involves an awareness of the enormous range of human desires, of the permanence of one's 'natural' desires, and of "the need to hold all of one's desires together in an integrated unity"; it involves the passage from such awareness towards the construction of an informed 'life-plan' and, thus, towards "an integrated system of hierarchically organised preferences"; and it involves having those capacities and dispositions which allow one to effectively pursue such a life-plan. Then, given that we accept that the good of society requires that individuals have some concern for each other, he argues that the requirement of 'psychical unity' (through the possession of a life-plan and system of preferences)

entails that the good of the individual further requires that the needs and interests of others be accomodated **within** that life-plan and system.

I would argue that we face ultimate choices and commitments here. I do not believe that it is possible to discover rationales for ways of life or forms of society unless those 'rationales' are founded upon such choices. J.White (1982 p.129) agrees that "justifications can't go on forever and that somewhere one reaches bedrock commitments" - for example, that one should attend to the well-being of others. In a later work (J.White 1990) he makes a similar point when he argues that the value of personal autonomy is relative to the type of society in which we find ourselves. But, I believe, we reach bedrock much sooner than he would allow.

J.White (1982 p.50 and p.58) claims that each individual has permanent natural desires ("to be loved, to be secure, etc.") and other wants fostered by institutions and culture; and that he "has to" learn to cope with the conflict between such desires by integrating them "within a single scheme". Each of us, he argues (1990 p.31), has to organise our desires, "to impose a hierarchical structure on them and resolve conflicts between them". It is this commitment to thinking through one's desires and arriving at an integrated scheme which is central to his view. It is true that he also stresses the importance of unreflective pursuit of enthusiasms; and of a disposition to act upon one's desires and to approach one's projects with enthusiasm; but it is the notion of an autonomous and integrated (albeit evolving) life-plan which, in its elaboration, gives rise to a particular view of the broad aims of education.

Do we have to regard some of our desires as an inevitable and permanent feature of our lives? And, if so, then

which desires? Must we really include, for example, the desire to be loved? Do we have to integrate all of our desires within a single scheme? Is that the only (rational) response to the inevitable conflict which results from the diversity of those desires? What constitutes a 'single scheme'? We may have many schemes and projects: to be a good teacher, contented, a caring parent, a researcher in the philosophy of education, a responsive friend, and so on. Do we have to prioritise and integrate them within a single scheme?

Faced with conflict between such schemes and concerns, some may devise priorities, schedules, and timetables; but others may simply 'muddle along' - responding to opportunity, external pressure, and changing (and unforeseen) circumstance. Are the latter less rational; must they resolve these conflicts by means of an integrated life-plan? It is here, I believe, that we reach bedrock in J.White's justification for a particular view of the well-being of the individual: in a particular response to the difficulties inherent in leading a life (and in a particular view as to which desires are natural, inevitable and permanent).

However, there is no doubt that, if it were possible to discover a rationale for a particular view of the well-being of the individual (or of society) then that accomplishment **would** provide **'a philosophy of education'**. Our broad aims for education would simply be: to provide an education which contributed towards the achievement of that way of life (or form of society). The nature of those broad aims could be indicated as the features of that way of life were elaborated. If, for example, the well-being of the individual requires the construction of an integrated system of hierarchically organised preferences then appropriate broad aims for education would include the development of an ability to create such a construction.

It may then be possible to further elaborate the view of education beyond that which is straightforwardly implied by the offered view of the 'good life' for the individual, in order to present a more detailed view of the education which would contribute most towards the achievement of such a life for each individual. But, to return to the central point of this introduction, would an approach which did not draw heavily on knowledge and understanding derived from a range of disciplines other than philosophy be sufficient to **that** task? If education is a means to 'the good life' or 'the good society (or to any other end) then the **elaboration** of broad aims for education has to draw upon knowledge and understanding derived from a wide range of disciplines - the philosopher of education may well contribute to that task but cannot expect to have a central, or crucial, role.

The different task of **advocating** one particular set of broad aims for education could, as we have seen, be pursued by means of the search for rationales for a favoured view of the well-being of the individual. But I do not believe that such a search can result in a justification for only one such view. We reach differing bedrock commitments at an early stage in the justification process.

Having claimed that **this route**, to a rationale for one particular set of broad aims for education, will not succeed; we may go on to question whether that **goal** is, in fact, necessary. J.White (1982 p.3) argues that unless those who are involved in education can come to a reasoned conclusion as to which broad aims are acceptable, then cohesion between the different parts of the educational system is endangered. The work of primary schools, secondary schools, colleges, universities, teacher-training institutions, and staff within each institution must "mesh together"; and that is

to be achieved by means of the shared aims which (it is hoped) would result from rational discussion. Thus: shared aims are necessary, that requires agreement about aims, and that requires a shared rationale.

But, firstly, does the work of all the institutions listed (and of parents and others involved in education) have to mesh together in this way? Must the primary school, university, parent, and polytechnic share the same aims? There are many aspects to human development and I see no fundamental reason why different institutions and groups should not focus on different aspects of that development; or have very differnt priorities within a range of broad aims; or have aims which are, to some extent, conflicting.

Furthermore, such institutions may set out to meet different needs or to meet the same needs in distinctive ways. Individual secondary schools, for example, may deliberately seek to have or to emphasise different, albeit overlapping, aims. Far from seeking to mesh together in the sense of having identical aims, ethos and priorities, they may seek to mesh together by means of a collaboration aimed at offering a range of distinct alternatives to the local community (see Jenkins 1991 Ch.9).

Secondly, even if shared aims were necessary would that require agreement? It is perhaps possible for institutions to work to coherent, shared aims and yet tolerate a large measure of disagreement over those aims; or, perhaps more likely, without there being any great measure of explicit agreement. However, even though agreement may be overrated, it may be desirable. If so then we can ask, thirdly: do shared aims and agreement require a shared rationale?

J.White (1982 p.2) lists a few of the aims which are "currently at large in the world of education": to promote the growth of understanding (or knowledge, or reason, or the mind) for its own sake; to help each pupil to develop his potentialities to the full; to enhance personal autonomy; to promote all-round development (in a balance between intellectual and practical achievements or between the arts and sciences); to promote excellence within specialisms; to ensure a literate and numerate work-force; to ensure an intelligent participatory democracy; to foster art and culture; to develop (moral) character. As he says, the list of aims is almost endless. Furthermore, we might add, some immediately conflict and some will conflict when we begin to interpret and pursue them in detail.

We are thus likely to be faced by conflict between the aims which are favoured by different individuals and, indeed, by conflict between the aims which each of us, as a single individual, would wish to favour. Yet as individuals we may, nevertheless, persist in regarding all (or some subset) of these aims as desirable and respond to our dilemma simply by choosing to foster that educational system which we believe will maximise achievement of all of those aims.

Likewise a group of individuals who cannot agree upon aims, **can** agree to compromise and to choose an educational system which will allow a measure of achievement for each of the favoured sets of aims. Such courses of action do not require a rationale which provides an over-riding aim and a resulting system of priorities. Shared rationales are not the only route to agreed systems and objectives. Furthermore, unless we are so optimistic as to believe that rational enquiry can lead to one set of broad aims which all rational people involved in education **must** pursue, then such compromise is inevitable.

Consensus decisions reached through such compromise need not be seen as mediocre decisions which necessarily affect the quality of the outcome. As Caldwell and Spinks (1988 p.197) point out, the quality of the outcome will depend upon: a) how acceptable the decision is to those who must implement it or who will be affected by it; and b) how effective the chosen systems and objectives are in achieving the aims.

For example, in order to reach a view as to how we can maximise achievement of a set (or several sets) of aims which involve conflict, we will have to determine the impact which the achievement of each aim would have upon the achievement of each of the others. This process may result, say, in our assigning low priority to some aims in order to enhance overall achievement; and will thus result in a system of priorities. That system of priorities will be effective only if the nature of the conflict between the aims has been correctly understood. This type of route to consensus will require knowledge and skills derived from a wide range of specialisms.

Furthermore, whatever our route to a system of priorities (whether through a rationale involving an over-riding principle, or through compromise, or through steadfastly pursuing personal preference, ...), the final choice of aims will require consideration of 'practical' factors. We will have to take into account, for example, the time, resources and skills which we can realistically hope to make available; the nature and extent of the alterations to our society which may be required in order to achieve any great measure of success for our chosen priorities; the lessons which can be learned from other times and places as to the likely consequences of adopting this or that set of aims; and so on.

Those who reach a settled system of priorities through consensus and compromise may well incorporate consideration of such factors into the process by which they achieve a consensus. However, those who seek a rationale for one particular set of broad aims for education are likely to see such factors as involving conditions which need to be altered or sustained in order to achieve those aims; and will thus see consideration of such factors as external to the debate about aims. In the latter case, the broad aims may be seen as somehow 'intact' even though they may not, in fact, be pursued as stated. Our stated aim is, say, to help each pupil to achieve independence and to develop their potential to the full; but our aim, in practice, is to help pupils to develop independence in some respects and to develop some of their potential.

This is not a trivial point. If, in practice, we cannot hope to achieve a satisfactory measure of success for each of our favoured set of aims then it may sometimes be sensible not merely to trim our sails but rather to change tack somewhat - ie. adopt different priorities from amongst the range of aims which we find acceptable.

Consideration of such factors, and the ability to make informed judgments as to whether aims need to be altered (despite our rationale), will once again require knowledge and skills derived from a wide range of specialisms. The task of **settling** aims (not merely the task of elaborating those aims, nor merely the task of discovering how aims are to be realised) is extremely complex. To see that task as the special province of the philosopher of education, or as just what one should expect philosophy of education to engage in, is, I believe, to obscure the fact that considered and informed discussion of the issues involved will require skills and knowledge derived from many different specialisms.

A denial that the elaboration and advocacy of aims and priorities in education is just what the philosophy of education should engage in, need not be based on the view that such activities amount to casuistry, moralising, or mush. We do not have to choose between allowing the title 'philosophy of education' or accepting such pejorative terms as 'mush'. Nor do we need to rely on a false contrast between 'analysis' and 'going beyond analysis' to advocate and elaborate broad aims. The true contrast is between a philosophy of education which attempts such an advocacy, and one which restricts itself to attempts to draw on and develop work done in established branches of philosophy in ways which are relevant to educational issues (and the latter need not, of course, restrict itself to 'analysis').

I have tried to argue that the former task, **of advocating and elaborating broad aims**, cannot be achieved without drawing upon knowledge and understanding derived from a wide range of disciplines. If that task is tackled purely from a philosopher's perspective then, I believe, it will at best articulate a particular set of favoured commitments.

The latter approach, **of drawing on and developing work done in established branches of philosophy**, certainly need not confine itself to conceptual analysis but it may be that it cannot easily escape the same objection: perhaps little can be achieved without drawing upon knowledge and understanding derived from a wide range of disciplines.

It is the latter approach which I shall be following in this thesis. In the first chapter I shall, therefore, discuss whether that approach can escape the same objection. In order to highlight some of the issues which are involved in answering that question I shall consider, in some detail, the relationship between the

philosophy of mathematics and the aims of mathematical education. The issues raised will then be used in order to begin a consideration of the topic with which this thesis is concerned: the relationship between moral philosophy and moral education.

CHAPTER 1.

Philosophy of X and educational aims for X.

The relevance of established branches of philosophy.
The philosophy of mathematics.
Aims, objectives and methodology of mathematics teachers.
Moral philosophy and moral education.

The relevance of established branches of philosophy.

Most would now deny that there are logical links between broad theories of education and broad philosophical theories. For example, Passmore (1980 ch.1) rejects the notion that to each philosophical school there corresponds a philosophy of education. The conclusions which traditional philosophers of education sought to sustain cannot, he says, be derived from epistemological, ontological, or metaphysical premisses. "That is exactly why Feigl and Russell divorced their educational from their philosophical writings. There is no possible passage from logical atomism to Russell's radical educational innovations, from logical empiricism to Feigl's defence of liberal education.".

If this is so then is there any relationship between philosophical theory and educational theory? And if philosophical theory **is not** directly relevant to the task of establishing broad educational doctrines or aims (but psychological, sociological and other factors **are** directly relevant to that task) then do philosophers of education have no choice but to articulate a particular set of favoured commitments (whilst making use of the work and views of psychologists, sociologists and others)?

Philosophy of X and educational aims for X.

An alternative is for the philosopher of education to engage in a different task. Broad doctrines and aims are not the only issue in education. Even if broad educational doctrines cannot be deduced from philosophical theories, and even if it is neither possible nor necessary to 'settle' broad educational aims through the use of philosophical argument, this does not of course mean that work done in established branches of philosophy has no bearing upon controversies in education. Perhaps no-one would deny that philosophy may be helpful in such a context. But are there controversies in education which can be settled by drawing on and developing work done in established branches of philosophy and without equally detailed reference to work done in other disciplines?

Passmore (1980 p.12-15) claims that there are many ways in which the philosopher, qua philosopher, can make, and has made, a direct contribution:

> Problems in social, political, and moral philosophy are, for example, directly relevant to such controversies as those over the selection of students, government of schools, grading, and the relation between schooling and employment.
>
> Many controversies in education involve philosophical concepts which call for close analysis - there has been too little "discussion of such questions as the circumstances in which we can properly say of a child, for example, that he has been 'well-trained'; that he has learnt to 'appreciate' literature; that he acquired the ability to 'think for himself'; that a particular form of teaching will develop his understanding or imagination".

> Epistemology may not form the basis for a general doctrine of education but it may well be directly relevant to pedagogical issues - for example, "to reject the view that all knowledge is based on sense-impressions is to deny that, insofar as it aims at the imparting of knowledge, teaching must proceed by giving children sense-impressions".
>
> Also, Passmore says, such branches of philosophy as the philosophy of mathematics, the philosophy of science, and so on, will bear "directly upon controversies about the processes of teaching". "Every 'philosophy of' aids the teacher to see more clearly what he is doing when he teaches the subject that 'philosophy of' is about".

The first two of these examples illustrate the way in which philosophy may usefully contribute to controversies which arise from discussions about different forms of education. They would serve as examples of the way in which those tackling educational issues might draw on work done in philosophy **as well as** work done in other disciplines. The role of philosophy may be seen as similar to that required in debates over broad aims for education - ie. as one of many contributions. There may be no implied claim that the work of the philosopher may (of itself) determine the form of education which we will (or ought to) engage in, or that it may (of itself) settle controversies about specific aspects of education.

However, the last two examples perhaps imply a more direct and crucial role for the philosopher. Just what, for example, does the claim that the philosophy of mathematics, or science, will 'bear directly' upon controversies about the processes of teaching amount to? A philosophy of mathematics may well be a contributing factor in rational decision making about objectives and methods, but is it ever a determining factor?

Philosophy of X and educational aims for X.

Some are willing to claim that a philosophy of mathematics is, or ought to be, a determining factor in settling issues about the teaching of mathematics. For example, Clark (1987) asks the question: "Why **ought** a teacher to avoid pointing around the environment if she holds that mathematics is analytic; and, conversely, why is a teacher who proceeds in this way **committed** to an empiricist view of the subject, even if she has never heard of such a view?". The answer which he gives rests on the claim that to do otherwise is to be "afflicted with a kind of incoherence".

Clark claims that when we teach a subject we ought to direct the attention of the learner to the "place where the propositions are verified". Thus, if we believe that the truth of mathematical propositions is determined simply by the meaning of the terms used in those propositions then, as teachers of mathematics, we ought not to proceed by directing the learner's attention around the environment. Furthermore, Clark claims, 'genuine' teaching requires that the teacher **has** a view as to where the propositions of the subject are verified (the attention of the learner must be deliberately directed to the place of verification **as such**); thus teachers cannot avoid incoherence simply by declining to be committed to any particular philosophy of mathematics.

I.Scheffler (1973 p.34-40) also claims that a philosophy of a given subject is necessary to the teaching of that subject. A philosophy of - an analysis and understanding of the form of thought embodied by a subject - is necessary for several reasons. For example:

> facilitating the acquisition of habits and methods appropriate to a given subject requires the ability to analyse and articulate them, and to understand their point;

> the teacher must be prepared to justify the perpetuation or alteration of those habits and methods, and that requires the ability to criticise and evaluate them.

Thus the teacher, Scheffler claims, needs to have not only a facility in such methods but also the ability to analyse and evaluate them. Such analysis and evaluation is precisely what a philosophy of a given subject involves.

Such views perhaps go further than the claims made by Passmore (1980 p.13): for example "the philosopher of science, by giving students a better grasp of the connection between science and commonsense, can help to prevent the teaching of science from becoming a kind of magic", and "the philosopher of history can make it plain just how history teaching must differ, in its criteria of success, from the teaching of the social sciences".

Here Passmore may merely be asserting that it can be helpful to be clear about the nature of the subject taught. But how clear? Is it really necessary for teachers to take sides in the kind of controversies which occupy philosophers when they wrestle with problems in the philosophy of mathematics, science, or history? Must teachers decide between different 'philosophies of'? Is there a direct and obvious relationship between a particular philosophy of, say, mathematics and a particular set of aims, objectives and methods for teaching mathematics?

The Philosophy of Mathematics.

Korner (1960 p.9-10) points out that the "apparent contrast between the indefinite flux of sense-impressions and the precise and timeless truths of mathematics has been among the earliest perplexities and problems not of

the philosophy of mathematics only, but of philosophy in general". Philosophers have asked: Why is it that mathematical propositions appear to be necessarily, self-evidently or indubitably true? Are they true in this peculiar way because they are asserted about objects of some special type, or because they are asserted about objects in general or 'as such', or because of their not being asserted of any objects at all? Is their truth due to the particular method by which they are reached or are verifiable - for example, an immediate and incorrigible act of intuition or of understanding?

Korner goes on to consider three schools within the philosophy of mathematics: the logicist, the formalist, and the intuitionist. For the purposes of this discussion it will be convenient to concentrate our attention upon just one of these 'schools'.

The logicist believes that (pure) mathematics deals exclusively with concepts definable in terms of a very small number of fundamental logical concepts, and that all its propositions are deducible from a very small number of logical principles (by means of a small number of methods of inference). In attempting to derive mathematics from logic the logicist makes use of truth-functional tautologies (eg. p or not p), postulates from the logic of classes (eg. a U b = b U a), and postulates relating to the use of the terms 'all' and 'some' (eg. in a universe of discourse consisting of a finite number of objects, say $a_1, a_2, .. a_n$, '(x)f(x)' is equivalent to '$f(a_1)$ and $f(a_2)$ .. and $f(a_n)$'). As Korner (1960 p.50) says, "every logicist system draws its list of postulates and rules of inference from the logic of truth-functions, the extended logic of classes and the logic of quantification".

The first point to note is the claim that "the list of postulates and the list of inference rules are not

independent .. [eg.] a suitably large list of postulates enable one to economize in inference rules". Thus we can have different systems which aim to derive mathematical propositions from different sets of logical postulates. The second point (which Korner notes earlier) is that problems with the logic of classes seem to require that some of the postulates are not 'logical' principles at all. For example, if **any** class is admitted then the logic of classes leads to contradictions. One such contradiction Russell (1919 p.136) describes: "Form now the assemblage of all classes which are not members of themselves. This is a class: is it a member of itself or not? If it is, it is one of those classes that are not members of themselves, ie. it is not a member of itself. If it is not, it is one of those classes that are not members of themselves, ie. it is a member of itself.". Either way we have a contradiction. Russell's response is to adopt various rules for stratifying classes into types in order to avoid the possibility, within the logic of classes, of a class containing itself as a member. Others have adopted similar rules and would not claim that such rules are themselves **logical** principles.

Further difficulties arise when the logic of quantification is extended to universes of discourse consisting of an infinite number of objects. With regard to a finite universe consisting of objects $a_1..a_n$ the proposition '$f(a_1)$ and $f(a_2)$ .. and $f(a_n)$' can be written '$(x)f(x)$' and then since, in our logic of truth-functions, the proposition '$(f(a_1)$ and $f(a_2)$ .. and $f(a_n))$ -> $f(a_i)$' is a tautology it follows that the proposition '$(x)f(x)$ -> $f(a_i)$' is also a tautology within that finite universe. But if we wish to regard the proposition '$(x)f(x)$ -> $f(a_i)$' as valid for an infinite universe then we cannot do so for the same reasons; rather we must adopt this as a postulate or adopt such postulates as will ensure that this proposition will be deducible as a theorem. Furthermore, those propositions

which involve quantification over infinite ranges can be interpreted in many different ways: for example, as mere technical devices, or as inadmissible, or as involving an empirical assumption about the world. Once again we may be involved in making postulates (that there is an infinite class of individuals in the universe) which we would not wish to claim are logical principles.

The responses which logicists make to these difficulties will differ. Furthermore, it is clear - as Korner (1960 p.34) points out - that if the logicist is to derive particular theorems of arithmetic from an initial set of logical propositions then he will need to change symbols en route: "somewhere in the path leading from the premisses to, say, '1+1=2', the transition from obviously logical symbols to symbols not obviously logical must occur". The transition will, therefore, be mediated by definitions and the account which logicists give of those definitions may be very different. The definition of number, for instance, may be regarded (with Russell) as a mere device of notation which declares that a newly introduced combination of symbols is to mean the same as another combination of symbols whose meaning is already known, or it may be regarded (with Frege) as an attempt to demarcate a class of objects whose members exist as independent entities. These two accounts of definition lie at the heart of two very different branches of logicism: the nominalistic, in which the propositions of mathematics are seen as not being 'assertions' at all; and the realistic, in which the propositions of mathematics are regarded as making assertions about 'logical' objects.

The philosopher of mathematics who wishes to adopt a logicist position must decide, amongst other things, in what way to avoid the antimonies generated by the logic of classes, how to interpret propositions which quantify over infinite ranges, and what account to give of the

definitions which permit the transition from logical principles to mathematical propositions.

Does the teacher of mathematics also have to make such decisions? Perhaps those who insist upon the importance, to the mathematics teacher, of the philosophy of mathematics would be satisfied if the teacher were to decide between broad schools of mathematical philosophy and would not require the teacher to decide between the various different philosophical theories which have arisen within each 'school'. Perhaps it is sufficient to decide between, say, logicism, formalism, intuitionism, or empiricism. As a logicist, for example, I may have at least decided upon what I believe to be the method of verification for mathematical propositions (derivation from logical principles). But a brief look at the development of the 'logicist school' reveals that even this may not be as clear as we might suppose.

Some logicists have held not only that mathematical propositions can all be derived from a small number of logical principles but also that such principles share some fundamental feature which clearly demarcates them as 'logical' - for example, that they are known a priori, or are true by definition, or are indubitable, or are non-empirical. Other logicists have doubted that it is possible to make clear a distinction between, say, empirical and non-empirical propositions and have therefore held only that mathematical propositions can be derived from a small number of principles (no attempt being made to characterise those principles in some special way).

If we adopt the latter position, and if we believe that mathematical propositions are verified by deriving them from some specific set of principles, then the question arises: 'How are those principles themselves verified?'. For example, in classical logic

'(p or q) and r -> (p and r) or (q and r)'
is a truth-functional tautology. But is there a method of verification for this principle? If so, then is that 'method' such that the logical principle could turn out to be false?

Putnam (1975 p.174) asks the question "could some of the 'necessary truths' of logic ever turn out to be false **for empirical reasons**?" and he goes on to argue that "the answer to this question is in the affirmative and that logic is, in a certain sense, a natural science". Putnam claims, for example, that anomalies which have arisen within quantum mechanics can be resolved if the distributive laws, which form part of classical logic, are given up. Alternative methods of resolving such anomalies involve (roughly speaking) the claim that the process of measuring (say, the energy level of an atom) affects the value obtained by measurement: "there is a mysterious 'disturbance by the measurement'". But, says Putnam (1975 p.183), there are two problems with the latter resolution: firstly, no **theory** of this disturbance is offered; and, secondly, "if a procedure distorts the very thing it seeks to measure, it is peculiar that it should be accepted as a good measurement, and fantastic that a relatively simple theory should predict the **disturbed** values when it can say nothing about the **undisturbed** values". Therefore the resolution which involves our abandoning the distributive laws of classical logic is, according to Putnam, to be preferred.

Putnam concludes that logic is empirical; it is, in a certain sense, a natural science. As soon as we recognise that alternative logics might have serious physical application, then the a prioricity of logic vanishes.

However, one who accepted Putnam's views as to the nature of logic might remain, in a sense, a logicist. It would

still be possible to maintain that the business of the mathematician is to derive logical consequences from a limited set of logical principles; and to believe that mathematics as a whole can be derived from a particular set of such principles. Some 'logicists' (those who believe that all mathematical propositions can be derived from a small number of logical principles) may then be, in a sense, empiricists.

Does the teacher of mathematics have to decide whether to espouse or reject this particular strand of logicism? If he espouses it then does he, when teaching mathematics, 'direct the learner's attention around the environment' or refrain from so doing? If the claim that to each philosophy of mathematics there must correspond a particular methodology for teaching mathematics were correct then the task of determining **what** methodology for teaching was appropriate to a particular philosophy of mathematics would be far from straightforward.

We might go on to ask, at this point, not only whether teachers of mathematics need to adopt a philosophy of mathematics but also whether mathematicians need to do the same. Is the mathematician who uses 'empirical' methods committed to a certain philosophy of his subject?

Putnam points out that mathematicians have often used what he calls 'quasi-empirical' methods and have felt that their belief in certain theorems has been justified as a result of those methods. Such methods involve arriving at a hypothesis by means of intuitively plausible though not certain analogies, checking the results of the hypothesis to see if any counter-examples are generated, and then demonstrating that the hypothesis has important consequences for mathematics and science. (Putnam offers several examples - 1975 p.64-69.)

Putnam points to these examples as a means of support for his claim that the methods of proof and of quasi-empirical inference are complementary - and for the further claim that much of mathematics is 'empirical'. Putnam (1975 p.76) accepts that proof (rigorous deduction from axioms) is the primary method of mathematical verification but he insists that quasi-empirical methods can also verify mathematical theorems - we often **know** such theorems to be true before we succeed in finding a proof.

The non-empiricist may insist, on the contrary, that such methods are not methods of verification - they may suggest hypotheses but we do not **know** those hypotheses to be true until a proof is found, proof is the only method of mathematical verification.

The point I wish to make is that the mathematician who makes use of such methods is not committed to either view, nor is it necessary that he should make such a commitment. If the theorem has application and if counter-examples are not found then the mathematician may adopt it and work with it (eg. determine its implications and apply it to the 'solution' of specific problems) and leave aside the question of whether it has been verified. Of course, if a mathematician aims to 'verify' theorems then he **will** have to commit himself to a view as to the methods of verification appropriate to mathematics, but that may not be his aim. Furthermore, even if a mathematician believed that the ultimate aim of mathematical activity is to verify theorems, and that such verification requires, say, derivation from logical principles, he may nevertheless choose not to devote himself to that task and may employ quasi-empirical (and other) methods in the belief that the propositions he arrives at may be useful and significant, or that his work will prepare the ground for others with more rigorous and less speculative leanings.

As a mathematician one will need a firm view as to methods of verification only if one is firmly in the business of 'verifying' and claiming verification. The problem of the nature of verification for mathematical propositions is one of many difficult problems in the philosophy of mathematics but a decision as to which solution to adopt (to this and other such problems) is not a necessary condition of engaging in mathematics.

Nevertheless, some (eg. I.Scheffler 1973 p.35) might insist that it is a necessary condition of **teaching** mathematics. The mathematician may arrive at true (and useful) mathematical propositions by quasi-empirical methods (or manipulation of objects, or guesswork) but if it is the case that **the** method of verification is, for example, derivation from logical principles then that is an important fact about mathematics. Ought not the teacher, who believes this to be the case, to teach accordingly and avoid encouraging pupils to use quasi-empirical methods (or manipulation of objects, or guesswork)? And if that is so, if a belief about methods of verification determines the appropriate approach to teaching, then does not the teacher have to decide what his belief is?

If the teacher's aim is (or has to be) to teach pupils how mathematical propositions are verified, or to give pupils skill in verifying mathematical propositions, then the answer to both of these questions clearly has to be in the affirmative. But there are other aims which a mathematics teacher might have.

Aims, objectives and methodology of mathematics teachers.

The Dainton Committee (1968) listed the following reasons for the study of mathematics:

Philosophy of X and educational aims for X.

as a means of communicating quantifiable ideas;
as a training for discipline of thought and for logical reasoning;
as a tool in activities arising from the developing needs of engineering, technology, science, organisation, economics, sociology, etc.;
as a study in itself.

Just as one might accept the legitimacy of each and every one of the aims for education in general which J.White pointed out as being "currently at large", so too one might accept that each of the aims above are legitimate aims for the teacher of mathematics. One might also feel that each of them is not only legitimate but also has value. But, even if this were the case, one might ask whether they have equal priority and whether, given limited time, they are in practice compatible.

The decisions we make about priorities may depend upon a very wide range of factors. Some of these may be to do with the nature of our aims for education in general but others may relate to our views, for example, as to whether particular aims are in fact achievable, or as to what objectives are most appropriate to children of a particular age or level of achievement, or as to the intrinsic difficulties involved in achieving objectives appropriate to this or that aim, or as to ways in which the achievement of objectives appropriate to one aim help (or hinder) our achieving objectives appropriate to another aim, and so on.

The aims we adopt, and the priorities we decide upon, will then influence our choice of methodology. For example, the methods appropriate to mathematics teaching which primarily aims to train pupils for general discipline of thought may be very different from those which are appropriate if we place a high priority on the

ability of our pupils to use mathematics as a tool in activities arising from the developing needs of engineering, technology, science, organisation, economics, sociology, etc. Furthermore, both of **these** aims might be seen as making use of the study of mathematics as a means to something else. Thus if either (or both) of these aims is prioritised then there need not be any connection between methodology of teaching and our espoused philosophy of mathematics (in particular the method of verification which we believe to be appropriate to mathematics).

Even if our chosen aim **exclusively** involves teaching mathematics 'for itself' (and we believe, say, that mathematical verification involves rigorous derivation from logical principles) we may still feel that a methodology of teaching which involves practical activity, investigative work, and the use and application of mathematics, would be more likely to awaken the pupil's interest. So that, in the early stages at least, our choice of methods may once again have little or no connection with our philosophy of mathematics. Encouraging pupils to use mathematics in concrete situations, or to investigate and speculate about mathematical relationships between features of their environment, may be the best introduction to a study of mathematics 'itself'. If mathematics 'itself' is primarily a matter of verifying mathematical propositions (and I am not sure that it is) then pupils may be much more likely to take an interest in it if they learn that the verified propositions can be useful and illuminating.

Not only is our choice of methodology likely to be influenced by a very wide range of factors but the process of using those methods is a learning process and the discoveries we make about the success of this or that methodology may sometimes influence the choice of aims. The teaching of mathematics used to involve (especially

in the teaching of geometry) a rigorously deductive approach. In part this was, perhaps, because of beliefs about the nature of mathematics but it was also a result of the belief (shared by the Dainton Committee) that such methods would train pupils for discipline of thought and for logical reasoning. This latter belief has often been challenged; for example, the Director's Report (1962-3) for the SMP O-level mathematics course claimed that "for the majority of pupils, formal geometry offers little training in logical reasoning and emphasises, instead, practice in the memorising of theorems and proofs of no particular worth".

Most would now agree in rejecting the idea that exposure to rigorous mathematical reasoning is a means to achieving general discipline of thought or to encouraging pupils to be more logical in their thinking. This aim, as an aim for mathematics teaching, is now rarely mentioned - the methods employed did not seem to achieve the stated aim and few believed that the methods could be altered in a way which would improve success in that aim.

If we look at the aims, objectives, and programmes of study for mathematics in the 1989 orders for the National Curriculum, we find no mention of mathematics teaching as a means of training for discipline of thought. We also find a shift away from rigour and towards a much greater emphasis upon investigative and speculative work; the use and application of mathematical concepts, knowledge, understanding and techniques; and practical activities relevant to the pupil's interests.

Most teachers would agree (for a range of reasons) that it is desirable to place a high priority on endeavouring to ensure that school mathematics is interesting and enjoyable, and there is now a large measure of agreement as to the objectives and methods which will excite interest and cause enjoyment in school pupils ('relevant'

skills and topics, a practical and intuitive rather than a rigorous approach). There also seems to be a consensus in the wider society that the pupil's ability to use and apply mathematics ought to have a much greater priority. Such factors as these have (and ought to have) influenced the teacher's choice of objectives and methods, and they have done so independently of any views as to how the propositions of mathematics are verified.

The process of deciding (or coming to agreement) upon aims, objectives, and methods for mathematics teaching is, of course, one which ought to involve a good deal of reflection and debate. In the course of that process problems will arise the solution of which may require skill in philosophy, or knowledge and understanding derived from established branches of philosophy. But I do not believe that philosophical theories about the nature of mathematics can determine the methods or aims we ought to adopt for mathematics teaching, any more than I believe that broad philosophical theories (about the nature of knowledge or reality) can determine our broad aims for education.

Just as informed debate about the broad aims of education requires skills and knowledge derived from a range of specialisms; so too, I believe, consideration of aims, objectives and methodology for mathematics teaching will need to draw upon an equally wide range of skills. The relationship between a philosophy of mathematics and decisions about mathematics teaching is far from straightforward. Furthermore, the degree of relevance which the former has to the latter will depend upon the broad aims we adopt for mathematics teaching and for education in general.

The same point could be made with respect to the philosophy of any discipline or area of the curriculum. We cannot simply assume that proposed solutions to the

problems which interest a philosopher of this or that discipline will be relevant to issues relating to the teaching of that discipline: that they will (as Passmore claims, 1980 p.13) "bear directly upon controversies about the processes of teaching".

The philosopher who wishes to investigate the relationship between education and a philosophy of a particular discipline or area of the curriculum must, I believe, outline the conditions under which proposed solutions to philosophical problems **would be** relevant to the choice of aims, objectives, content, or methodology.

The area of the curriculum which I wish to investigate at length is that of moral education. The aim of this thesis will be to consider some proposed solutions to certain problems in moral philosophy, to criticise and modify those proposed solutions, and to attempt to determine the relationship between those proposed solutions and issues in moral education. I shall, therefore, conclude this chapter by briefly discussing the conditions under which certain aspects of moral philosophy would be relevant to issues in moral education. Later chapters will, hopefully, make clear in detail how the proposed solutions to problems in moral philosophy are related to those issues (given those conditions).

Moral philosophy and moral education.

Central issues in the philosophy of mathematics relate to difficulties in establishing the nature of the truth conditions, and methods of verification, for mathematical propositions. In moral philosophy we have the added difficulty of establishing the nature of the judgments made: are they statements, prescriptions, or expressions of feeling? Those who characterise moral judgments as

statements will then have to go on to explain the nature of the truth conditions, and methods of verification, for those statements. Those who claim that moral judgments are prescriptions may go on to offer descriptions of methods by which those prescriptions are arrived at, or may be arrived at, or must be arrived at if they are to be 'moral' judgments. Those who claim that moral judgments are expressions of feeling may go on to describe what they believe to be the psychological or sociological causes of those expressions of feeling which are characterised as 'moral'.

Even though such issues are central to moral philosophy, we cannot assume that the solutions which are proposed to these (and other) problems in moral philosophy will be straightforwardly relevant to issues in moral education. For example, let us suppose that the only aim of a particular group engaged in providing moral education is to ensure that the actions of the educatees accord with the moral judgments of the educators. (And let us suppose that an argument to the effect that an activity with such an aim could not be referred to as moral 'education' would not be particularly illuminating.) If this were the aim then proposed solutions to the problems outlined above would not be straightforwardly relevant to problems of teaching.

Given such an aim, moral education would be a matter of ensuring conformity to the expectations and demands of the educators. Such an authoritarian moral education would require, as Dearden (1968 p.170) says, that the pupil acquires the 'virtues' of unquestioning obedience, conscientious compliance and deference; and that the motivation to obedience be sustained by the 'impressiveness' of authority. The choice of detailed aims, objectives, and methods would then largely be guided by an understanding of the psychological and sociological factors involved in maintaining the

impressiveness of authority and a disposition to obedience. It may well be the case that philosophical problems will be raised in this context - for example, how do certain forms of constraint and punishment accord with our notion of justice, how can we determine when a pupil is 'disposed' to comply with a given moral principle - but proposed solutions to the question of the existence and nature of methods by which moral judgments are to be arrived at would not (given this aim) be relevant.

However, if the judgments of the educators were not, themselves, such as would be arrived at by a method appropriate to moral judgment then, we could claim, they are not 'moral' judgments at all. Thus we might insist that the aim of moral education cannot merely be to ensure that the actions of the educatees accord with the judgments of the educators. Rather the aim must be (at least) to ensure that the actions of the educatees accord with those moral judgments **which would be arrived at by methods appropriate to such judgments.**

An aim of this sort clearly presupposes that there are such methods and therefore raises just those problems in moral philosophy which we have been discussing. But it is important to note that although proposed solutions to those problems may thus determine the nature of the judgments to be transmitted (that is, the 'content' of moral education), they would not necessarily determine the detailed objectives and methods appropriate to that education. The aim as stated is still compatible with the educators having the role of merely transmitting, and ensuring behaviour in accordance with, those moral judgments which they (or others skilled in such methods) make. In order to achieve this aim, the educators must, of course, now ensure that the judgments made are such as would be arrived at by such methods, but the educatees need not be party to this process and may therefore be

encouraged to accept such judgments on 'authority'. The choice of detailed objectives and methods would still be guided by an understanding of the factors involved in ensuring obedience, deference, and behaviour according to the judgments made (or passed on) by the educators.

It may well be the case that those who believe that there **are** methods appropriate to the making of those moral judgments which relate to action would not be satisfied with an educational aim which made no reference to the ability of the educatees to themselves use such methods. Such dissatisfaction may be due to a commitment to the wider educational aim of achieving autonomy for the educatees (in all areas and perhaps because such autonomy is seen as desirable in itself) or it may be due to a conviction that the inscrutability of authority is especially dangerous in the context of morality. As Dearden says (1968 p.171) if obedience can be relied upon then the demands of authority "can safely be extended to cover unfair privileges" and "the temptation to .. an abuse of trained gullibility must be very great").

The point which Dearden, and others, make is not, however, one which need necessarily lead us to modify the aim described above. If the aim **is achieved** then the actions of the educatees will accord with those moral judgments which would be arrived at by methods appropriate to such judgments. The 'authority' appealed to by that aim is not the authority of a particular group and the judgments which they happen to make, but rather is that of the methods and the judgments which result from their use.

But we may, nevertheless, feel that educatees should be able to make their own moral judgments, and thus scrutinise and see the merit of the judgments to which they are being asked to conform. We may have various reasons for believing this. For example:

In the absence of such ability it may well be the case that, even if the educators are able to influence the behaviour of the educatees, that influence will be temporary and will also not extend to behaviour which is not witnessed by the educators.

The constraints upon the educators (and others entrusted with the task of making moral judgments by means of appropriate methods) may not be sufficient to ensure that the stated aim is achieved - ie. those who are entrusted with responsibility for ensuring that the authority of the methods is maintained are likely to abuse the authority which that role gives them.

Just as aesthetic development is not merely a matter of learning what are the 'correct' judgments, or even of making choices based on such judgments, but is fundamentally about appreciating beauty; so too moral development involves the ability to make moral judgments oneself, to appreciate their significance, to understand what is right or good and not simply to accept it on authority.

To educate people in a way which fails to enable them to make their own moral judgments is morally wrong.

These examples illustrate the very different types of reason which may lead us to modify our aims for moral education. The first would need to refer to the particular nature of the society in which the educators and educatees found themselves. The elaboration of the second would require an understanding of human psychology. The third of these reasons would need to be justified by philosophical argument. The last appeals directly to a moral judgment.

The examples illustrate the way in which discussions of aims may appeal to a wide range of insights. As in the case of mathematical education, informed criticism of our aims for moral education may require us to draw upon knowledge and understanding derived from a range of specialisms. But the last of the examples points to a very different possibility. In the case of moral education, our moral judgments may themselves influence our choice of aims.

We might elaborate the last reason in the following way: to aim merely to mould the behaviour of educatees is to ignore their capacity for moral judgment and is to ignore the fact that their behaviour could stem from such judgment; the pursuit of such an aim fails to treat educatees as morally responsible agents; such an education would treat them as means to our ends (albeit worthy ends) and would, therefore, be morally wrong. The possibility of this type of consideration means that our moral judgments may directly influence our choice not only of the 'content' of moral education but also of its aims, objectives and methodology.

I shall return to this point in a later chapter but, for the moment, we can thus envisage a range of considerations which may lead us to favour a moral education which has at least two aims:

a) to impart an ability and willingness to arrive at (by appropriate methods) moral judgments;

b) to encourage an inclination to act in accordance with moral judgments arrived at (by appropriate methods) by self or others.

If such aims were adopted then problems as to the nature and existence of methods appropriate to moral judgment would become straightforwardly relevant to the process of

moral education. The solutions which are proposed to these central problems in moral philosophy would bear directly upon controversies about detailed objectives and methods of teaching - just as solutions to problems in mathematical philosophy would "bear directly" if our aims included that of imparting the ability to verify certain types of mathematical propositions.

Furthermore, the particular nature of the solutions proposed to these philosophical problems may provide additional reason for adopting those aims in preference to the aims of a more authoritarian moral education. Equally, however, such solutions may also lead us in the **opposite** direction - **towards** authoritarian aims. We can see how either of these may be the case if we consider a moral theory such as that given by Hare (1981).

Hare's theory yields a characterisation of moral thinking which involves two elements: ascertaining the consequences of alternative actions in a given situation and imaginatively identifying with the preferences of all those involved in that situation. According to Hare, action in accordance with such judgment would maximise preference satisfaction in each situation. Thus Hare's theory yields a form of utilitarianism. The form of thinking which leads to utilitarian judgment also, it is claimed, results in an inclination to act accordingly. Thus the achievement of aim a) would bring with it achievement of aim b). This last claim may well lead us to adopt aim a) as the primary focus for moral education.

However, as we shall see in later chapters, Hare's elaboration of that theory leads to a view of moral education in which aim a) and (a form of) aim b) have equal importance. That elaboration takes account of the fact that we are seldom able to engage in the form of thinking which Hare describes - that is, to determine which action would maximise preference satisfaction in a

particular situation. Thus, Hare claims, moral education should involve an aim to impart a disposition to act according to general principles which would ensure maximisation of preference satisfaction over a range of situations. Hence a form of aim b) is necessary. Aim a) will also be necessary for if we had such dispositions but lacked an ability to engage in such thinking we would not, for example, be able to determine how we ought to act in those situations in which the general principles conflict. Thus moral education will focus upon both aims.

But we may now be only a small step away from excluding any focus upon aim a) for all but a 'gifted' few (those destined to be in authority). Moral thinking of the type described is, as Hare says, very difficult. We may believe that (all or some group of) our educatees are not capable of acquiring sufficient skill in ascertaining consequences and imaginatively identifying with the preferences of others. We may also believe that they are not capable of avoiding a tendency to give undue weight to their own preferences (or to the preferences of those to whom they are 'close'). We may, therefore, decide that **their** moral education should **not** include an aim along the lines of a). The nature of the moral theory, and our knowledge of human abilities and weaknesses, may lead to the abandonment of aim a).

For Hare it is the maximisation of preference satisfaction which matters. If (for the reasons given) the pursuit of an aim along the lines of a) would result in a decrease in overall preference satisfaction then that aim should be avoided. A capacity and disposition to engage in moral thinking has (for Hare) no intrinsic moral worth. The fact that the theory involves an ideal outcome of action (the maximisation of preference satisfaction) does not directly yield particular aims for moral education. There may be different ways of

achieving those outcomes and each may involve very different aims for moral education.

Thus far we have only considered moral worth and moral education in relation to action and outcome. Some moral theories incorporate such a focus. In such theories the nature of agent is not centre stage; the agent has moral worth only insofar as the nature of that agent is such that it ensures morally worthy actions and morally good outcomes. But not all moral theories have such a focus.

If our moral theory were to focus primarily upon the moral worth of the agent rather than upon the value of the outcomes of action, if it were to centre upon a notion of an ideal **agent** rather than upon an ideal **outcome**, then the link to aims for moral education may turn out to be much more direct. If, for example, our moral theory yielded the view that a form of thinking similar to that outlined by Hare had **intrinsic** moral worth then it would entail that all **ought to have** the capacity and disposition to engage in such thinking. A moral theory of this type, which incorporates a notion of an ideal agent, will thus directly involve aims for moral education - the statement of such an ideal **is** a statement of an aim for moral education.

I claimed, in the last section, that the degree of relevance which our philosophy of mathematics has to decisions about mathematical education will depend upon the broad aims we adopt for mathematical education and for education in general. Here we see that the nature and extent of the relevance which moral philosophy has to decisions about moral education will depend upon the particular nature of the moral judgments we make and of the moral theory we espouse.

In this thesis I will consider and criticise Hare's moral theory. I will offer an elaboration and modification of

Hare's theory and in so doing I will propose a theory which incorporates a shift of focus from outcomes to agent. The nature of the resulting theory will be such that it will directly involve aims for moral education.

I will begin with Hare's theory not only because the form of thinking which it describes is very similar to that which I will recommend as being an essential feature of an agent having moral worth, but also because criticism of that theory will allow me to make clear the contrast between Hare's 'rationalist' approach and my own approach.

Hare claims that the form of moral thinking which he describes can be used to underpin and systematise those moral views which most of us share. His rationalist approach involves the further claim that a fully rational agent must, when attempting to form his own moral judgments, employ such a form of thinking. I hope to support the former claim, but I will argue against the latter.

I will argue that a moral theory which involves such a form of thinking is just **one way** in which we can provide a perspective upon those moral views which we share. If we are to decide which such perspective to adopt then we will need to adopt criteria of selection which cannot be derived from a consideration of the nature of rationality. I will claim that our choice of such a moral perspective/theory/philosophy may ultimately (and legitimately) be determined by a consideration of issues in education. In particular: 'How does that perspective enhance the ability of the educator to develop those moral views in self and others?' and 'How does that perspective relate to our experience of educational development?'.

Philosophy of X and educational aims for X.

I will thus be claiming that the relationship between our response to issues in moral philosophy and our response to issues in moral education may not only be very direct but may also be a relationship of interdependence. On the one hand, the moral theory we espouse may directly entail aims for moral education; on the other hand, our consideration of aspects of educational development may be a crucial factor in the selection of that moral theory.

CHAPTER 2.

Moral judgment and an inclination to act.

Hare's characterisation of critical thinking.
Critical thinking and moral education.
Hare's route to Utilitarianism.

Hare's characterisation of critical thinking.

Hare (1981 p.20) says that he hopes "that by investigating the meanings of moral words we shall manage to generate logical canons which will govern our moral thinking". The two characteristics, which Hare claims are features of the meanings of moral words and which are central to this project, are prescriptivity and universalisability.

Hare hopes to show that these features entail a method of moral thinking which all rational agents, as rational, must adopt. This method is that of 'critical thinking'.

Very roughly, the argument which generates Hare's characterisation of critical thinking is, I believe, as follows:

a. a fully rational agent only makes a prescription for action in a given situation if he knows what are the consequences of the various possible actions in that situation;
b. the prescription which results from such knowledge is rational only if it depends upon what are the preferences of that agent with regard to those consequences;
c. some moral judgments are prescriptive;

d. prescriptive moral judgments are 'universalisable' - that is, the same prescription must also be made for all situations (actual and possible) which are identical save with regard to the numerical identity of those involved;
thus, the same prescription must be made for each of those **hypothetical** situations in which the agent **is** each of the persons involved;

e. a fully rational agent only makes a prescription which is universalised in this way if he knows what are the consequences of the various possible actions in those situations;

f. the prescription which results from such knowledge is rational only if it depends upon what are the preferences of that agent for each of those situations;

g. in order that the prescription may so depend, the agent must acquire knowledge of what his preferences would be if he were each of the persons involved;

h. such knowledge requires an imaginative identification with those persons such that the agent actually acquires their preferences;

i. the resulting prescription (if rational) will thus depend upon what are the preferences of each of those involved with regard to the consequences of the various possible actions.

Steps e. to i. yield the description of critical thinking. Thus: I acquire the preferences of all those involved in a particular situation (by identifying imaginatively with each), I consider the consequences of alternative actions in that situation, and I reach a decision in the light of all the preferences which I now have (after imaginative identification).

Steps a. to d. make claims with regard to rationality, prescriptivity, and universalisability, which are intended to justify the claim that moral thinking

logically requires critical thinking. The argument, as it stands, involves claims about the logic of moral words (made in c. and d.), but it also makes claims about the rationality of choice (made in a. and b., repeated in e. and f.) and about necessary conditions for knowledge of the preferences of others (h.). The argument generates a characterisation of the method of arriving at moral judgments.

When rational agents are engaged in an attempt to reach a moral judgment, this method would require each rational agent to consider the same facts in the same way. Hence, insofar as such agents have knowledge of the relevant facts, they will all reach the same judgment.

But not only will full employment of the method of critical thinking yield agreement as to what is the morally right thing to do in a certain situation; it will also result in a preference to act in the appropriate way. The prescription which is the result of moral/critical thinking is an expression of the preference which the rational agent has after imaginatively identifying with the preferences of all those involved, and acquiring knowledge of the consequences of alternative actions, in a given situation.

Thus if people understand the meanings of the words they use, if they are rational, and if they have the necessary knowledge when making moral judgments then they will not only agree in their moral judgments but will also (as a result of reason alone) prefer to act accordingly. Such a theory of moral thinking would, if correct, have important implications for central issues in a moral education involving aims of the sort outlined in the last chapter.

Moral judgment and an inclination to act.

Critical thinking and moral education.

If Hare is correct, then if **those who make moral judgments** could be made more rational and could be given the ability to acquire knowledge of the relevant facts, then that of itself could ensure not only agreement in judgment but also right action. Rationality and the ability to acquire knowledge may, in a sense, be sufficient to virtue.

On this view, reason is not the mere "slave of the passions" (Hume), rather it is the case that the exercise of reason can "determine the will" (Kant). There are important caveats to be made here, if we are to do justice to Hare's theory, but it remains true to say that, for Hare, reason can lead to right action without there having to be present a 'good disposition' which merely employs reason as its tool.

There are other moral philosophies in which there is a direct link between the means of arriving at a moral judgment and the disposition to act. For example, moral realists such as Platts (1979 p.261) may claim that the "distinctive feature of clear moral perception is that it gives us a **compelling** reason to act". But what the realists seem not to do is to show how we should deal with someone who fails to see the 'moral facts'. How do we achieve clear moral perception and how do we help others to achieve it? Platts (1979 p.247) says that "We detect moral aspects [of a situation] in the same way we detect .. other aspects: by looking and seeing". But then how do we account for, and educate, those who are able to detect just those other aspects which the realist detects but are (mysteriously) unable to detect the 'moral aspects' which the realist claims to detect - ie. do not share the realist's moral views and do not receive a compelling reason to act in the way which the moral realist favours?

Moral judgment and an inclination to act.

The realist is likely to say that that person is simply not 'sensitive' to the moral features of a situation. That much follows from the theory: perception of moral features implies that one has a compelling reason to act, therefore if one lacks such a reason to act then one has not perceived the moral features. But the theory offers no means of distinguishing those who can perceive moral features from those who cannot, other than their sharing the realist's views and having an 'appropriate' disposition to act. From within an educational context we will wish to be offered some further elucidation of this skill. Without such an elucidation we have no clue as to the type of education which is likely to produce an improvement (in skill and thus, according to the theory, in behaviour). In practice, such a theory is likely to give rise to the view either that we can do nothing, or that we can (at best) ensure that our educatees behave in a way which conforms to the judgments of those who are blessed with the required skill. In either case this would mean that we were unable to pursue the aims involved in the type of moral education which we are considering.

Hare's claim that moral judgment results in a disposition to act follows from the detailed description which he offers of the process of moral judgment. That description not only provides the link to action but also clarifies the essential features of a moral education which aims to impart an ability to make such judgments. The educatee must acquire the ability to:

determine the facts in a specific situation,
ascertain the range of alternative actions,
establish the consequences of those possible actions,
determine the preferences of those involved,
imaginatively identify with those preferences,
relate consequences to the preferences thus acquired.

Moral judgment and an inclination to act.

The exercise of such abilities would result in a preference which was based upon the facts in a given situation and the logic of our moral language.

In short, the broad outlines of a moral 'education' (of the sort we are considering) are spelt out in the theory. If those engaged in moral education aim to develop the ability of the educatees to themselves arrive at moral judgments, and aim to encourage an inclination to act according to such judgments, then this philosophy of morality has clear implications for the nature of the detailed objectives involved.

Like Kant, Hare not only claims a form of 'objectivity' for moral judgments but also offers a description of a method of arriving at those judgments. That method is one which even those who oppose the claim of objectivity (as I shall) can understand and utilise. It also involves skills which we all have and which we can, given appropriate education, improve. There is no appeal to a form of perception or intuition which, as well as being a mystery to the opponents of realism or intuitionism, is such that the proponents of those theories can give little guidance as to how we might educate those who are deficient in it.

Hare's view does not, of course, entail that the method of critical thinking need be the only route to right action. A central feature of Hare's overall position is his 'two-level' theory whereby the importance of general principles and appropriate dispositions is constantly emphasised. When our action is the result of a disposition to be guided by this or that general principle then we are operating at the intuitive level (the level of the 'prole'); when our action is the result of deliberations of the sort outlined above then we are operating at the level of critical thinking (the level of the 'archangel').

Moral judgment and an inclination to act.

Unfortunately we are seldom able to operate at the level of critical thinking; we have limited knowledge of consequences and limited skill in determining the preferences of others. Reason alone may determine choice and may not, therefore, need the 'guidance' of good dispositions but such dispositions will nevertheless be important in those situations where we lack the ability or opportunity to fully exercise reason - and that is nearly always.

There are thus, according to Hare (1981 p.45), two ways in which each of us can achieve virtue in our actions: as proles we can act on the basis of our good dispositions; as archangels we can act as a result of critical thinking. If someone had all the characteristics of the archangel (was capable of perfect critical thinking) then "everything would be done by reason in a moment of time" and that person would not "need the sound general principles, the good dispositions, the intuitions which guide the rest of us". But in fact "we all share the characteristics of both [archangel and prole] to limited and varying degrees and at different times."

Both archangel and prole are ideals. The archangel clearly so - he can, when confronted with a novel situation, instantly fulfil all the requirements of critical thinking. But the prole is also an ideal - he has **good** dispositions, **sound** principles. As moral agents we will need (if Hare is correct) to strive after both ideals - to have the qualities of the archangel so that we can exercise them when we have the opportunity, and to have the qualities of the prole so that we can cope with more pressing situations.

The outlines of a moral education (of the type we are considering) are, thus, further spelt out by the theory. As moral educators we may see our role as falling into

Moral judgment and an inclination to act.

two parts. First, to instil dispositions to act according to general principles which have resulted from the limited critical thinking which we, as a community, have been able to undertake; and, second, to give our pupils the ability to acquire knowledge and exercise reason in order that they might think critically themselves.

As Hare (1952 p.76) says there are, according to his view, two aspects to moral education. Firstly, a child must be provided with "a solid basis of principles"; and, secondly, it is necessary to provide ample opportunity to engage in the decision-making process by which such principles "are modified, improved, adapted to changed circumstances, or even abandoned". The first is necessary because we do not, generally speaking, have the time or the skill to engage in full decision-making. The second is necessary because a body of principles will not meet all circumstances in a complex and changing world, and because those principles may sometimes yield conflicting judgments; so that if a child is to achieve autonomy in such situations, then that child will need to acquire skill in the decision-making process.

The view which Hare puts forward is one in which these two aspects of moral education are both essential to the development of virtue; but it is the second which is seen to be the ultimate guarantor. For without such skill the individual could not arrive at a correct moral judgment for those situations which are not dealt with, or are dealt with in conflicting ways, by the body of principles; and the community as a whole could not build up such a body of principles.

Furthermore, as has been said, central to this view is the claim that the skills of critical thinking may be, in a sense, sufficient to virtue. **When** critical thinking is employed there is no need for 'good' dispositions. The

Moral judgment and an inclination to act.

canons of thinking which are generated when we understand the meaning of moral words, and which will govern our thought insofar as we are rational, would be sufficient to yield a preference to act in the right way.

If Hare were correct, then those involved in education would be right in thinking that the development of those capacities required by critical thinking would, when employed to make moral judgments, result in improved behaviour. Such a view would give educators the hope that, when aiming to develop virtue, they could do something more than to simply 'mould' behaviour. They could encourage capacities which would, of themselves, improve behaviour and which would, equally importantly, be immensely useful in other contexts (eg. in making decisions of prudence). The successful moulding of character and dispositions could then be seen as only a part, and ultimately a secondary part, of a wider, less 'authoritarian' project.

Educators would be able to give genuine reasons and explanations for the moral judgments they hold and, most importantly, give to their pupils the means of arriving at just the same judgments. Such a possibility would mean that those engaged in moral education could live up to the ideal of rationality which philosophers such as I.Scheffler (1973) describe - according to which the teacher's central task is to encourage the pupil to exercise his own judgment. "Teaching is, in this standard sense, an initiation into open rational discussion"; it is not merely the passing on of views and attitudes from teacher to pupil. It is not then a question of those who have a highly developed 'moral intuition', or who are peculiarly perceptive and sensitive to the 'moral features' of situations, passing on the results of their skills to others less fortunate, and ensuring that the latter are so disposed as to act according to the judgments received.

Moral judgment and an inclination to act.

The type of moral education which we are considering aims not only to impart an ability to make moral judgments but also to encourage a disposition to act accordingly (and thus to improve behaviour). The moral theory which we are considering describes the process of moral judgment in a way which makes clear not only the requirements of the first aim but also the way in which the achievement of that aim will contribute to the achievement of the second. If it is not possible to establish such a moral theory then it may be that a moral education which includes the second aim cannot avoid having as a **primary** task the imparting of certain fundamental attitudes or dispositions.

Hare's moral theory purports to avoid that consequence. But some commentators would claim that such a fundamental attitude (namely a sentiment of generalised benevolence) is surreptitiously appealed to by Hare.

Hare's route to Utilitarianism.

There are, as I understand it, two main propositions which the theory of critical thinking would yield:

> a. **as a result of the process of** making moral judgments, fully rational agents who know all the relevant facts will agree as to what is the morally right thing to do in a given situation;
>
> b. **as a result of the process of** making moral judgments, fully rational agents who know all the relevant facts will have a preference that the right action should be performed.

The first proposition says that moral disagreements must be the result of misuse of moral concepts or of ignorance

of 'non-moral' facts. As Nagel (1982) says, it implies that "there are no disagreements which are moral all the way down" and that "all fundamental moral disagreements are in a sense illusory".

One would expect someone who believes that moral judgments 'describe' moral facts to make the claim that full knowledge must result in agreement, but what is unusual about Hare's position is that this claim is made from a non-descriptivist position. Hare is not a naturalist, he does not believe that moral words are tied by virtue of their meanings to fixed non-moral properties; nor is he a realist or an intuitionist, he does not believe that there are 'moral facts' which can be perceived or intuited.

Both of the propositions above (a and b) are derived from claims about the meanings of moral words - viz. that they involve universalisability and prescriptivity. As Nagel says, Hare extracts a very large moral rabbit from what looks at first like a very small and empty linguistic hat.

This, perhaps, makes the second proposition even more surprising than the first. Hare claims that his theory (which initially concerns only the formal, logical properties of moral words) yields "a system of moral reasoning whose conclusions have a content identical with that of a certain kind of utilitarianism". But Hare's theory involves the further claim that the exercise of reason (according to the canons of moral reasoning) will, of itself, yield a preference to act in the way which such reasoning dictates. In the light of this additional claim, it will be useful to contrast Hare's route to Utilitarianism with earlier approaches.

Utilitarianism involves the assertion that an action is morally right insofar as it results in a maximisation of

Moral judgment and an inclination to act.

utility (or happiness or preference satisfaction). Such an assertion raises many questions. Two fundamental questions are: 'Why should we judge the morality of an action in terms of utility?' and 'Why should we act (or be inclined to act) according to judgments of morality/utility?'.

Bentham's response to the first question rests, in part, upon his belief that we do (to a great extent) judge the morality of actions in this way and that, insofar as we do not, we simply express our own personal and unfounded sentiments. The principle of utility, Bentham claims (1789 Chapter 1), offers the only means of giving meaning to the words 'ought', and 'right' and 'wrong'; and there has never been a "human creature breathing, however stupid or perverse, who has not on many, perhaps on most occasions of his life, deferred to it". If someone does not defer to that principle then he "expresses neither more nor less than the mere averment of his own unfounded sentiments"; and where the sentiments of two such people differ they can say no more than 'I like it' and 'I do not like it'.

Part of the appeal of Utilitarianism may be that, through the principle of utility, it appears to offer not only a means of justifying those moral views we share but also a means of resolving disagreements between our moral views. Utilitarianism seems to provide a simple, coherent **foundation** for particular moral views. According to Bentham, it provides the only such foundation.

Williams (1988) contrasts an approach in which we seek a 'foundation' for our moral opinions with an approach in which we are "merely .. concerned with the implications, presuppositions, and incoherences of those opinions". Williams favours the latter approach and denies that there is any need to 'go back to foundations' - in a 'Cartesian sense'. He criticises Hare for rejecting the

latter approach and for treating as mere prejudice moral views which have not been derived from such a foundation. Thus Williams attributes to Hare an approach which is certainly attributable to Bentham (and, perhaps, other Utilitarians).

Whether Hare would plead guilty to the charge of 'foundationalism' would depend upon how that expression is used. As Hare (1988 p.291-2) makes clear in his response to Williams, he would reject a 'Cartesian' approach based upon an appeal to **substantive moral opinions** which are claimed to be clear, distinct and self-evident. He would reject such an approach because as he says (1981 p.12) the opinions or convictions which are appealed to may indeed have been generated by prejudice and will merely reflect the moral environment in which each of us have grown up. This is the case whether the appeal is to those convictions favoured by a moral intuitionist or whether the appeal is to a single 'utilitarian' moral principle. None of these convictions are, according to Hare, shared by all and, far from providing foundations which offer the means of resolving moral disagreements, they themselves represent fundamental moral disagreements. If we are to provide 'foundations' then we cannot do so by means of substantive moral convictions.

Hare (1990 p.292) seeks "a secure method of moral reasoning", "based on an understanding of what we are up to when we are thinking morally", and "achieved by a thorough examination of the concepts we are using in our thought". According to Hare, we do not all share moral convictions but we do all share a use of certain words and concepts. Our use of those words and concepts involves our acceptance of the universalisability and prescriptivity of moral judgments and, Hare argues, if we examine carefully what that entails then we will see that moral reasoning must yield conclusions which have a

Moral judgment and an inclination to act.

content identical to that of a certain kind of Utilitarianism.

Hare claims to provide a rationally unavoidable and very distinctive route to Utilitarianism. An understanding of the proposed route allows us to see that Nagel's early criticism is not adequately argued. Nagel (1982) says that Hare, in order to perform the conjuring trick of producing a moral rabbit from a linguistic hat, has smuggled in a substantial moral intuition "the same one that Sidgwick saw to be the basis of utilitarianism: 'I **ought** not to prefer my own lesser good to the greater good of another'". This intuition, Nagel claims, is what allows Hare to derive a utilitarian position from limited claims about the logic of moral words. However, "there are those who do not share it".

Hare would agree - there are those whose upbringing has not resulted in a conviction that "the good of any one individual is of no more importance .. than the good of any other" (Sidgwick 1874 Book 3 Chapter 13) and there are those who have not understood the logic of their use of certain moral words and concepts.

Hare does not claim (like Bentham) that the principle of utility offers the only means of justifying moral opinion nor does he claim (like Sidgwick) that a substantive Utilitarian principle is intuitively self-evident. This is not the way in which Hare derives a utilitarian position. Hare's thesis is that: in each particular situation, and as a result of moral reasoning in accordance with the canons generated by the logic of moral terms, I **will** not prefer my own lesser good to the greater good of another. **The utilitarian preference for the greater good of another is, case by case, the result of moral reasoning** - this preference has its source in such reasoning. It is this fact, if it is a fact, which means that the results of moral reasoning will accord

Moral judgment and an inclination to act.

with the conclusions of a certain kind of utilitarianism. Nagel's claim (undoubtedly correct) that there are those who do not share Sidgwick's moral intuition is, as a criticism of Hare, misplaced.

We can see in more detail the distinctiveness of Hare's route to Utilitarianism if we consider the nature of the response which it allows to our earlier question 'Why should we act (or be inclined to act) according to judgments of morality/utility?'. That route would allow us to say (along the lines of the second proposition given at the beginning of this section) that those who make moral judgments **will** act (or be inclined to act) accordingly.

It could be claimed that if someone makes a **sincere** moral judgment then that logically entails that they have a preference that action should accord with that judgment. So that there is a sense in which the second proposition is true by definition. But what is important about that proposition in the context of Hare's theory, is the claim that the source of that preference is the process of moral reasoning. One could perfectly well perform utilitarian calculations out of idle curiosity and with no resulting inclination to act; but one could not perform critical thinking (as described by Hare, ie. as involving imaginative identification) in a 'disinterested' way - and it is critical thinking which, Hare argues, the rational agent making a moral judgment has to perform if he understands the logic of moral terms.

Utilitarians (following a different route) may argue that an appeal to **benevolence** is required in order to yield an inclination to act in accordance with the dictates of moral/utilitarian reasoning. For example, Smart (1973 p.7) says that the Utilitarian must, when addressing others, appeal to a shared sentiment of generalized

Moral judgment and an inclination to act.

benevolence - ie. 'the disposition to seek happiness, or at any rate, in some sense or other, good consequences for all mankind'.

Bentham (1834), on the other hand, argues that a genuine understanding of **our own self-interest** is what is needed in order to generate an inclination to act according to the conclusions of utilitarian reasoning. The aim of 'deontology', Bentham (1834 p.123) says, is to point out "to each man on each occasion what course of conduct promises to be in the highest degree conducive to his happiness: to his own happiness, first and last; to the happiness of others, no further than insofar as his happiness is promoted by promoting theirs; ... what will also be shown is in how many different ways, more than is very generally understood, each man's happiness is ultimately promoted by an intermediate regard shown in practice for the happiness of others".

Bentham (1834 p.148) claims that "the intrinsic and ultimate object of pursuit to every man at all times" is his **own** well-being. He regards "as an incontrovertible fact, that no man ever has done or ever can do any act which at the moment of action is not .. , in his own eyes at least, his interest to do" (p.175). Given this claim, and given the assertion that the morality of our actions is measured by their utility, then the only purpose of ethics must be to point out the extent to which the pursuit of the greatest happiness of the greatest number (extra-regarding interest) serves the pursuit of our own happiness (self-regarding interest) (p.192). The extent to which the one serves the other is (Bentham insists) much greater than is commonly thought.

It is in this context that Bentham makes some appeal to benevolence. Many of us, on many occasions, will feel sympathy towards others and will thus gain pleasure through acting in ways which bring about the pleasure, or

prevent the pain, of others. But Bentham goes on to refer to two other factors which will provide a motive for acting virtuously. Firstly, through acting in ways which show our regard for the well-being of others we will gain a reputation which will increase the regard which others show for our well-being. Secondly, through acting in ways which benefit those with whom we have regular dealings we increase the chance that the individuals benefited will reward us at some later date. These two factors Bentham (1834 p.184) sees as providing an inducement which "is of the same sort as that which the husbandman has for the sowing of his seed, or that which the frugal man has for laying up his money". "By every act of virtuous beneficence which a man exercises, he contributes to a sort of fund - a sort of Saving Bank - of general good-will, out of which services of all sorts may be looked for as about to flow on occasion out of other hands into his".

I have considered Bentham's route to Utilitarianism in some detail because it presents very starkly the problems which are raised when we try to discover factors relating to our **existing** motivations and preferences which could provide an inducement for acting according to judgments of morality/utility. If our motivations and preferences cannot be influenced by moral perception or by self-evident substantive moral intuition, and if they do not already include a sentiment of generalised benevolence, then we appear to be driven towards factors which seem very unsatisfactory both in extent and nature. My rational desire to build up substantial funds in a Saving Bank of good will seems to provide an insufficient and **highly inappropriate** inducement to moral virtue.

There are alternatives and one is to claim, with Hare, that the **process** of making rational moral judgments involves a form of thinking which brings with it an inclination to act. By identifying moral thinking with a

form of thinking which involves acquiring the preferences of others, Hare seems to avoid the problem of discovering motivations or preferences which could provide an inducement for acting accordingly. The motivation to act is provided by those preferences which I acquire through the imaginative identification which is required by moral thinking.

But, as Hare himself points out and as we shall see in a later chapter, the problem is merely shifted so that the question becomes: 'Why engage in moral thinking?'. That, in turn, makes central the question: 'Why educate ourselves and others to be inclined to engage in moral thinking?'.

I too will seek to argue for an identification between moral thinking and critical thinking. However, firstly, the description of critical thinking which I shall offer will be somewhat different to Hare's. That difference will stem from my criticism of Hare's argument. Secondly, the answer I shall attempt to give to the central educational question ('Why educate ourselves and others to be inclined to engage in moral/critical thinking?') will be very different to Hare's. That difference will stem from my rejection of the form of Utilitarianism which arises from Hare's argument. It will involve a shift of focus (in morality) from outcome to agent and (in moral education) from performance to motivation.

CHAPTER 3.

Consequentialist and non-consequentialist moral theories.

Rational choice and morality.
Consequentialism and non-consequentialism.
Consequentialism as involving a ranking of consequences.
Consequentialism as involving no agent-relative features.
Consequentialism and non-consequentialism redefined.

Rational choice and morality.

The first steps in Hare's argument (as presented in the previous chapter) concern the rationality of choice - 'What makes a prescription rational?'. It is claimed that:

a. a fully rational agent only makes a prescription for action in a given situation if he knows what are the consequences of the various possible actions in that situation;
b. the prescription which results from such knowledge is rational only if it depends upon what are the preferences of that agent with regard to those consequences.

At that stage of the argument, we are concerned only with prescriptions which are a response to the question 'What shall I do now?'; we are not yet concerned with prescriptions as a response to the question 'What ought I to do now?' (where that 'ought' is a moral ought).

Hare follows Brandt (1979) in saying that 'rational' refers to "actions, desires, or moral systems which survive maximal criticism by facts and logic". In the context of actions, this definition requires that a

rational agent gains knowledge of any facts which might affect the decision as to how to act. A prescription will fail to be fully rational insofar as the agent lacks knowledge such that having such knowledge might have resulted in a different prescription.

Prescriptions express choices and decisions, and we cannot rationally decide what to do unless we know what we would be doing if we did this or that. As Hare (1952 p.56) says, "The whole point about a decision is that it makes a difference to what happens; and this difference is the difference between the effects [consequences] of deciding one way, and the effects [consequences] of deciding the other.". The prescription made will then depend upon what my preferences are with regard to those consequences. Any consequences which relate to those preferences, and might thus make a difference to the decision, will need to be considered if the prescription is to be fully rational.

When moral considerations are not involved then rational consideration of the question 'What shall I do?' requires knowledge of what acting in different ways would entail in the given circumstances (knowledge of consequences). Furthermore, the consequences which we, as rational agents, are required to consider are those which relate to our preferences. If some consequences are not related to my preferences then I may not need to consider them; but if other consequences are related to my preferences then I may be foolish to neglect to consider them.

When moral considerations are not involved then preferences and consequences are not only relevant to the appraisal of actions, they are central. If we aim to give our educatees the ability to appraise actions and to come to informed decisions (rather than relying upon the guidance of others) then we will aim to develop an ability to consider consequences and preferences.

Consequentialist and non-consequentialist moral theories.

Now if morality has anything to do with looking at things from a wider perspective than that of our individual concerns and preferences, then it would seem natural to extend what has so far been said in such a way as to relate it to the preferences of all concerned rather than to the preferences of a single agent. Thus: when **moral** considerations are involved then rational consideration of the question 'What **ought** I do?' requires knowledge of what acting in different ways would entail in the given circumstances (knowledge of consequences); and, furthermore, the consequences which we are required, as rational agents, to consider are those which relate to the preferences of **all** those involved.

As Rawls (1971 p.23 and p.27) says, when describing the attraction of classical utilitarianism, "why should not a society act on precisely the same principle applied to the group and therefore regard that which is rational for one man as right for an association of men?". Thus "the most natural way .. of arriving at utilitarianism .. is to adopt for society as a whole the principle of rational choice for one man".

Whether such a parallel between the approach to questions of prudence and questions of morality does seem 'natural' will, of course, depend upon the way in which one is disposed to approach 'moral' questions about actions. But what is true is that most of us do, on some occasions, consider how alternative actions might affect others - we sometimes do appraise possible actions in terms of what they are likely to entail in the circumstances and in the light of the preferences of others. What is also true is that most of us are, on some occasions, motivated by such considerations - you go to see someone because they are expecting you and you know that they will be disappointed if you do not go, I buy my daughter roller skates because I know that she

will be pleased, the doctor tells someone that they are dying because he believes that they would prefer to be told, and so on.

Can we not say that we are, in such cases, deciding what we (morally) ought to do and that we are motivated by moral considerations? We may not wish to go so far as to claim that moral appraisal of actions, and moral motivation, **only** involves consideration of consequences in the light of the preferences of others. But, surely, it would not be unreasonable to investigate deliberations about consequences, in the light of the preferences of others, on the grounds that they may have some relevance to a moral theory.

Yet some would claim that investigation of such deliberations cannot have a central place in a moral theory because, when it comes to answering the moral question 'What ought I to do?', considerations of consequences are (sometimes) simply not relevant. As Anscombe says (1958 p.192): "there are certain things forbidden **whatever** consequences threaten".

Indeed in some cases, it is claimed, such deliberation would not only be irrelevant but would also be morally wrong. If someone thinks, Anscombe says, that it is **open to question** "whether such an action as procuring the judicial execution of the innocent should be quite excluded from consideration - I do not want to argue with him; he shows a corrupt mind". Those who are willing to suspend judgment until a consideration of consequences and preferences has taken place sometimes show a corrupt mind.

If this were so then those who educate others in a way which results in them (always) making 'moral' decisions through a consideration of consequences and preferences would, presumably, be responsible for that corruption.

The acquisition of that ability may contribute to other aspects of a person's education; but, in order for moral education to take place, the educatee must learn that the exercise of such abilities is (at least sometimes) not a means to moral judgment.

The sort of moral intuition which Anscombe is expressing has been used to mark a general contrast between 'consequentialist' and non-consequentialist' moral theories. The way in which that contrast has been made has, however, varied a great deal. Some contrasts fail and others focus upon different aspects of moral evaluation of actions.

Since, throughout this thesis, I will not be using those terms in the way in which they are currently used in debates over moral theory, I now wish to consider the ways in which that contrast has been made and to give reasons for the focus which I shall recommend.

Consequentialism and non-consequentialism.

Consequentialists are often contrasted with deontologists - 'one must do one's duty regardless of the consequences'. The deontologist is sometimes characterised as one who believes that the moral value of an action is a feature of the action itself as opposed to being a feature of the consequences (or effects, or extrinsic features) of the action. For example, Hudson (1970 p.87) characterises the deontologist as one who holds that "the rightness or wrongness, goodness or evil, of an action is intrinsic to the action itself".

We might begin by pointing out that if the deontologist were to accept that the expression 'the consequences of a particular action' is equivalent to something like 'that which is the case given the performance of the action in

Consequentialist and non-consequentialist moral theories.

the present circumstances', and if it were the case that consequences so defined are not relevant, then it would be difficult to see how the fact that one has done one's duty could itself be morally relevant. As Hare says (1952 p.57), "Even to do our duty - insofar as it is **doing** something - is effecting certain changes in the situation.".

If 'consequences' are defined in the same way that Hare defines 'effects' (ie. so broadly that it refers to anything which is the case given the performance of the action in the present circumstances) then it may appear necessary to make a distinction between **types** of consequence. The deontologist, who (say) claims that we always have a duty to tell the truth or that we are always forbidden to kill an innocent person, singles out some types of consequences as relevant to moral appraisal. So too the 'consequentialist', who (say) claims that one ought to act so as to maximally satisfy the preferences of all concerned, singles out other types of consequence. Thus we might attempt to find a way of contrasting such types of consequence other than by merely listing them.

Mackie (1977 chapter 7) characterises the consequentialist as one who builds a moral system around the notion of some 'goal' to be attained. He also, like Hudson, seems to identify the non-consequentialist with the deontologist; and sees the deontologist as one who builds a moral system around the notion of rules, principles, duties, rights, or virtues. The consequentialist, Mackie suggests, sees as central the prescription: 'Act so as to bring about X' (where X is the goal, or a disjunction of goals, to be attained); he may give some place to rules, principles, etc. and thus prescribe 'Do things of kind Y', but only insofar as such things are conducive to the goal(s) specified. The

deontologist, on the other hand, sees as central the prescription 'Do things of kind Y'.

But, once again, if any consequence of an action (as defined above) can be taken as a 'goal' then this notion does not seem to be particularly helpful. If an action can be described, say, as one of telling the truth then we can say that a consequence of that action is that the truth is told. The deontologist who tells us that we ought to tell the truth ('Do things of kind Y') is telling us that those actions in which we make an assertion ought to have the consequence that the truth is told ('Act so as to bring about X'). If the idea of a 'goal' is to be the basis of a genuine distinction between the consequentialist and the non-consequentialist then it will have to be defined more narrowly so that some consequences are not specifiable as goals.

Perhaps there are other ways in which we can make a contrast between types of consequence. For example, 'intrinsic' consequences (features of the action itself as Hudson might call them) as opposed to 'extrinsic' consequences.

This approach would seem to require that, from amongst all the possible descriptions of an action, we are able to distinguish those that are a description of 'the action itself'. Other statements would either not concern the action, or would describe extrinsic features of the action. The non-consequentialist might then be characterised as one who would claim that the only consequences which are morally relevant are those consequences which are described in statements which are logically entailed by a description of 'the action itself'.

But it is doubtful whether a distinction can be made which will serve the purpose. If, for example, it is

true that if a small tumour is bombarded with radiation then it will, in the following days and weeks, shrink and disappear. Then we might say of a radiologist treating a patient that he is:

1. destroying the tumour;
2. bombarding the tumour with radiation;
3. pointing the radiation device at the tumour and switching it on;
4. turning the device in such a direction and pressing this lever;
5. holding the device, flexing such and such muscles..

It would seem to be possible to describe the 'action itself' in any of these ways; and, given the approach outlined here, what is or is not an intrinsic feature of the action (and what is or is not an extrinsic feature or effect of the action) will depend upon which description is employed. If we cannot distinguish descriptions of an action which describe 'the action itself', then the required distinction between the morally relevant and the morally irrelevant features of an action will depend upon which description is being considered.

We can make a point about action-talk which is similar to that made by Melden (1961). The conviction that we can clearly distinguish the consequences of an action from the action itself may be based upon a false picture of the way in which we talk about actions. That picture is one in which 'the action' is some element, or combination of elements, from amongst a causal chain involving various 'happenings' - some concurrent and some consecutive - for example, a decision to destroy the tumour, various bodily movements, the pointing of the device, more bodily movements, the switching on of the device, emission of radiation, bombardment of the tumour, absorption of the radiation, shrinking of the tumour, disappearance of the tumour. 'The action' is then seen

to occur at some point in this causal chain and to have various consequences which occur at later points in the chain - 'it' brings about the emission of radiation, the bombardment of the tumour, and the ultimate disappearance of the tumour.

It does indeed seem natural to say, for example, that **by** turning the device in this direction and switching it on he **brought about** the emission of radiation and the bombardment of the tumour and so on. But we may say (equally naturally and with reference to the very 'same action') that **by** bombarding the tumour with radiation he **brought about** its destruction.

The distinction between what is done and the consequences of what is done is entirely relative to the way in which we choose to describe the action. If the terms 'consequentialism' and 'non-consequentialism' are to mark some generally significant distinction then we will need to consider features of moral judgment other than the fact that it involves evaluation of whatever is the case given the performance of the action.

I shall consider three approaches to that distinction. Given those different approaches **consequentialist** theories will be those in which (very roughly) evaluation of action involves:

a. a ranking of overall consequences;
b. no agent-relative features;
c. no motive-relative features.

Consequentialism as involving a ranking of consequences.

The consequentialist is here characterised as one who claims that the moral appraisal of an action involves a comparison between the consequences of that action and those of alternative actions according to some general

principle of evaluation. The application of such a principle determines a ranking of those sets of consequences such that one (or more) set is 'best', and the 'right' action is then the action with that set of consequences.

The non-consequentialist might then be characterised as one who claims that, on the contrary, some actions are just right or wrong regardless of any such comparison. The applicability of one of certain descriptions (eg. truth told, innocent person killed) to a consequence of an action is **always** sufficient to determine the nature of the moral appraisal.

As Hampshire says (1978 p.7), there are certain moral impossibilities which belong to the very notion of morality, "a morality is, at the very least, the regulation of the taking of life and the regulation of sexual relations, and it also includes rules of distributive and corrective justice; family duties; almost always duties of friendship" and so on. Some things are just wrong and others are just right; morality involves certain prohibitions and duties. Once we know that a certain description would apply to the consequences of a particular action then the moral judgment is not open to question.

It is the consequentialist's willingness to consider all the actions which are possible in a particular situation, to fail to rule out of court (or to immediately accept) some possible actions despite the appropriateness of certain descriptions, which is felt (by some) to be not only incorrect but also unacceptable. The consequentialists show 'corrupt minds' because they are willing to withhold judgment until they have considered all the consequences of that action and compared them with those of other possible actions in the circumstances. They may agree that the fact that the

consequences of an action may be described as 'the killing of an innocent person' is highly relevant to moral appraisal but deny that that is, of itself, sufficient to settle the issue - the consequences of all alternative actions may be considered and compared.

Williams (1973) initially characterises consequentialism by its focus on the causal properties of an action (the states of affairs, or sets of consequences, brought about by an action) as opposed to a focus on the action (itself). Williams points out, and we have already noted, the difficulties in clarifying the terms used here. He then goes on to offer a characterisation of consequentialism along the lines suggested here.

Consider the following statements:

a. in S, x did the right thing in doing A,

b. the state of affairs P is better than any other state of affairs accessible to x,

(in which P may be what is brought about by A and/or may consist of x's doing A).

A consequentialist view will be one in which b. is given as a reason for a., and (perhaps) a. only has sense because b. has sense. As Williams points out (1973 p.88/89) a non-consequentialist view may then involve one of three responses:

no sense is attached to b.;

b. has sense only insofar as a. is true;

b. has some independent sense but is not relevant to moral appraisal of the action.

As Williams says (1973 p.88), the non-consequentialist (who responds in the first or second way) "may have no general way of comparing states of affairs from a moral point of view at all", and the "emphasis on the necessary comparability of situations is a peculiar feature of

consequentialism in general, and of utilitarianism in particular".

Some forms of utilitarianism appeal to only one feature of the consequences of actions when making a comparative evaluation - the overall utility (or, in Hare's case the preference satisfaction) which is involved. Such a theory can, therefore, be called a 'monistic' consequentialist theory.

But a consequentialist theory may also be 'pluralistic' - it may refer to several features of the consequences of actions when making a comparative evaluation. For example, S.Scheffler (1982 Chapter 2) describes a form of pluralistic consequentialism in which the best set of consequences is worked out thus: maximise the well-being of the group which is worst off; if sets of consequences are identical in this respect then minimise the number in that group by moving them up; and so on. Thus we would have a 'distribution-sensitive' form of consequentialism in which moral evaluation requires consideration not only of well-being but also of the distribution of well-being.

We could now incorporate references to other features in such a way that a pluralistic 'consequentialist' theory began to look very much like a 'non-consequentialist' theory. For example, we could regard a set of consequences in which innocent people are killed as worse than any other set of consequences (regardless of, say, the overall level of well-being which is associated with those alternative sets of consequences). Thus our ranking of consequences may entail a set of prohibitions (and duties) such that the applicability of certain descriptions to the consequences of an action **is** (sometimes) sufficient to determine moral appraisal of the action.

However, a ranking which entails such a set of prohibitions and duties may, nevertheless, fail to ensure that the applicability of a certain description is **always** sufficient to determine moral appraisal. If, say, we regard a set of consequences in which innocent people are killed as worse than any other set of consequences then we may also regard a set of consequences in which ten innocent people are killed as worse than one in which one innocent person is killed.

Thus, if (to take a standard example) someone threatens to kill ten innocent people unless I kill this one innocent person (and if it is assumed - unrealistically - that I know it to be the case that the threat is genuine, and I have no other means of preventing it from being carried out, and so on) then, presumably, a consequentialist view entails that I ought to kill an innocent person.

There can be different 'non-consequentialist' responses to this particular example. These responses are the same as those given (in general terms) above:

'one innocent person killed is better than ten innocent people killed' has no sense;

one innocent person killed would be 'best' only if it were right in this situation (which it is not) to kill one innocent person;

'one innocent person killed is better than ten innocent people killed' may be true but it is not relevant to moral appraisal of the action.

The non-consequentialist's claim that some actions are just right or wrong regardless of any comparative evaluation of sets of consequences may involve a rejection of the possibility of such a comparison **or** a rejection of the relevance of such a comparison to moral appraisal of action. Furthermore, both of these stances can be adopted in general or, as we have seen in the

example, in particular contexts. Finally, both stances can be adopted in response to a very wide range of different types of comparative evaluation.

We have here started from a characterisation of the consequentialist as one who claims that the moral appraisal of an action involves a comparison between the consequences of that action and those of alternative actions according to some general principle of evaluation. We then characterised the non-consequentialist by describing responses to a consequentialist position. We could approach the contrast from the other side. However, we will begin not with a general stance but with a stance in response to a particular context.

Consequentialism as involving no agent-relative features.

In the context of the example (in which someone threatens to kill ten innocent people unless I kill one), we can accept that the death of ten innocents is worse than the death of one but deny that I ought to kill one **only if** we can point to some other feature which is relevant to moral judgment.

In the example I am faced with a choice between bringing it about that **I** kill one innocent person or bringing it about that **someone else** kills ten innocent people. We can resist the move from 'best' to 'ought' if we insist that (in the context of killing innocent people) 'ought' is 'agent-relative':

(x) (x ought to bring it about that **x** does not kill innocent people)

or, more simply:

(x) (x ought not to kill innocent people).

We would then be claiming that (in this context) we cannot derive a moral judgment about action from our comparative moral evaluation of sets of consequences since an agent-relative moral judgment already applies.

We could follow Parfit (1984 p.27) in putting this response in terms of a claim that each of us ought to have the aim that **he** does not kill innocent people - as opposed to the claim that each of us ought to have the aim that there is no, or less, killing of innocent people.

If our moral aims were **all** of this type then, in all contexts, comparative 'moral' evaluation of sets of consequences would cease to be relevant to moral judgments about action. Such a position would also be compatible with one involving a claim that 'moral' evaluations of that type are not possible. Thus both responses to the consequentialist (given at the end of the previous section) could be expressed in terms of the role of agent-relative features in our moral judgment.

If the contrast between consequentialism and non-consequentialism is drawn in these terms then we are able to characterise the consequentialist theory as one which makes no appeal to agent-relative features (and, perhaps, appeals only to a ranking of sets of consequences) and the non-consequentialist theory as one which appeals only to agent-relative features (and makes no appeal to a ranking of sets of consequences). We can also characterise 'hybrid' theories (to use S.Scheffler's expression) as those which involve appeal to some agent-relative features but otherwise appeal to a ranking of sets of consequences.

[Scheffler also points out that non-consequentialist or hybrid theories can incorporate agent-relative features which, rather than **forbidding** me to perform the action

which has the best outcome, grant me **permission not** to perform that action.]

If we combine the contrast here with that in the previous section we have:
consequentialism -

a. makes no appeal to agent-relative features,

b. appeals only to a ranking of sets of consequences;

non-consequentialism -

c. appeals only to agent-relative features,

d. makes no appeal to a ranking of sets of consequences.

Now both Scheffler (1988 p.5) and Parfit (1984 p.26/27) seem to claim that these features go hand in hand. However, it seems to me that this is not so.

We could have a moral view according to which it is always wrong to bring it about that an innocent person is killed, someone is tortured, ten innocent people are killed, ten people are tortured, and so on. We could see each of these as wrong regardless of whether I bring about these consequences or I bring it about that someone else brings about these consequences (ie. no appeal to agent-relative features). We could see each of them as just wrong regardless of any comparison with the consequences of alternative actions (ie. no appeal to a ranking of sets of consequences).

In the context of the example, such a view would yield the conclusion that my action would be morally wrong whatever I did. The non-consequentialist who appeals to agent-relative features says that I ought to not kill one (I thus bring it about that the other person does kill ten). The consequentialist who appeals to ranking says I ought to kill one. The moral view considered here yields the conclusion that I ought not to do either - but, alas, I must. Such a view would be non-consequentialist

according to the characterisation in the last section and consequentialist according to the characterisation in this section.

Such a view is, I believe, coherent. Furthermore, if we combine it with a view of blame (according to which to do wrong is not always to be worthy of blame) it yields a plausible interpretation of our moral intuitions - our "spontaneous convictions, moderately reflective but not yet theorized" - as Williams (1985 p.94) describes them. This is not an interpretation which I would support but, when faced with the dreadful situation in which if I do not kill an innocent person someone else would **definitely** kill ten, our **moderately** reflective convictions are not (I believe) so clear as to rule out such an interpretation.

There are many different ways in which we can interpret our moral intuitions with respect to actions which we find repugnant. I shall offer my own interpretation in a later chapter but here I shall roughly describe a Utilitarian interpretation. According to Hare, for example, although a Utilitarian moral theory may yield the conclusion that in some situations killing an innocent person would be morally right, it is also true, firstly, that such situations would be extremely rare (especially if wider consequences are taken into account) and, secondly, we could never be sure that a given situation was of that type and, thirdly, a Utilitarian theory can provide very good reasons for educating people to find such actions extremely repugnant.

Scheffler (1988 p.9) claims that this leaves a gap between the interpretation and the intuition, and that most would agree that agent-relative moralities "mirror our everyday moral thought much more closely than consequentialism does". Does this 'most' and 'our' refer to (moderately reflective) moral philosophers? What does

'more closely' mean? Does it mean that an agent-relative interpretation of moral intuition is somehow part of that intuition? Or does it mean that when we try to state those intuitions we do so in agent-relative terms? A proponent of agent-relative morality will do so; but a Utilitarian will not.

There are different ways of drawing the boundary between consequentialists and non-consequentialists - some are not clear, others draw the boundary in somewhat different places. If the boundary is meant to divide those who **do** from those who **do not** have difficulty with moral intuitions with respect to actions which we find repugnant (or with, say, moral intuitions about personal responsibility) then, I believe, the attempt to draw the boundary in this way may do two things.

Firstly, it may obscure the fact that there are a range of coherent and plausible ways in which we may interpret those moral intuitions. These interpretations may fall on one side of the boundary, on the other, on both, or on neither. Monism, pluralism, ranking of sets of consequences, agent-relativism are all important features of moral theories but to draw a line through those features (to act as a boundary between consequentialism and non-consequentialism) is to make an unnecessary or a false contrast. The contrast is unnecessary if it is made in terms of one feature (non-consequentialism is just, say, agent-relativism). The contrast is false if it implies that several features go hand in hand on one side of the boundary. Furthermore, all such features have something in common: they relate to an evaluation of what is the case given the performance of an action - they relate to **consequences**.

Secondly, it may prevent a focus upon aspects of moral appraisal which are at least as important as those relating to consequences. It may obscure the fact that

in some moral theories all such evaluation is derivative. In a 'non-consequentialist' moral theory the agent, and the complexity and quality of motivation which leads to his action, has centre-stage.

Consequentialism and non-consequentialism redefined.

Throughout the previous sections it was assumed that moral appraisal of an action was centrally concerned with what is the case given that an action is performed. Non-consequentialism has so far been characterised as involving an insistence that agent-relative features are central, or as involving a rejection of certain ways of making comparative evaluations between consequences. But perhaps non-consequentialism is better seen as a demand for a shift of focus from the world, and the way it is being altered by action, to the agent and the sources and 'springs' of his action.

The focus on what is the case given that an action is performed is a focus on consequences, and the insistence that moral appraisal is **primarily** a matter of evaluating consequences (as good or bad, better or worse) is perhaps what distinguishes the consequentialist. Agent-relative theories require a shift of focus; but the focus is not shifted away from consequences, it is merely narrowed so that moral appraisal concerns, say, whether I bring it about that **I** kill an innocent person rather than whether I bring it about that **I or another person** kills an innocent person.

All consequentialist theories may lead us to see actions only in terms of what is the case given the performance of an action but I shall, for the moment, confine my remarks to Hare's theory.

Hare's analysis centres upon consequences, possible and actual, which agents are interested in because those consequences would satisfy, or fail to satisfy, their preferences. The agent's struggles are all with the world and the aim of those struggles is to make the world such that those preferences are satisfied. The struggle becomes a moral struggle when the prescriptions involved are not only those of the agent but also those of other people.

There is little mention of the agent's struggle with himself or others, where that struggle is seen in terms of people who are (in themselves) morally imperfect and in need of improvement. It is then easy to be tempted into seeing self-improvement only as a matter of learning and modifying skills and principles in the light of changes in the world and of improvements in our knowledge of the world - ways of doing better in the business of improving the world.

When Hare (1952 p.72-75) discusses the possibility of the instability of moral principles, or the appropriateness of passing on (unaltered) our moral principles to our children, the explanation of instability and the questioning of appropriateness is all in terms of changes in the world or changes in our knowledge of the world. What Hare sees as central are the implications of such changes for the way in which we can successfully bring about the preferred consequences: principles become inappropriate because, in the altered circumstances, the same way of behaving no longer brings about those preferred consequences.

Hare does not seem to envisage the possibility that principles could come to be seen as inappropriate because of a rejection of the preferences which they are designed to satisfy. Or, at least, he does not seem to allow that the question 'What ought I to prefer?' may be central,

and independent of, the question 'How can I satisfy these preferences?'.

Hare's later discussions (1981) of the possibility of altering preferences turn on answers to the question 'How would alteration of this preference affect the satisfaction of these other preferences?'. The moral theory he outlines places no overall constraints on preferences, it concerns only the problem of how we must behave if we are to bring about those consequences which would maximally satisfy the preferences which people actually have. "The effect of universalisability is to compel us to find principles which impartially maximise the satisfaction of .. preferences. It does not constrain the preferences themselves." (1981 p.226).

If Hare's theory is offered as a 'complete' theory of moral thinking then this implies that questions about whether there are independent constraints upon what people prefer (and, if so, then what are those constraints) not only can not be answered by moral thinking but also **are not moral questions at all.**

I believe that such questions may be as central to morality as the particular form of 'concern for others' which critical thinking represents. If that is so then moral education may have to **centrally** involve an effort to shape not merely the behaviour of the agent but the character of the agent. Moral education will involve aims which do not merely relate to what the agent achieves through action but also, and more importantly, which relate to what the agent does as a result of who he is.

In this section I merely wish to suggest that we might characterise the consequentialist as one who (like Hare) sees moral appraisal as primarily a matter of evaluating consequences. The non-consequentialist may then be

characterised as one who insists that moral appraisal is primarily about agents (not primarily about what happens in the world as a consequence of action) - the motives, intentions, preferences, dispositions, capacities of the agent are the primary focus.

But now we may note, with Moore (1966 p.95), that although it may be admitted that "it is right and proper that a man's motives should ... influence our judgment ... the question is: What **sort** of moral judgment is it right and proper that they should influence? Should it influence our view as to whether the **action** in question is right or wrong? It seems very doubtful whether ... it actually does ... for we are quite accustomed to judge that a man sometimes acts **wrongly** from the best of motives". And again, with Mill (1863), "the motive has nothing to do with the morality of the action, though much with the worth of the agent".

The consequentialist may in fact agree that moral appraisal of **agents** should focus on motives etc., but nevertheless insist that moral appraisal of **actions** is a matter of evaluating consequences. However, my point here is that we should see the non-consequentialist as demanding a shift of focus away from consequences and that shift can be achieved by taking a certain view of **the relationship between the two forms of appraisal.**

Thus I wish to suggest that the consequentialist may be usefully characterised as one who insists that:

> a. moral appraisal of actions is primary and depends upon an evaluation of the consequences of the action;
>
> b. moral appraisal of agents is secondary and derivative because the dispositions, motives, etc. of the agent have moral value only insofar as they

are conducive to (or would generally result in) those consequences which have value.

The non-consequentialist may then be seen as insisting, on the contrary, that:

c. moral appraisal of agents is primary and depends upon an evaluation of the characteristics of the agent;

d. moral appraisal of actions is secondary and derivative because actions have moral value only insofar as they are such as would be performed by an agent with those characteristics which have value

[or, with Aristotle perhaps, insofar as they are such as would be performed by an agent acting in those ways which would result in his acquiring those characteristics].

The non-consequentialist may see characteristics such as honesty, benevolence, compassion, loyalty, sympathy, understanding, as good in themselves; but for the consequentialist such virtues only have value because they generally lead to consequences which have value.

In this thesis I shall offer a moral theory which sees benevolence, non-malevolence, understanding and humility as the primary focus of morality and of moral education. That theory will not only be a non-consequentialist theory, it will be a 'radically' non-consequentialist theory. I shall end this chapter by characterising such 'radical non-consequentialism'.

The characterisations of consequentialism and non-consequentialism given above permits us to make a further useful distinction. The characterisation of non-consequentialism implies (through d.) that a **particular**

action may have moral value even if it is **not** performed **as a result of** the characteristics which have value. However, one might wish to claim that if characteristics of the agent are the primary focus in moral appraisal then a particular action has moral value only if it is due to those characteristics. That is, we replace d. with:

> d'. moral appraisal of actions is secondary and derivative because actions have moral value only insofar as they are performed by an agent **as a result of** those characteristics which have value.

According to this view the moral appraisal of action is not a matter of what is done but may be wholly a matter of why it is done. The moral appraisal of actions is not merely secondary to, and derived from, the moral appraisal of agents, it is directly dependent. This view I shall call **radical** non-consequentialism.

Non-consequentialist theories may incorporate an ideal - the ideal agent who possesses all those characteristics having positive moral value and lacks all those characteristics having negative moral value. Such an ideal may generate a view of right action for a given situation - the action which would be performed by an ideal agent. If the theory is radically non-consequentialist then it will yield the view that an action which is right but which does not result from such characteristics is **merely** right - it lacks moral worth.

Kant offers such a theory. For Kant the characteristic of the agent which has moral value is his apprehension of duty. An action has moral value only insofar as it is brought about by that apprehension of duty. An action which conforms to duty but does not arise from duty has mere 'legality'.

To encourage educatees to do right because, say, they believe it is in their own interest, or they wish to please their educators, or they expect reward and fear punishment would not only contribute nothing to their moral development but would also achieve nothing of **moral** worth. In the context of Kant's moral theory, moral education must aim to reveal to educatees that apprehension of duty can have more power than "all allurements arising from enjoyments and everything which may be counted as happiness or from all threats of pain and harm" (1788 p.155).

I too will offer a radically non-consequentialist moral theory. That theory will generate a view of right action, and of the moral worth of the agent, which has much in common with Hare's view. The structure of the theory, and features which result from that structure, will have much in common with Kant's view. Both Hare and Kant offer arguments for the 'correctness' of their moral theories. I shall consider both theories, and both arguments, before presenting my own.

In the next chapter, I shall look at various features of Kant's moral theory and, especially, of the argument which gives rise to his radical non-consequentialism.

CHAPTER 4.

Moral thinking, moral motivation and moral worth.

Kant's radical non-consequentialism.
Morality and freedom.
Moral experience and moral education.
Objections to Kant's theory.
Choosing to act against inclination.

Kant's radical non-consequentialism.

Kant (1785 p.11-18) claims that nothing "can be called good without qualification, except a **good will**" and that "a good will is good not because of what it performs or effects, not by its aptness for the attainment of some proposed end, but simply by virtue of the volition".

The good will is a source of action. To act from a good will is to act from duty (the notion of duty "includes that of a good will"), and that is to act not out of inclination but simply because of the moral law. Action "done from duty must wholly exclude the influence of inclination, and with it every object of the will, so that nothing remains which can determine the will except objectively the law".

The moral worth of **the agent** is measured by the source of his actions - he has moral worth insofar as he acts from duty. But the moral worth of **his actions** is also, for Kant, measured in the same way. To have moral worth an action must be done from duty, and "an action done from duty derives its moral worth, not from the purpose which is to be attained by it, but from the maxim by which it is determined, and therefore does not depend on the

realisation of the object of the action, but merely on the principle of volition by which the action has taken place".

Not only is the satisfaction of inclinations irrelevant but also what is **in fact** achieved by an action. If my respect for the moral law 'One must always tell the truth.' determines my action, then that, of itself, establishes the moral worth of my action. Whether I have, in fact, told the truth cannot therefore be relevant to the moral appraisal of the action.

If I choose to act out of respect for the moral law 'Always tell the truth.' then actions which are the result of that choice have moral worth (if it is indeed a moral law). Knowledge of the source of an action is sufficient to moral appraisal. If, for example, I believe that my brother is in the garden and if, because I choose to act out of respect for moral law, I say 'My brother is in the garden.' then my action has moral worth. But my belief may be mistaken (he may now be elsewhere) and, in that case, I have not told the truth. My action is determined by the law but it is not, in fact, according to the law. If the former is sufficient to moral appraisal then the latter cannot be relevant. Consequences, even in the broadest sense (according to which that I have not told the truth is here a consequence of my action), are not relevant to the moral appraisal of the action.

The particular nature of an action (what it performs or effects) will depend upon the circumstances and upon my beliefs; the moral worth of the action depends entirely upon the nature of the motivation which gives rise to it. Kant's **radical** non-consequentialism can be seen clearly if we contrast this with his view of the 'prudential worth' of an action.

Kant (1788 p.17-19) makes a distinction between practical principles as maxims and as laws. A practical principle is "a proposition which contains a general determination of the will" and might be something like:

'One should provide for one's old age.'

It may "have under it several practical rules" (or precepts); and here, I take it, Kant is referring to principles like:

'One should save when young in order to provide for old age.';

'One should befriend rich people in order to provide for old age.'

(The first is a "correct and important practical precept of the will")

A practical principle is a subjective principle (a maxim) when "the condition is regarded by the subject as valid only for his own will". This will be so, according to Kant, if the subject recognises that the principle is not valid simply of itself but rather because the end it specifies (provision for old age) is something the agent wishes to achieve.

The maxim and the practical precepts which it 'has under it' yield imperatives:

1. provide for your old age;
2. save when young;
3. befriend rich people.

Insofar as reason determines the will, and given

a. I wish to provide for my old age,

b. I believe I will live to an old age,

then I will choose to provide for my old age

ie. I will recognise imperative 1. to be valid for me.

Insofar as reason determines the will, and given

a. and b. and
c. I believe I can save something,
d. I believe I cannot expect much from rich friends,
etc.
then I will choose to save when young
ie. I will recognise imperative 2. to be valid for me.

The maxim is a subjective principle but the imperatives are valid 'objectively' in that they must govern my choice insofar as reason determines my will. But the imperatives are 'hypothetical' because they only apply given various subjective conditions - viz a,b,c,d etc..

Provided I do not recognise other conflicting imperatives as valid for me then, insofar as reason determines the will, action will follow. The resulting action is appropriate if in fact I am providing for my old age, if I am in fact saving. Such an action has 'prudential worth' **because of what it performs or effects.** That is, the action is appropriate and has worth if I **am** achieving the end specified in the maxim or precept; if my action is, **in fact**, in accordance with the imperative. This is so because if it does not so accord then I will not have achieved what I wished to achieve; and the maxim was valid for me **only because** the end it specified was something I wished to achieve.

All this is contrasted with what is the case if a practical principle is an objective principle (a practical law). In this case the principle is valid of itself and not given any subjective conditions. That is, insofar as reason determines my will then I would (if this principle were a law) choose to provide for my old age. I would recognise imperative 1. to be valid for me **irrespective of any subjective conditions.** Provided I do not recognise other conflicting imperatives as valid for me (provided subjective causes do not hinder my action)

then, insofar as reason determines the will, action will follow.

But now, "only the volition is completely determined" by a moral law. The specific nature of the action that follows **will depend** upon subjective conditions. If providing for my old age were a moral duty, and I recognised it as such, then what I actually did would depend upon such things as my belief that I cannot rely on rich friends. We can begin to trace the same picture of the recognition of other imperatives as was traced in the case of a subjective principle. Subjective conditions are, once more, relevant.

However, although the particular nature of the action (what it performs or effects) will depend upon the circumstances and upon my beliefs; the moral worth of the action depends entirely upon the nature of the motivation which gives rise to it. The action has moral worth if and only if it stems from my recognition of duty. Kant's view is **radically** non-consequentialist.

The view of prudential worth of an action, in which worth is dependent upon outcome, is clearly contrasted with the view of moral worth of an action, in which worth is dependent only upon motivation. That contrast rests upon a contrast between a will which is determined by a principle because of a desire to achieve the end stated in the principle, and a will which is determined only because it recognises the principle to be law.

For Kant, the recognition of a principle as a law concerns the 'form' of the principle; it does not concern the worth of the end which is stated in the principle (the 'matter' of the principle). Kant (1788 p.35) explicitly rejects the idea (which he attributes to all other moral theories) that a moral law could be such because of the moral worth of the end stated in that law.

"If it were, the maxim could not be presented as giving universal law, because then the expectation of the existence of the object would be the determining cause of the choice, the dependence of the faculty of desire on the existence of some thing would have to be made basic to the volition, and this dependence would have to be sought out in empirical conditions".

If my choice were determined by a recognition of the worth of the end stated in a law then, according to Kant (1788 p.19), that would involve it being the case that I conceived the end, I wanted the end, and I sought its realisation. But, crucially, we have no grounds for attributing such a want to all rational beings - **not even as rational**. We would have, therefore, no grounds for claiming that the principle was objectively necessary or that the imperative was categorical.

What we **can** attribute to all rational beings, as such, is an ability to recognise whether a maxim does or does not have the **form** of law. That, says Kant, is a case of determining whether one can rationally will that the maxim of one's action "should become a universal law". The maxim has the form of law if we can (as rational beings) think of it as a law. The ability to distinguish maxims in this way is an ability which cannot be denied to any rational being. (Whether such an ability will serve to distinguish what is required is, of course, another matter.)

[Here we can see the similarities with Hare's rejection of a moral theory which is based upon substantive moral conviction. We do not all share such convictions - not even as rational. We do share a use of certain words and concepts. What we can attribute to all rational beings, as such, is an ability to understand the language which they share. The logic of that language demands that, as rational, we universalise our prescriptive moral

judgments. Thus Hare's theory, like Kant's, rests firmly upon the role of universalisation in the moral thinking of a rational agent.]

The central question in moral appraisal, for Kant, is whether the rational being has acted merely because he has recognised the universal validity of his maxim and therefore recognised its validity for him as rational being. If choice is determined by the recognition of moral law, without reference to anything one might wish to achieve, then "the law directly determines the will; [and] action in accordance with it is in itself good" (1788 p.64). "What is essential in the moral worth of actions is that the moral law should directly determine the will." (1788 p.74).

If this is so, it cannot be the case that if the action does not in fact accord with the law then it does not have moral worth, nor can it be the case that its moral worth is diminished. The source of the action is all that matters for its moral appraisal; the achievement of the end stated in the principle, what is the case, what the action actually entails in the circumstances, what are the consequences of the action (in the broadest sense) is irrelevant.

It has to be said that some of what Kant says does appear to conflict with this radical non-consequentialism. For example, he says that the highest worth of humans lies in their intentions and **not in their actions only** (1788 p.74). The fact that, when discussing moral worth, he is willing to refer at all to what is done, as well as why it is done, seems to imply that actions could have moral worth independently of the motive which lay behind them.

It might seem reasonable to argue that if moral worth is in some way due to laws which require us to promote certain ends, then the achievement of the end specified

by the law must have some moral worth. But this, I think, would be to miss Kant's central point, which is that moral worth is entirely a matter of whether the will is determined by a maxim which is a law. If it were the case that the end itself (albeit derivatively) had moral worth then it might be argued that an action which achieved that end would have moral worth irrespective of whether it were done out of respect for moral law - but this Kant explicitly rejects. "It is of the utmost importance in all moral judging to pay strictest attention to the subjective principle of every maxim, so that all the morality of actions may be placed in their necessity from duty and from respect for the law, and not from love or leaning toward that which the action is to produce." (1788 p.84); "there are many minds so sympathetically constituted that .. they find pleasure in spreading joy around them .. but I maintain that .. an action of this kind, however proper, however amiable it may be, has nevertheless no true moral worth" (1785 p.16).

This radical view is tempered somewhat by the assertion that such actions would be honourable and, even, deserving of praise and encouragement. But, says Kant, the inclination to act in such a way does not warrant our esteem and the actions have no moral worth (such actions have mere 'legality').

Thus (as was stated earlier) an action which stems from respect for moral law but is, in fact, not in accordance with moral law has undiminished moral worth; and (as stated here) an action which is in accordance with moral law but does not stem from respect for moral law has no moral worth.

Now Sullivan (1989) believes that Kant simply fails to make proper use of the distinction between the moral worth of the action and the moral worth of the agent.

Moral thinking, moral motivation and moral worth.

According to Sullivan, Kant could (and should) have said that an **action** has moral worth (morality) if it accords with a moral law (if it has legality) and that the **agent** has moral worth insofar as he is motivated by moral law. As Sullivan (1989 p.29) says, the two questions 'What makes an action morally right?' and 'What makes an agent morally good?' "can be considered to be logically distinct"; this "is shown by the fact that we can conceive of a morally evil person performing a morally acceptable action .. [and] we also can conceive of a morally good person mistakenly believing that he is acting rightly when in fact he is not".

Sullivan maintains that Kant fails to clearly distinguish the two questions. That lack of clarity is, according to Sullivan (1989 p.30), partly due to the fact that Kant "often thinks of an 'action' as a person's intention rather than as a person's behaviour and, when he does, he describes actions so as to include the agent's end and motive". Whether or not that is so, we can, surely, refer to the agent's motive when describing **the action**. Just as we can say: 'that action involved this behaviour', 'that action brought about these consequences'; so too we can say: 'that action arose from this motive'. Having said that, we can say (with the consequentialist): 'that action has moral worth because **it** brought about these consequences'; or (with the non-consequentialist) 'that action has moral worth because **it** arose from this motive'. The latter statement need not be due to a lack of clarity.

Kant does answer the two questions differently: an action is morally right (has legality) when it is according to law; an agent is morally good when he is motivated by law. But Kant explicitly denies that the **moral rightness** (legality) of an action entails (or is equivalent to) the **moral worth** of the action. There is a third question: 'What makes an action morally good?' and the answer to

that question need not be the same as the answer to the question: 'What makes an action morally right?'. Kant insists that an action which is according to law but is not motivated by respect for the law may be morally right but it has no moral worth. An action which is according to moral law is (simply as such) not **morally better** than an action which is **not** according to law because the legality or illegality of an action has nothing to do with the moral worth of that action.

Morality, for Kant, is about the quality of the agent's motivation. Actions have moral worth, have morality, only **derivatively**: they have moral worth only insofar as they stem from the agent's respect for the moral law.

Furthermore, in insisting upon a distinction between the moral rightness of an action and the moral worth of an action Kant is, I believe, expressing a moral view which many share. When we consider an action simply as 'the person's behaviour', or as bringing about such and such consequences, we do not consider its moral worth; in order to make a moral evaluation of an action we need to know more.

Suppose: dentists have two types of anaesthetics - A and B; when A is used I still suffer some pain and I feel sick later. Suppose: the morally right thing to do is to use B when giving me dental treatment. If the dentist does indeed use B when treating me does his action have moral worth? According to the moral view now outlined, we need to know more. Suppose the dentist:

a. did not know about the effects upon me of using A, meant to use A, but picked up B by mistake;
b. had used A with a succession of patients, had run out of A, and had to use B;
c. had obtained a special discount when purchasing B, and was using it with all patients in order to save money.

In each case the dentist's action is morally right; but in each case, according to the view outlined, the action has no moral worth.

A world in which people acted unintentionally, or because they could not do otherwise, or simply out of greed and selfishness would be a world without moral worth. This would be true **even if** such people always did (or happened to do) the morally right thing.

According to this view, the language of morality is used to mark out something distinctive and fundamentally important about our approach to decision making. The use of that language in relation to actions simply as 'the person's behaviour', or as bringing about such and such consequences, is entirely derivative. Such a view is not adequately expressed in terms of a distinction between the moral worth of actions and outcomes, and the moral worth of agents and motives. For, according to this view, actions and outcomes (simply as such) do not have moral worth.

[We can draw a parallel with words such as 'shrewd'. We may speak of a 'shrewd move' but we mean by that (something like) 'a move resulting from the use of intelligence and foresight'. If we use these expressions in this way then we may say of a checkmate move in chess that it was not a shrewd move because, say, the move was made accidentally or the move was made in order to make a pleasing pattern of pieces. Some might, alternatively, use the expression 'shrewd move' to refer to a very successful move or a move which is such as a shrewd person would perform. Such a person might then say to us: '**You** are confusing the shrewdness of a move with the shrewdness of a person.'. But such a remark would show a misunderstanding of the way in which we use such expressions - to mark out something distinctive about our approach to problem solving.]

Moral thinking, moral motivation and moral worth.

Radical non-consequentialism of this sort is, I believe, central to Kant's view of morality. Like other aspects of that view it stems from Kant's assessment of 'ordinary moral consciousness'. As Sullivan says (1989 p.19), Kant offers an analysis of the features and presuppositions of ordinary moral consciousness and ordinary moral reasoning. But it is his response to the problem of 'freedom' which, I believe, largely determines the particular nature of his approach to that analysis.

Morality and freedom.

Sullivan (1989 p.76) states that "because the new Newtonian science regarded the natural world of which human beings are part as completely governed by inexorable causal laws, it inevitably generated moral scepticism. In a causally determined world the notion of 'ought' can have no meaning.". Kant (1781 p.218) himself maintains that "Everything that happens, ie. begins to be, presupposes something upon which it follows according to a rule.", "All alterations take place in conformity with the law of connection of cause and effect." and he goes on to offer a proof of such universal causation in the Second Analogy. In the context of his moral theory he has thus set himself the task of giving an account of freedom (and hence of the possibility of morality) which is compatible with this belief that every event (in the phenomenal world) has a cause.

Walker (1978 p.136) says that Kant believes that "it is a condition of being morally accountable that one's will be free, in a sense incompatible with its being determined by antecedent causes. If I belonged only to the phenomenal world I could not therefore be free, or morally responsible.". We might add that, for Kant, this is because it is the case that **as** belonging to the

phenomenal world I am not free and cannot therefore be morally responsible; **as** an event (a wholly determined event) in the phenomenal world my action cannot be morally appraised, the question of its moral worth does not arise.

Kant begins his consideration of freedom when he outlines the thesis and antithesis which make up the third conflict of the transcendental ideas. If freedom, Kant says (1781 p.411), were a feature of the world of phenomena then it would be mere lawlessness; for if freedom had laws and yet were a feature of the phenomenal world then those laws would be laws of nature and not laws of freedom - freedom "would simply be nature under a different name". If, however, freedom were lawlessness and a feature of the world of phenomena then nature (as an ordered system and as thus distinguishable from dreams) would be "hardly thinkable; the influences of [lawless freedom] .. would so unceasingly alter the laws of [nature] .. that the appearances which in their natural course are regular and uniform would be reduced to disorder and incoherence" (1781 p.414). Therefore there can be no freedom in the world of phenomena and everything in that world must take place entirely according to the laws of nature.

Thus one half of the antithesis within the third antimony is, according to Kant, demonstrated. But this does not imply that there is no freedom; it may yet be the case, Kant claims (1781 p.466-7), that phenomena may be grounded in freedom even though as events in the world of sense they are entirely determined by the laws of nature. We can conceive of this possibility only if we regard phenomena not as having absolute reality but rather as the appearances of things in themselves. If phenomena "are not things in themselves then they must rest upon a transcendental object which determines them as mere representations"; and if this is so then there is nothing

to prevent our attributing to that transcendental object, besides the quality through which it becomes phenomenal, a causality which is also not phenomenal.

There would, according to Kant, be no contradiction in thus viewing one and the same event as being in one aspect merely an effect of nature and in another aspect an effect due to freedom. But even if there were no contradiction involved in such an idea, we would not yet have been given any grounds for believing that there is such freedom, or even for believing that anyone other than Kant entertains such an idea of freedom.

However we all do in fact, Kant claims (1781 p.472), entertain the idea that there is a form of causality other than that which we observe in the world of the senses. We know ourselves not only through our senses, as phenomena, but also as possessing reason and understanding; and we do believe that our reason possesses causality. When a man tells a malicious lie that act is **entirely determined** by his education, the society he keeps, the dispositions of his nature, and the occasioning causes at the time; yet he is nevertheless blamed. This, according to Kant (1781 p.477), is because we regard reason as being completely free, "we regard reason as a cause that irrespective of all the above-mentioned empirical conditions could have determined, and ought to have determined, the agent to act otherwise". The act is imputed entirely to a fault (or to the default) of reason.

Such an imputation implies that we imagine that reason does not belong to the series of sensuous conditions, it lies outside the phenomenal world but yet is capable of determining events in that world - and when it does so determine such events it does so according to the laws which it apprehends. Thus, if freedom is to be found at

all, it is to be found in those actions which are determined by the apprehension of law.

The arguments in the Critique of Pure Reason do not, however, establish the reality of freedom. All that Kant claims to have shown thus far is that "causality through freedom is at least not incompatible with nature" (1781 p.479).

When Kant turns to his moral theory he offers a different type of example taken from our experience of morality. That example, Kant claims, not only gives grounds for believing in the objective reality of freedom but also corroborates his claims at to what the nature of that freedom must be.

Moral experience and moral education.

What we discover, when we consider our experience of morality, is that it is **possible** to choose to act in a certain way **despite the fact** that such an action would lead to something we wanted to avoid **more than** we wanted anything else. Ask someone, for example, "whether he thinks it would be possible for him to overcome his love of life, however great it may be, if his sovereign threatened him with .. sudden death unless he made a false deposition against an honourable man whom the ruler wished to destroy", "that it would be possible for him he would certainly admit without hesitation" (1788 p.30).

According to Kant, here we have someone recognising he is free **because** he knows he could do something simply because he ought, and regardless of what he wants. This glimpse of freedom takes us outside the phenomenal world of outcomes and inclinations. Within the context of inclinations we cannot imagine deliberately choosing to act in a way which led to something we wanted to avoid,

unless it also led to something we wanted more. The experience encapsulated in this example shows us something distinctive and that has to do with our apprehension of moral law.

This example and Kant's interpretation of it play, I believe, a central role not only in his moral theory but in the wider theory presented in the Critique of Pure Reason. It is this example (and others like it) which shows us that pure reason can be practical - that it is at least possible for it to move us to action - and "with the pure practical faculty of reason, the reality of transcendental freedom is .. confirmed", and the concept of freedom "is the keystone of the whole architecture of the system of pure reason and even of speculative reason" (1788 p.3).

In insisting that we can do what we do not want to do (even what we least want to do) and that we can resist doing what we want to do (even what we most want to do) Kant is, I believe, expressing a view of moral experience which many share. The belief that we can in some way overcome ourselves seems to me to be central to our experience of morality. But Kant's attempt to analyse and illuminate this belief is, once again, inextricably linked with his response to the problem of freedom.

What Kant needs to claim is that here we have the possibility of a choice which is made regardless of what is the case in the phenomenal world - regardless of our circumstances, history, preferences and inclinations. It is a choice which has its source outside of the world of empirical conditions, and can therefore be free of the causal links which bind every aspect of that world. But this is not a mere negative freedom from physical necessity - that would simply be a lawlessness, or arbitrariness, which would bring us no closer to solving the problem of free will and moral responsibility -

rather the example shows a choice determined in a different way by laws of a peculiar kind (1785 p.63). Our choices can be determined simply by the laws which reason presents to us.

The type of example of our moral experience which Kant here offers shows, he claims, not only the reality of our freedom but also the nature of that freedom - our freedom lies in our responding to the laws of morality as opposed to the laws of nature.

[Or, alternatively: our freedom lies in our **ability to** respond to the laws of morality as opposed to the laws of nature. The difficulties in establishing which of these claims Kant is making, and the link with problems as to the role of 'blame' in Kant's theory, will be briefly discussed in the next section.]

Our freedom, and hence our morality, lies in our response to the laws of morality. Insofar as we are motivated by our recognition of such laws **we** have moral worth; insofar as our actions stem from such recognition **they** have moral worth. If we and our actions are to have moral worth then we must acquire the ability to recognise and respond to those laws. To simply act according to law (to do right) is not enough. To do right because, say, we believe it is in our own interest, or we wish to please those in authority, or we expect reward and fear punishment would achieve nothing of **moral** worth.

Thus the **moral** education of an individual must involve two aspects. First, the individual must acquire the ability to determine whether a maxim of action does or does not have the form of law; that is, the rationality of the individual must be developed. That ability will enable the individual to apprehend some maxims as law. The acquisition of that ability demands, as Sullivan says

(1989 p.287), the renunciation of dependence on external authority and a willingness to think for oneself.

But that is not sufficient; for we are also subject to inclination and those inclinations, despite our apprehension of law, may determine our action. So, secondly, as Kant says (1788 p.156) the individual must become receptive to a pure moral interest. If this receptivity is developed then the apprehension of law will prove stronger than "all allurements arising from enjoyment and ... all threats of pain and harm". Each must develop this receptivity not only to avoid acting wrongly but also to ensure that right action stems not from inclination but from apprehension of moral law.

Such receptivity is to be developed by exposure to examples of a pure moral interest; examples in which we see actions performed with no regard to inclination. If we are shown such examples then we will admire them and wish to emulate them (1788 p.160); and through them we will become conscious of the possibility of freedom from "the impetuous importunity of inclinations" (1788 p.165).

Moral education, therefore, involves holding the pupil's attention to the consciousness of freedom. There can be no freedom, no morality and no moral worth insofar as our actions are motivated by our inclinations.

It is against this background that we can, perhaps, interpret Kant's analysis of a third central feature of morality. This is the belief expressed in the second formula of the Categorical Imperative: "Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end."

Kant is not, I believe, merely expressing the view that others have preferences which ought to be respected and

that we should not, therefore, treat others merely as means to the satisfaction of our own preferences. He is expressing the view that each of us can make choices and act not merely from inclination but against inclination; each of us can respond to moral law and overcome those inclinations; each can, as rational, be free and self-determining; and that to ignore this is to ignore our humanity and to ignore our true worth (our moral worth). We ignore the humanity of others not only when we treat others merely as a means to the satisfaction of our own preferences but also when we treat others merely as a means to morally right action - that is, when we attempt to control and influence behaviour merely through reward, punishment and subjection to authority.

An education which **merely** aimed to control behaviour in such ways, and thus ensure morally right action, would not be a **moral** education.

I have now claimed that there are three features of morality and moral experience which are central to Kant's view:

always treat humanity never simply as means but always as an end;
nothing can be called good without qualification save a good will;
each of us is capable of overcoming the impetuous importunity of our inclinations.

I shall claim that a moral theory which does not give adequate expression to these features is unsatisfactory. Kant's theory illuminates those features but it does so by means of an analysis of freedom - our freedom as involving a response to our apprehension of moral law. But that analysis is beset with problems; both in terms of the means by which we are supposed to apprehend those laws and in terms of the notion of freedom which is involved in the analysis.

Moral thinking, moral motivation and moral worth.

Objections to Kant's theory.

Kant's theory ultimately rests, firstly, upon the coherence of the contrast between actions as stemming from choices made by the self as noumenon and actions which are simply performed by the self as phenomenon; and, secondly, upon its capacity to deliver substantial moral laws by the means described.

A law is a moral law if one can rationally will that it should become a universal law. We discover which maxims are laws simply by attempting to conceive them as laws. Everything must be done by reason alone and so, presumably, we can make no use of any facts about the world. Yet, as Korner says (1955 p.138), it is clear from Kant's examples that this test does not consist merely in replacing the 'I' of the maxim with 'everybody' and then seeing whether the result is or is not **logically self-contradictory.**

Kant (1785 p.39) considers four possible maxims in order to show how the method would eliminate them. The four maxims are:

1. from self-love I will shorten my life when its longer duration is likely to bring more evil than satisfaction;
2. when I think of myself in want of money, I will borrow and promise to repay, although I know that I can never do so;
3. I shall neglect the cultivation of my natural gifts whenever it agrees with my inclination to indulge in pleasure;
4. I will take nothing from anyone or even envy them, but neither will I contribute to other's welfare or assist them when in distress.

Korner points out that if we universalise these maxims then we do not get logical contradictions; so, either Kant needs to extend the notion of contradiction to cover 'moral absurdity', or he must use other true statements about the world in order to derive a contradiction from their conjunction with the universalised maxim.

If he does the former then his test would be circular or superfluous. If he does the latter then some of the statements used will need to be empirical statements whose truth might be doubtful - for example, Kant's discussion of the fourth seems to assume that we are none of us capable of deliberately depriving ourselves of all hope of support in times of need. Korner concludes that Kant's test for moral law is not adequate.

Clearly Kant does need to appeal to something more than the logically self-contradictory nature of the universalised maxim, and clearly he is not able to appeal to a posteriori truths in order to yield a contradiction - for that would be inconsistent with his whole theory. However, he can, without inconsistency, appeal to a priori truths. He could claim, for example, that the statements 'If people expect promises to be broken then they will place no reliance on any statement purporting to be a promise.' and 'No-one can deliberately deprive himself of all aid.' are not only true but also known a priori.

The danger then is that Kant will judge to be a priori just those statements which will assist in deriving the moral 'laws' which he subscribes to. As Walker says (1978 p.158), "What he actually does is to build into the conception of rationality all his substantive moral views".

It is also worth mentioning at this point that even if the required truths are known a priori, and can thus be

used to refute the putative moral laws derived from examples 1 to 4, the derivation of the contrary (eg. never commit suicide) does not straightforwardly follow.

From:

it is not possible to prescribe for all x (x do A),

it does not follow that:

it is necessary to prescribe for all x (x not do A);

we can only deduce that:

it is necessary not to prescribe for all x (x do A).

From the fact (if it is a fact) that 'Everyone should break their promises.' cannot be a moral law, it does not straightforwardly follow that 'Everyone should keep their promises.' must be a law.

It may be true that, given Kant's test, the latter can be a law whereas the former cannot. But are there not other maxims concerning promising which could survive the test? Suppose, as McIntyre suggests (1967 p.198), we take the maxim 'I may break my promises only when ...' and we fill the gap by a description which will apply to my present circumstances but to very few others (eg. the promise concerns some borrowed money, the borrower is a teacher, the amount is fifteen pounds, etc.) then such a maxim seems to pass the test. That test "imposes restrictions only on those insufficiently equipped with ingenuity".

We could object that such a maxim would only be proposed and adopted by someone who wished to avoid the settling of this particular debt (ie. myself). As Sullivan says (1989 p.163) "we may not appeal to the desirability of the possible or probable empirical consequences of the (universal) adoption of a particular kind of action". The proposal of such a maxim as a counter-example is not in tune with Kant's notion that what we are seeking are laws which can determine the choices of rational agents, **simply as rational**. Such an objection would shift the

ground to Kant's claim that there can be such choices, and to a consideration of the coherence of Kant's contrast between actions as stemming from choices made by the self as noumenon and actions which are simply performed by the self as phenomenon.

According to Kant, everything is subject to law. Our actions are either the effect of choices made according to our conception of the moral law and/or they are the effect of empirical conditions which are subject to the laws of physical necessity. Insofar as we are part of the 'world of understanding' our actions can be caused by choices which are not themselves 'caused' by anything else; but as part of the 'world of sense' our actions are caused by choices made in the light of desires and inclinations, and those desires and inclinations are themselves caused by other phenomena, and so on (in an endless causal chain).

We now have two systems which can explain the determination of choice, two systems of laws which must, according to Kant, apply to separate aspects of our nature - self as noumenon and self as phenomenon. Free will has reality only for the self as noumenon. If the self were only a noumenal self then it would always choose in conformity with moral law - it would always be free, ie. its choices would be determined by its apprehension of moral law (1785 p.70). Insofar as the self is also a phenomenal self there is, presumably, a form of conflict between the two systems of determination.

If that conflict results in a choice determined by the conception of law then the choice has moral worth (the will is a good will). If the choice is simply determined by desires, inclinations, and other empirical conditions then it has no moral worth.

An action has moral worth if and only if it stems from a choice which has moral worth. An action has worth only derivatively because an action is only known as a phenomenon and every event in the world of phenomena is causally determined. **As** an event in the world of sense the action is not free - and therefore the question of moral worth does not arise.

But now if freedom, morality and moral worth are all manifested in our responding to the laws of morality, as opposed to the laws of nature, then a conflict which results in a choice determined by the laws of nature not only has no moral worth but also has no freedom. The implications of such a view would seem to present serious difficulties for Kant's moral theory. All wrong actions (and all right actions not determined by recognition of moral law) would lack freedom and would seem, therefore, not to be blameworthy and (as wholly determined) to be not immoral but **amoral**.

Furthermore, it is difficult to see how there can be any **conflict of choice**. The choices of the self as noumenon are determined by recognition of the laws of morality, the choices of the self as phenomenon are determined by the laws of nature. If freedom is just the determination of choice by a recognition of moral law then there is no sense in which I am **free to choose between** duty and inclination.

Walker argues (1978 p.148) that Kant, in his later works, does not identify freedom with obedience to law but rather identifies freedom with the ability to make such a choice between duty and inclination. Such a choice must be made by the self as noumenon (unless we are willing to take the step of suggesting that there is a more ultimate self behind the phenomenal-noumenal viewpoints). But then why would I (as noumenal self) "choose the evil course rather than the good one"? "If I choose the evil

course then I freely decide to let my inclinations have their way". It cannot be the case that my inclinations "are too strong for me and determine my decision, for then I should not be properly free and accountable for my choice."

Both interpretations of Kant's view of freedom present considerable difficulties. Kant's analysis of and response to the problem of freedom, according to Sullivan, may at times seem incoherent and, according to Walker, is a hopeless failure. Certainly that response seems to make it extremely difficult for him to provide, as he intended, an analysis of our ordinary moral views and moral experience.

However, I would not wish to claim that, to paraphrase Williams (1985 p.65) we cannot travel far enough into Kant's territory to bring back central aspects of his moral philosophy without bringing back "the more extravagant metaphysical luggage of the noumenal self". For I would like to develop a moral theory which, like Kant's, is radically non-consequentialist; provides firm links between moral thinking, moral motivation and moral worth; and incorporates the three features of morality and moral experience which I believe to be central to Kant's view.

In the final section of this chapter I would like to make some preliminary remarks relating to the third of those features:

> each of us is capable of overcoming the impetuous importunity of our inclinations.

Choosing to act against inclination.

If it were the case that we could not avoid Kant's interpretation of our **experience of morality** then the

Moral thinking, moral motivation and moral worth.

temptation to subscribe to his moral theory might be greater; but I do not believe that this is the case.

In his interpretation of the example in which we see the possibility of choosing death rather than the making of a false declaration, Kant eventually makes the transition from a claim that here is a choice made against the agent's own inclinations, to a claim that here is a choice made regardless of all empirical conditions, to a claim that here is a choice made by the self as noumenon.

These transitions are themselves made in the light of his wider theory and are part of the quest for a freedom which, Kant says, can find no place in the world of phenomena. The second transition stands or falls with that wider theory but the first can, I think, be approached more directly.

The first point to make about the example is that we could see the claim that it is possible to choose death rather than lying as simply a claim that it is possible to have an inclination not to lie which is stronger than the inclination to preserve one's life. Kant's response would presumably be that this misses the point of the example: the point is that every agent will admit the possibility of such a choice (the same choice) despite the fact that few agents have such a strong inclination against lying. That is: it is possible for me **as I am** not merely possible for me to be other than I am (and thus to make such a choice from inclination). There must therefore be a means by which each of us (as we are) could be brought to such a choice despite lacking that inclination.

Kant sees only one possibility: we apprehend moral law and acknowledge that law as a duty - we have knowledge that such a choice would be in conformity with a moral law which applies to us. Such knowledge would have

nothing to do with what is the case, nothing to do with outcomes or inclinations or any other empirical conditions. Thus the first transition is made. But it relies on the claim that there is no other means by which we could be brought to such a choice - namely one in which we choose to act in a way which leads to something we want to avoid more than we want anything else. This claim seems to me to be false.

Each of us can acquire inclinations and can discount, or give different weight to, the inclinations which we have. We can acquire inclinations through imaginative identification with others - I now want you to have this book because I know how much you want it. We can discount an inclination as a result of a wish to be other than we are (through imaginative identification with an 'ideal self') - I now discount my desire to see you suffer because I wish I were not vengeful. Thus we can deliberate as if we had preferences other than those we in fact have.

In later chapters I shall elaborate further the notions of imaginative identification with an ideal self and benevolent (or malevolent) identification with others. All I wish to claim here is that the process of deliberation can alter one's inclinations and/or can alter the weight which one gives to one's actual inclinations when making decisions. This fact is one way of interpreting the possibility (of choosing to act against my actual inclinations) in such a way that it becomes a genuine possiblity for each of us (as we are). It is possible for each of us to discount, for example, all our preferences save the preference not to lie.

Kant might reject such an approach on the grounds that such acquisition or discounting of inclinations may alter my inclinations and thus determine my judgment, but not every agent will reflect, or see the need to reflect, in

this way (and each agent may not reflect in this way on all occasions). Kant might make the same point here as was made earlier: every agent will admit the possibility of the choice in the example, despite the fact that some agents do not deliberate in this way. That is: it is possible for me **as I am** not merely possible for me to deliberate in a way I do not (and thus to make such a choice from the inclinations I then have).

For Kant, it is simply as rational agents that we recognise the possibility of our making choices regardless of our inclinations. All rational agents, Kant claims (1788 p.30), apprehend moral law and acknowledge that law as duty. The possibility of acting against inclination is simply the possibility of our choosing to act according to those laws we all do apprehend and acknowledge. It is a possibility for each of us as we are - as rational. We recognise that we can act against inclination **because** we know that we ought.

However, if we do not presuppose Kant's metaphysics, then this is merely to claim that this is the only explanation of the possibility of our acting against inclination **and** it does apply to us all. I have denied the former claim and will now deny the latter. We do not all apprehend 'moral law' in the (unsatisfactory) way which Kant suggests nor do we acknowledge as a duty the particular 'moral laws' which Kant acknowledges.

Some do simply have a very strong inclination not to lie; some do deliberate in a way which results in their acquiring or discounting, or giving altered weight to, inclinations; some do (perhaps) apprehend and acknowledge moral law as Kant suggests. Each of these offers a means of interpreting the claim that it is possible for each of us to act against our inclination. None of these applies to us all - not even as rational.

[Each of these might apply to us all as **morally worthy** agents. I shall claim that the second does. We **ought** to acquire the preferences of others through benevolent identification and to discount all malevolent preferences. Morally worthy agents will then sometimes act against (their own) inclination.]

There are, I am claiming, alternative ways of interpreting the belief that we can deliberately act against our inclination - a belief which I, like Kant, take to be central to moral experience. Furthermore, such an interpretation can be incorporated into an alternative moral theory. Kant's interpretation derives, I believe, from his moral theory; it is not the only means of doing justice to that experience.

At the heart of Kant's theory is the claim that:

for all

morality gives reason to act in a certain way.

This 'unqualified' universal quantification would enable Kant to answer Bradley's (1876) question 'Why should I be moral?' by saying that the question would not arise for any rational agent - no defence of morality is required. So too the intuitionist, and the realist, can claim that the question would not arise for any agent capable of clearly intuiting, or perceiving, moral facts - it would arise only for the morally 'blind'.

However, Kant's quantification is not in fact unqualified - it ranges over all '**rational**' agents. So too, the intuitionist's, and realist's, quantification ranges over all those **capable** of clearly intuiting, or perceiving, moral facts.

For the philosopher the interesting question will be 'How far can we push the range of quantification?'. But for the educationalist the pressing question will be 'How do

we ensure that an educatee falls within the range of quantification?'. I claimed (in chapter 2) that the intuitionist, and the realist, fail to give an answer to that question. It is that failure which, I believe, makes it reasonable for us to question whether such a theory has any genuine significance; or whether it merely ensures that the description used in the range of quantification picks out all, and only, those who share the theorist's moral judgments.

Kant does offer an answer to the latter question. In order to fall within Kant's range of quantification we will need to be 'rational'. But that involves apprehension of moral law and acknowlegement of that law as duty. Once again the danger is that the ability to apprehend, and the disposition to acknowledge, such laws will be attributable only to those who share Kant's particular moral views.

The educationalist's question provides a way of determining how far we can push the range of quantification without merely ensuring that the description used in that range picks out all, and only, those who share the theorist's moral judgments.

At the heart of Hare's theory is the claim that:

> **for all those willing and able to make moral judgments**
>
> morality gives reason to act in a certain way.

And Hare can, and does, give a clear answer to the educationalist's question. Furthermore, the link between morality and action is maintained. But it no longer holds for all rational beings; the agent must be **willing and able** to make moral judgments.

If there are, in this way, limits to how far we can push the range of quantification then Bradley's question

becomes more difficult. This too is a question which will deeply concern, and will have a special significance for, the educationalist: 'Can we justify an attempt to ensure that our educatees fall within the range of quantification; that morality determines **their** actions?'.

The question of justifying the imparting of a **willingness** to engage in moral judgment will be examined later. But, according to Hare, the **ability** to make moral judgments has certain **necessary** features; and it is this claim which will be examined now.

CHAPTER 5.

Critical thinking, universalisability and impartiality.

Hare's claims with regard to universalisability.
Mackie's characterisation of universalisability.
Hare's characterisation of universalisability.
The use of universalisability.

Hare's claims with regard to universalisability.

Hare identifies (first level) moral thinking with critical thinking. His argument for that identification and his characterisation of critical thinking turn upon his views with regard to the universalisability of prescriptive moral judgments. Hare claims that:

1. universalisability is a necessary feature of those judgments and that, therefore, a rational agent making such judgments should engage in a form of thinking which reflects that feature;
2. the nature of universalisability is such that it places certain very specific constraints upon a rational agent which are such as to demand that form of thinking which he calls critical thinking;
3. the use of the form of thinking which he describes yields judgments which have a content identical with that of a certain form of Utilitarianism.

As stated in Chapter 2, I too will seek to argue for an identification between moral thinking and critical thinking (somewhat modified). That identification will also form a part of a moral theory which has features in common with Utilitarianism. But, although the judgments yielded by that form of thinking will be Utilitarian, the

view of moral worth will be radically non-consequentialist. The moral theory which I shall propose will focus upon the agent and, as in Kant, the links between moral thinking, moral motivation and moral worth will be central.

My route to that theory will involve an attempt to show that that alternative form of Utilitarianism can underpin and inspire those (moderately reflective) moral views which we share; and that it can do so more satisfactorily than the form which Hare proposes. But if Hare's argument were correct then such a route would not be legitimate. The logic of our moral language dictates the form of thinking which should determine our moral judgments and, Hare claims, those judgments happen to have a content identical with that of a certain form of Utilitarianism. If that is so then to propose an alternative form is to ignore the logic of our language.

In order to reach an alternative we must reject Hare's claims. In this chapter I shall examine Mackie's rejection of Hare's first and second claim (above). I believe that Mackie's criticism fails but that an examination of his criticism will allow a further clarification of the central features of Hare's route to Utilitarianism. In the next chapter I shall then attempt to reject Hare's second and third claims (above).

Mackie's characterisation of universalisability.

Hare's argument has two requirements with respect to universalisability:

1. the universalisation of moral judgments must be a logical requirement and not merely a requirement which is derived from a moral intuition which some do not share;

2. the nature of that requirement must be such that Hare's conclusions can be drawn without appeal to any moral intuitions (eg. intuitions relating to fairness or impartiality).

Mackie (1977) claims that one of these has to be false. So I shall begin with a detailed look at Mackie's characterisation of possible notions of universalisability.

Mackie argues that there are three stages in the universalisation of moral judgments, and that the third stage is necessary to the derivation of Hare's type of utilitarianism. He further argues (1977 p.83) that the thesis that universalisability is a logically necessary characteristic of moral judgment is progressively more dubious as we move through the stages, and that for the third stage it is "plainly false".

He agrees (1977 p.83) with the proposition that moral judgments are universalisable: "Anyone who says, meaning it, that a certain action .. is morally right or wrong .. is thereby committed to taking the same view about any other relevantly similar action". But he says "the key phrase is 'relevantly similar'". "In practice no two cases will ever be exactly alike" so that "universalisability would be trivial and useless, therefore, if we could not rule out many of the inevitable differences as irrelevant". The crucial question, for Mackie, then becomes: 'Which features are, or are not, relevant?'.

In the first stage of universalisation, as characterised by Mackie, it is simple numerical difference which is treated as being not relevant. In making our universalised judgment we cannot specify one individual, or set of individuals; the judgment must apply equally to

all, it cannot be made only for all those situations in which that individual is, or is not, involved.

Mackie is saying that we must specify the circumstances in a way which will give general application (otherwise the universalised judgment would be trivial and useless) but that first stage universalisation does not permit us to use a specification which deliberately picks out an individual or set of individuals.

The universalisation of the judgment 'You ought to look after your aged father.' could legitimately yield:

for all x,y (if x is son of y and y is aged then x look after y)

but it cannot legitimately yield:

for all x,y (if x is not Jeff Wardle of Woodley and x is son of y and y is aged then x look after y)

The test of a moral judgment is then whether we can subscribe to a judgment which has been universalised in a legitimate way.

Such a test may be both an intuitive and a logical requirement. That it is an intuitive requirement could be argued on the basis that it is a requirement which conforms to the intuition that there must be impartiality between individuals (simply as individuals). That it is a logical requirement, Mackie says (1977 P.87), could be argued; but it is dubious "since we can understand as moral the view of the ascetic that something he does not condemn in others would be wrong for him".

But now, this 'stage' of universalisation may require impartiality between individuals (as such) but it does not require impartiality between types of individuals (according to qualitative differences such as those of sex, race, resources, ability etc). It allows the

derivation of principles which can be 'unfair' in many different ways. For example, it does not exclude:

for all x,y (if x is employer and y is employee and y is not a woman then x pay y a decent wage).

There are two different responses which Mackie envisages (although he does not clearly separate them). The first is to extend the notion of what cannot be relevant, what cannot be specified when we specify circumstances. This response would require us to say not only that universalisation does not permit us to use a specification which deliberately picks out an individual or set of individuals but also that it does not permit us to use a specification which deliberately picks out a type of individual (according to nationality, gender, race, etc.).

The problem then is that although this requirement would disallow principles which are unfair because they discriminate **against** certain groups, it would also disallow principles which (in the interests of fairness) discriminate **in favour of** certain groups. For example, it would allow:

for all x (x bear the full cost of his housing and medical treatment);

but it would disallow:

for all x (if x is not poor then x bear the full cost of his housing and medical treatment).

The second response is not just in terms of what **is not** relevant, but in terms of what **is** relevant. Mackie responds in this way and his second stage adds to the requirements of the first stage the requirement that we determine which principles we would subscribe to regardless of changes in the mental and physical qualities, resources, and social status of individuals. That is, we must envisage the alteration of those

features (with regard to ourselves and others) in order to determine what principles we can subscribe to. The specifications which are relevant are those which "look relevant from whichever side you consider them" (1977 p.91/92). The specification in the last principle (above) would then be relevant if I could subscribe to that principle after, for example, envisaging myself as rich and then as poor.

"The judgments that result will not, then, take unfair account of one's own special abilities or resources or social position, or of one's interests in so far as they are determined by these" (1977 p.92).

Such a test, says Mackie, may correspond with a generally used form of argument: 'How would you like it if .. ?'; but it does not seem to be a logical requirement - "it does not seem that moral terms are being misused if they are employed in judgments which are adhered to only because such challenges are brushed aside" (1977 p.96).

Furthermore, this second stage may require impartiality between individuals (as such), and between types of individuals (in terms of qualitative differences), but it does not require impartiality between those having different tastes and ideals. It allows the derivation of principles which can be unfair, for example, to those groups whose interests do not coincide with our own.

Therefore, Mackie argues, we have not yet reached a utilitarian view, because the utilitarian demands that **we take equal account of all actual interests.** If we are to achieve the sort of impartiality which utilitarianism demands then we must extend the second stage so that it requires us to determine which principles we would subscribe to regardless of changes in the desires, tastes, preferences and ideals of individuals. That is, we must envisage the alteration of those features (with

regard to ourselves and others) in order to determine what principles we can subscribe to.

But now, Mackie says, such a requirement is clearly not a logical requirement - we are not constrained "under penalty of being said not to be thinking morally or evaluatively, to give equal weight to all ideals, or even to respect ideals that we do not share" (1977 p.96). Nor is it an intuitive requirement: moral judgments commonly include a claim to objectivity and it is "all too easy to believe that the objective validity of one's own ideals provides an overwhelmingly strong reason for taking no account at all of ideals that conflict with them" (1977 p.97).

Mackie concludes that only the first stage of universalisation (at most) could be said to be a logical requirement, and it falls far short of yielding the consequence that moral thought accords equal weight to the interests of all persons. If Hare relies only on logical requirements then he cannot reach the desired conclusion; if he is to reach that conclusion he must appeal to notions of impartiality and fairness.

Hare's characterisation of universalisability.

Hare says (1981 p.108): "I wish to stress that there are not .. different stages of universalisation. Moral judgments are, I claim, universalisable in only one sense, namely that they entail identical judgments about all cases identical in their universal properties".

For Hare, universalisability amounts to the claim that it is, for example, contradictory to say that:

> "Jack did just the same as Jim, in just the same circumstances, and they are just the same sort of

> people, but Jack did what he ought and Jim did what he ought not." (1981 p.81)

This requirement is similar to what others have called the requirement of 'moral consistency' - the principle that if two situations are identical in respect of their 'non-moral' features then we have to give the same moral judgment in each case (see S.Blackburn 1971). It also corresponds directly with Mackie's first-stage universalisation.

However, Mackie says that universalisation would be "trivial and useless" if we did not replace 'identical in all universal properties' with 'identical in all **relevant** universal properties'. His reason for saying this is the fact that in practice no two cases will be exactly alike. This is undoubtedly true but what is its significance? That fact would only be relevant if we were attempting to reach a judgment which applied to more than one case. If we are merely trying to reach a moral judgment in respect of a particular situation then that fact is not relevant.

Mackie's mistake (with regard to his interpretation of Hare's argument) is his assumption that the requirement of universalisalisation is being offered as a direct means of generating and testing **general principles.** If this were so, and if the resulting principle were not general (did not apply to more than one situation) then it would be useless. But this is not so.

Hare (1963) does refer to 'relevant similarity' when discussing terms which have descriptive meaning and the rational constraints which apply to their use. In that context, similarity in relevant respects - the respects which govern the use of the term - is of central importance. But that is because we are unlikely to have much use for a descriptive term which does not have general application (although, of course, we may use such

terms to construct descriptive expressions which in fact only have application to a particular case).

Likewise, when discussing commendation of, say, motor-cars, he says (1952 p.129) that "the implication of the judgment 'That is a good motor-car' does not extend merely to motor-cars **exactly** like that one .. [since if] this were so , the implication would be for practical purposes useless; for nothing is exactly like anything else". The commendation extends to every motor-car that is like that one in the relevant respects - the respects for which I was commending it. But that is because in commending such things we are, typically, applying a standard for judging motor-cars in general.

In Moral Thinking the appeal is to exact similarity (1981 p.63). Universalisation, in conjunction with relevant or exact similarity, may be used to generate principles with general or particular application, but in critical thinking it is exact similarity which matters. Such thinking **is** useless as a means of guiding future choices, but then that is not its purpose; rather it is intended to be a means of determining choice in a particular situation - the one which confronts us.

The requirement of universalisation means that we cannot (logically cannot) subscribe to a **moral** judgment made in a specific situation with respect to a particular person unless we also subscribe to an 'equivalent' judgment which is quantified over agents. For example, we cannot subscribe to:

in situation S, Jim **ought** to do A,

unless we also subscribe to:

for **all** x (in situation S, x ought to do A).

If 'S' is a general description of a type of situation then the universal judgment will be a general judgment

which applies in all situations of that type. It may then be useful as a general guide to behaviour.

If 'S' refers to a particular situation (or is a descriptive expression such as 'a situation exactly like this one in all its universal properties') then the universal judgment will be a specific judgment which applies (in fact) only to that actual situation. It will not then be useful as a general guide to behaviour.

But Hare's aim is not to generate and test general principles, it is to test specific moral judgments. So the fact that a universalised judgment may or may not be general is not a fault in the characterisation of universalisation requiring rectification. We do not, therefore, have to "rule out many of the inevitable differences as irrelevant" and are not forced into the stages which Mackie describes. That is not the way the argument proceeds (as we shall see in the next section).

In fact, Mackie's analysis not only misrepresents Hare's argument it is also misleading. It implies that the generation of the three stages stems from the need to rule out as irrelevant some of the aspects of a situation in order to achieve generality. But this is not the way in which Mackie, himself, arrives at those stages. Apart from the initial appeal to the irrelevance of numerical difference (that it is this person rather than that person - simply as such), the appeal throughout is to notions of 'fairness' and 'impartiality' which supposedly require us to imagine ourselves with different qualities and different outlooks. We need to do this in order to avoid the generation of principles which are unfair to this or that group or individual. Mackie presents this requirement as if it were simply tacked on to the requirement with regard to numerical difference. It is not surprising, therefore, that he finds it so easy to unpick the stitching.

The use of universalisability.

The requirement of universalisation **can** be used to generate the sort of principle which Mackie has in mind. For example, consider:

> Jim has an aged father, so he ought to look after him.

If that 'ought' is a moral ought then:

> for all x (if x has an aged father then x ought to look after him).

From this we derive:

> if I have an aged father then I ought to look after him.

Universalisation can thus be used to generate a requirement that I consider the implications of a general universalised principle.

But it can also be used, Hare believes, to generate requirements by means of very particular judgments, and it is here that critical thinking begins.

For example:

> Jim, in his present situation S, ought to look after his aged father George.

If that 'ought' is a moral ought then:

> for all x,y (if x is in a situation identical to S and y is the aged father of x, then x ought to look after y).

From this we derive:

> if I were in a situation identical to S and George were my aged father, then I ought to look after George;

and:

> if Jim were in a situation identical to S and I were his aged father, then Jim ought to look after me.

If I judge that Jim, in his situation, ought to look after his father then, if that is a moral judgment, I am also committed to the two derived judgments. These judgments are prescriptions for (non-actual) situations in which I am involved and which are identical in all other respects to the actual situation in which Jim and his father find themselves. I am committed to making these judgments so, insofar as I am rational, I will (Hare argues) consider how it would be for me if those actions were performed in those situations.

It is first-stage universalisation which is thus used to generate a requirement that I consider the implications of a non-general universalised principle. This is the beginning of critical thinking.

We have here a requirement that I consider **how it would be for me.** In form this involves just the same appeal to self-interest as is made in decisions of prudence. But this is **not** an appeal to self-interest in the way which is often intended when we say 'How would you like it if..?'. When this is said, we often mean to appeal to **actual** self-interest. We argue thus: one day you may be old and in need of care, if Sam ought not to have to look after his aged father, then your children ought not to have to look after you, and how will you feel when you are old and neglected? Here we are appealing to the universalisability of a moral judgment in order to generate a **general** principle which **is** likely to have application to our listener's circumstances. It is open to the listener to say: '**I** won't get old.' or '**I** won't care if I'm neglected.'. It is precisely this sort of response which Kant is trying to avoid when he makes his appeal to the use of 'pure' reason - the listener's actual (or likely) circumstances and preferences are not relevant.

There are other arguments which use the generation of general principles in order to appeal to actual (or likely) circumstances and preferences. For example: 'What would it be like if we all did that?'. This also appeals to our preferences with regard to likely effects in the circumstances in which we live. Such forms of argument are standard forms of moral argument and may be used to make an appeal in terms of **actual** self-interest. These arguments may well be widely used but, as Mackie points out (and Hare would agree), appeal to them is no part of the meanings of moral terms.

The requirement for critical thinking is not derived in this way and does not involve an appeal to **actual** self-interest - I will never be George's son, and I will never be Sam's father. In considering how it would be for me, I am considering how it **is** for them because I am considering how it would be for me if I were them - but that is not something which has any likelihood of happening. The aim is to reach a moral judgment only for this actual situation, I am not thereby committed to a general principle which will apply to other actual situations. The appeal is not therefore to actual self-interest, it is merely an appeal to what is required of any rational agent who understands the logical nature of moral judgments.

We can further contrast this view with that of Mackie by considering another example.

Suppose a situation in which Claire, who is strong etc, is walking along and meets, on a narrow path, Sue, who is weak etc; and a proposed judgment which yields the prescription that Claire should push Sue aside. Now suppose another situation, a logically possible hypothetical situation, in which Sue is Claire and Claire is Sue. Sue, who is now strong etc, is walking along an identical path and meets Claire, who is now weak etc.

Then, assuming only first stage universalisation, the original judgment must, **if moral**, now yield the prescription that Sue should push Claire aside. But now, if the original judgment was a moral judgment and was proposed by Claire, then Claire is committed to the same judgment for that hypothetical situation which is identical in all its universal properties and, therefore (Hare argues), rationality requires that she consider the consequences of the proposed action in the hypothetical situation - a situation in which she is weak and pushed aside. Rationality requires that a judgment be made only after a full consideration of the facts, and these include what it would be like to be weak and pushed aside.

Claire, if rational, has to consider what it would be like to be Sue - what it would be like to be the weak party in this situation. But this is not because the strength or weakness of the parties concerned cannot be relevant to a moral judgment. It is not, therefore, because second stage universalisation demands that she test her judgment against a change in such qualities. Rather it is because first stage universalisation, together with the requirement of rational consideration of hypothetical situations, demands that she considers herself with these qualities.

If such consideration leads her to abandon the proposed judgment, then that does not mean that it was not a possible 'moral' judgment (because, in Mackie's terms, it turns out to have been proposed on the basis of Claire's possession of strength). Rather the assumption throughout is that the judgment **is** a moral judgment. It is the fact that it is a moral judgment which implies that it must be made with respect to all identical situations, including hypothetical situations; and this (it is claimed) implies that we must (insofar as we are rational) consider what such situations would be like;

and this hopefully ensures that (after a full consideration) we will abandon the judgment.

Hare's position relies on such a judgment being a moral judgment, otherwise such consideration would not be necessary. It is, for Hare, not a logical requirement that, for example, differences in strength be irrelevant to moral judgments; such irrelevance is (or may be) a substantial result of the sort of critical thinking which he is proposing - it is not built into the meaning of moral terms. Simple irrelevance of numerical difference, together with Hare's characterisation of the demands of rationality (and what that implies for the 'consideration' of situations), is what is doing the work.

We are required, Hare says (1981 p.221), "for the sake of rationality, to ascertain the facts, including facts about others' preferences" otherwise "our final moral judgment will be irrational". Such facts are made relevant because our prescription is universal. It applies to those hypothetical situations in which I occupy the role of the other person, and in which I have the qualities of that other person - it applies to all situations which are identical in all their universal descriptive properties. The 'rationality' requirement is the workhorse which, on the assumption of first stage universalisation only, eliminates 'unfairness'.

I have here tried to expound this part of Hare's argument in a way which makes it clear that it turns upon his claims concerning rationality. Hare's claims with regard to the nature of universalisability are minimal and would be generally accepted. The substantial claim which Hare is making concerns the applicability of the constraints of rationality. He is claiming that the sort of constraints, which apply to a rational agent considering the consequences of alternative actions in a situation

which will happen or may happen, also apply when that agent is considering a situation which will not happen - for example, a situation in which I am George's son. This claim, I shall argue, can be rejected directly and without appeal to a moral theory.

CHAPTER 6.

Rejection of Hare's position on logical requirements.

Critical thinking is not a logical requirement.

Hare's argument for identifying moral thinking with critical thinking involves an analysis of our use of moral language. Nagel (1982) says that Hare's analysis of moral language cannot be right. Firstly because "many people regard criticism of their moral views by this method as invalid" and, secondly, because even those who agree with Hare's moral position "would not regard those who reach moral views by a different method as misusing language". But that is merely to say that Hare's analysis and argument are wrong because others do not agree with it.

I shall argue that Hare's argument is not sound; but if it were sound then those who did not see the necessity for critical thinking as the means to making moral judgments would be failing to understand the implications of the logic of moral talk. We cannot, surely, simply assume that all those who have a different view as to the logic of moral talk have a clear and adequate grasp of that logic. Hare's work is an enquiry into that logic and an elaboration of the implications of the results of that enquiry. It is an attempt to establish a link

between substantive moral views and the form of reasoning which, Hare believes, a careful analysis of the logic of moral words shows to be necessary.

That analysis involves the claim that the sort of constraints which apply to a rational agent considering the consequences of an action in an actual situation (or a situation which may occur) also apply when that agent is considering a situation which will not occur. If objections I shall raise to that claim are correct then critical thinking is not a **logical** requirement of moral thinking.

In the chapter on universalisability I said that the requirement for critical thinking does not involve an appeal to **actual** self-interest - I will never be George's son, and I will never be Sam's father. In considering how it would be for me, I am indeed considering how it **is** for them; but I am doing so in terms of how it would be for me in a situation which will not occur.

When I say that in this situation Sam ought (morally ought) to look after his father George, I am (if universalisability is a logical requirement of the use of moral expressions) committed to making the same judgment for all identical situations and, therefore, to saying that: if I were Sam then I ought to look after George.

I am committed to a prescription for that situation which is like the actual situation in all respects save that I am Sam. This much is, **perhaps**, incontestable. Universalisability in this minimal sense may be a logical requirement of the way in which we (happen to) use moral expressions. As Hare says (1981 p.113) the moral judgment commits me to a moral principle and that principle applies to the hypothetical situation.

Rejection of Hare's position on logical requirements.

Hare must now go on to say not only that I am committed to that prescription but that I must, insofar as I am rational, consider what it would be like for me (as Sam) if that action were performed in this situation. Presumably I must do this because, as a rational agent, I do not make prescriptions without considering 'how it would be'. I am committed to a prescription for a situation in which I am Sam therefore I, as rational, consider my preferences for the consequences of acting according to that prescription in that hypothetical situation.

But, if I make a universalised prescription which logically entails prescriptions for non-actual and totally hypothetical situations, am I, simply as a rational agent, required to consider my preferences for the consequences of alternative actions in such situations and to modify or reaffirm the original prescription accordingly?

If I will never be Sam then why should it matter what prescription I make (or **implicitly** make by virtue of the logic of my language) for the situation in which I am Sam? If universalisability **is** a feature of the logic of moral language then it is true that when I say that Sam **ought** to look after George I am committed to the same prescription for the situation in which I am Sam. But why should it matter what that prescription is? If the prescription does not matter then why should I consider my preferences for the consequences of acting according to that prescription?

If universalisability were a logical requirement then that would **not** be sufficient to generate a requirement that the rational agent should engage in critical thinking. Mackie is, I believe, right to claim that first-stage universalisability is trivial. It is trivial

Rejection of Hare's position on logical requirements.

because a rational agent may take no interest in the prescriptions made for totally hypothetical situations.

In claiming that a rational agent may "refuse to consider the application of his moral principle" (Hare 1981 p.113) to totally hypothetical situations, I am **not** claiming that one may **refuse to apply** that principle to those situations - if universalisability is a feature of moral talk then that principle does apply to those situations and I have not disputed Hare's claims with regard to universalisability. Rather I am claiming that the rational agent may accept that application, he may accept that he is committed to a prescription for that situation, but he may, nevertheless, rationally decline to take into account how it would be if that prescription were acted upon.

It may be that the process of discovering how it would be **for others** is central to moral thinking; but that cannot be because moral thinking requires a rational agent to consider how it would be **for him** if he were those others. The rational agent may not be interested. He may say: 'I will not be those others so I do not mind what prescription I make (or have implicitly made) for those situations in which I am those others'.

However, Hare might claim that a '**rational**' agent cannot sincerely make a prescription (even implicitly) for any situation (even a totally hypothetical situation) unless he knows what his preferences are for the consequences of acting according to that prescription in that situation. If we accept a notion of 'rationality' which makes this the case then we cannot so easily combine an acceptance of the universalisability of moral judgments with a denial of the relevance of totally hypothetical situations. We shall have to examine whether in gaining knowledge of such situations the rational agent need find

any reason to revise a (universalised) prescription for the actual situation.

The fanatic and the amoralist.

Hare argues that critical thinking will yield conclusions which are the same as those of a certain type of utilitarianism. He claims that the only way that a rational agent can avoid such conclusions is to decline to engage in critical thinking and/or to decline to make moral judgments (ie. to be an 'amoralist'). But that may not be the only way in which we can, as rational agents, avoid reaching the conclusions which a certain type of utilitarian would reach.

Suppose that, in the example of Sam and George, I am disposed to judge that Sam **ought** to look after George. But suppose also that the utilitarian conclusion would be that Sam **ought not** to look after George - because, for example, George is such a terrible old fellow that he would completely disrupt Sam's life, Sam wants very much not to have his life disrupted, George could be quite content elsewhere, and so on. If I make the judgment that Sam ought to look after his father, and it is a moral judgment, then I am also committed to the same judgment for those situations in which I am Sam and in which I am George.

In the hypothetical situation in which I am Sam, I want what Sam actually wants. If I am committed to a prescription for this situation then, Hare argues, I must (as rational) attempt to discover just what it would be like for me.

The first way in which I may avoid reaching the utilitarian conclusion is to deny (as I did in the last section) that I have to discover this. I do not need to

do this because it does not matter what prescription I make (or have made implicitly) for a situation which will not occur. I am, it is true, logically committed to the prescription that I ought to look after George in that situation, but I do not really mind what I am committed to because that situation will not happen.

Thus someone who has a strong 'moral intuition' about this case (someone who is 'fanatical' in Hare's sense), and held that Sam ought to look after his aged father (however awful it would be for Sam and however little difference it would make to George), could rationally hold on to his judgement. The rational 'fanatic' **can** remain unmoved.

If this is correct then Hare's 'rationalist' project fails: reason, alone, cannot provide a route to agreement in moral judgments because rational agents may be 'fanatics'. But if the notion of rationality, given at the end of the last section and attributed to Hare, were appealed to then the project may remain intact.

The second way in which I may avoid reaching the utilitarian conclusion is described by Hare. In this case I **do** engage in critical thinking - I discover what it would be like for me to be Sam in this situation. In order to do this I gain knowledge of the preferences which Sam in fact has (the preferences I would have if I were Sam) and consider the consequences in the light of those preferences. In order to gain such knowledge (Hare claims) I must **acquire** preferences for the situation in which I am Sam which are identical to the preferences which Sam has for the actual situation.

This last claim is referred to, by Williams, as Hare's 'epistemological premiss' and I shall return to it later. It is the claim (Hare 1981 p.95) that I cannot **know:**

Rejection of Hare's position on logical requirements.

if I were Sam in this situation I would prefer with strength S that x should happen;

unless:

I **now prefer** with strength S that if I were Sam in this situation x should happen.

If I acquire preferences in this way then it will be the case that: **I** do not want, say, to suffer my father's dreadful habits, to put up with his disregard for the feelings of the rest of my family, and so on. I have (after critical thinking) a very strong desire not to look after George.

But note, it is not (according to the argument) that I do not want **Sam** to suffer in the way in which I now know he would suffer. My preference does not concern **Sam's actual** situation, it concerns **my hypothetical** situation. It is this preference which (according to the argument) may lead me to revise my (universalised) prescription for the actual situation.

However, I have still not reached the utilitarian judgment. In order to do that I must be willing to weigh the preferences I have for the hypothetical situation in which I am Sam against those I have for the hypothetical situation in which I am George and, as a result, make the same prescription for the actual and for the two hypothetical situations: 'Sam not look after George' in the actual situation, 'I not look after George' in the one hypothetical situation, 'Sam not look after me' in the other hypothetical situation.

But the rational agent can refuse, as Hare says (1981 p.183) to take this step; he can, for example, prescribe: 'I **not look after** George' in the one hypothetical situation, Sam **look after** me' in the other hypothetical situation, 'Sam **look after** George' in the actual

Rejection of Hare's position on logical requirements.

situation (this last being the prescription which was entailed by his original judgment).

As Hare puts it: the **amoralist** may, rationally, not accept any **universal** prescription for this situation, he may decline to make a **moral** judgment. He can no longer prescribe 'Sam **ought** to look after George'; but neither is it the case that he has to reach the utilitarian conclusion. Even after engaging in critical thinking the rational agent may prescribe 'Sam look after George'.

But now I wish to claim that, even after engaging in critical thinking, the rational agent can prescribe 'Sam **ought** to look after George'. Or, more precisely, I wish to claim that, even after gaining knowledge of the preferences I would have if I were Sam, I can, as rational agent, make that universalised prescription. This claim will involve challenging the role of the epistemological premiss in the context of deliberations about totally hypothetical situations.

Hare's epistemological premiss.

Hare's premiss implies that I cannot **know**:

a) if my house were on fire I would prefer, with the greatest possible intensity, that I should get out of it;

unless it is the case that:

b) I now prefer, with the greatest possible intensity, that if my house were on fire I should get out of it.

Williams (1985 p.90) gives this application and argues that it reveals the implausibility of the epistemological premiss. If I am making a prudential decision (for example, I am deciding whether to install smoke alarms) then, Williams says, there is no sense at all in which my

Rejection of Hare's position on logical requirements.

present preference is of the same strength as the preference I would have if the house were actually on fire, and it is not rational that it should be.

I do not agree with Williams as to the force of this example. In making prudential decisions of this sort, the rational agent does not require knowledge of a). I do not need to know the intensity of my preference for avoiding being trapped in a burning house; I merely need to know that that preference is much greater than the preference for saving the cost of a smoke alarm. The rational agent does not need to make an imaginative leap into such a situation because a decision such as this can be made without the knowledge which such a leap would yield. This application of the premiss will, however, be helpful in making clear my own challenge to Hare's use of that premiss.

Suppose that I claim to know a) but that I deny b) and claim instead that I do not now mind in the least if my house catches fire and I fail to get out. Given the epistemological premiss then it seems to be the case that either my claim to know a) is false, or my denial of b) is insincere. As Hare might say (1981 p.94): would not my lack of knowledge, or else my insincerity, be exposed if somebody said 'All right, if you don't mind, let's lock you in and set fire to the house'? If I protested then I would begin to reveal my insincerity; if I acquiesced but tried to break down the door when the flames spread then I would reveal my (previous) lack of knowledge.

However, suppose that I know my house **will not** catch fire. First, if my denial of b) is insincere then I also know that my insincerity need never be exposed. I have no need to protest against the threat since I know that it will not be carried out - I will not be locked into my

Rejection of Hare's position on logical requirements.

burning house. But, second, could I not, given this knowledge, **sincerely** deny b)? Thus:

> even though I know that if the house **were** on fire I would want to get out,
> I do not now mind in the least if my house were to catch fire and I failed to get out,
> **because** I know that my house **will not** catch fire.

Is there an inconsistency here? Are we failing to take account of the sense of 'know' which, Hare claims, entails the epistemological premiss; or does the further knowledge that my house **will not** catch fire make a difference? Here it is difficult to ignore the fact that I do **not** know that my house will not catch fire. However, things may be clearer if we consider the case involved in our example of critical thinking. Thus:

> even though I know that if I were Sam I would not want to look after my father,
> I do not now mind in the least if I were to be Sam and had to look after my father,
> because I know that I will not be Sam.

Despite knowledge of my preferences **in** the hypothetical situation, I have no preference **for** that hypothetical situation precisely **because** it is hypothetical - it will not happen.

If we are not here flouting a conceptual truth then such examples may be used to cast doubt upon the application of the epistemological premiss in the context of totally hypothetical situations. But there may be an alternative approach in which we preserve a form of that premiss.

In this approach we accept that knowledge of preferences in a given situation involves imaginative acquisition of preferences for that situation but maintain that one can then, rationally, take no account of or discount those preferences in one's deliberations eg. the deliberations involved in deciding upon a (universalised) moral

Rejection of Hare's position on logical requirements.

judgment for a particular situation. One can discount a preference for a totally hypothetical situation **because** that situation will not happen.

To discount a preference is to deliberate as if one did not have that preference. Deliberating as if one did not have a **preference** which one does have is not the same as deliberating as if one did not have **knowledge** which one does have. The latter may be irrational, but the ability to do the former is (I shall claim in a later chapter) a fundamental feature of human nature and to deliberate in that way may be rational.

The notion of rationality to which Hare appeals was outlined in chapter 3 and at the end of an earlier section in this chapter. It entails that an agent is irrational insofar as he **fails to gain and take account of knowledge** which may affect his judgment - in this case his (universalised) prescription for a particular situation. But that notion of rationality does not entail that an agent is irrational insofar as he **fails to count a preference** which may affect that judgment.

As a rational agent one may, for example, discount a preference for a cigarette because one is attempting to give up smoking; one may discount a preference for running away because one is ashamed of one's timidity; and one may discount a preference one has for a totally hypothetical situation because one makes (or prefers to make) universalised prescriptions on the basis of the preferences one has for situations which will happen or may happen.

In the next chapter I shall look more closely at the grounds which a rational agent might give for discounting preferences. Here I wish to maintain that, when considering a universalised prescription, a rational agent may decide to give no weight at all to the

Rejection of Hare's position on logical requirements.

preferences which he has for totally hypothetical situations and give as grounds for that decision the fact that such situations have no probability of occurring.

A rational agent may say: Perhaps (given the logic of moral language) my moral judgments are universalisable. Perhaps (given the nature of rationality) I need to know what my preferences are for the consequences of acting according to prescriptions which I, thus, implicitly make for totally hypothetical situations. Perhaps (given the epistemological premiss) in gaining such knowledge I shall acquire preferences for those totally hypothetical situations. But I do not need to count such preferences when deliberating upon a (universalised) moral judgment for a particular situation. Having discounted those preferences I can, when deliberating, sincerely claim that I do not now mind in the least if I were Sam and had to look after my father.

A rational agent adopting this stance can then point out that gaining knowledge of preferences and consequences in hypothetical situations in which I am those other persons involved in the actual situation turns out to be (given that stance) pointless. That agent can then say (as at the end of the first section): I am not interested in how it would be if I were those others; I will not be those others and so I don't mind what prescription I make (or have implicitly made) for those situations in which I am those others.

If this is so, then the second way of reaching a decision which does not agree with that of the utilitarian is open to the **moralist** as well as the amoralist. I can, despite the preferences I acquire when I envisage myself as Sam (ie. I want it to be the case that if I were Sam then I would not look after George), rationally prescribe for the situation in which I am Sam 'I look after George'. I do so because my prescription for the totally

Rejection of Hare's position on logical requirements.

hypothetical situation is not determined by my preferences for that situation (which are discounted) but is determined by the universalisation of my prescription for the actual situation. I can thus maintain the **moral** judgment 'Sam **ought** to look after George'. In gaining knowledge of totally hypothetical situations the rational agent need find no reason to revise a (universalised) prescription for the actual situation.

Even if we grant Hare's claims with regard to the logic of moral language, the nature of rationality and the epistemological premiss (each of which might be challenged directly), we can still claim that the moral judgments of a rational agent need not be determined by critical thinking.

Hare's rationalist project fails, I believe, because a rational agent may see totally hypothetical situations as being simply not relevant to deliberation about what to do or what one ought to do. I would not wish to claim that it is **irrational** to gain knowledge of totally hypothetical situations or to count our preferences for such situations. We might do the former as a means to acquiring preferences for the actual situation. We might do the latter because, like Hare, we are so inclined. But if we are not so inclined, and if our moral judgments are to reflect the preferences of others involved in the actual situation, then the link between our preferences and the preferences of those others must be forged in a different way.

If the possible suffering of Sam is to affect my deliberations then that will not be because I have an aversion to **my** suffering (in the totally hypothetical situation) as he would suffer (in the actual situation). If that possibility is to affect my decision as to what ought to be done then imaginative identification with Sam

Rejection of Hare's position on logical requirements.

must result in **my having an aversion to Sam suffering** as he would suffer in the actual situation.

My aversion to your suffering.

The first modification of the characterisation of 'critical thinking' which I am proposing stipulates that the imaginative identification involved in such thinking concerns, say, my aversion to **your** suffering and not my aversion to **my** suffering were I you.

Here it will be useful to clarify just what sort of 'identification' I shall be discussing later. Suppose that, in a particular situation 'S', Sue wants to eat an apple. Expressed in terms of phrastics and neustics (see Hare 1952 ch.2), we have:

Sue assents to,
'In S, Sue eats an apple, please';

The sort of identification, which Hare requires, relates to a hypothetical situation 'HS' in which, say, I am the person whom I am identifying with - eg. Sue. Thus:

I now assent to,
'In HS, **I** eat an apple, please'

The sort of identification, which I shall stipulate is part of critical thinking, relates to the actual situation and requires me to have the preferences which Sue has for that situation. Thus:

I now assent to,
'In S, **Sue** eats an apple, please';

Thus if I identify, in this way, with Sue when she faces the possibility of suffering then, given she assents to 'Sue not suffer, please', I will also assent to the very same statement. I want Sue not to suffer, just as she wants not to suffer.

Rejection of Hare's position on logical requirements.

Hare's argument with regard to the necessity of his sort of identification is based on the claim that I need to know how much Sue wants not to suffer (because I am prescribing for the situation in which I am Sue) and I cannot, Hare says (1981 p.95), know how much Sue wants not to suffer unless I now have an equal aversion to my suffering were I Sue. As I said above, Williams disputes the epistemological premiss which is employed here.

Williams (1985 p.91) gives a further argument for rejecting that premiss; an argument which was not mentioned above. The **cruel** person, he says, knows very well just how much Sue does not want to suffer and yet he has no preference to give help, to alleviate the suffering - on the contrary he is encouraged by his knowledge to act in just the way which will ensure the suffering. He certainly **knows**; but he does not assent to 'Sue not suffer, please'.

This objection clearly misses the point of Hare's argument. That argument does not rest upon a claim that knowledge of Sue's aversion to suffering requires assent to 'Sue not suffer, please' - it merely requires that I (with a vigour equal to that of Sue) assent to 'I not suffer, if I were Sue, please'. But that, as I have argued, is precisely why the argument does not succeed.

The employment of the epistemological premiss is fruitless in the context of totally hypothetical situations. My aversion to **my suffering were I Sue** need not figure in my moral (or any other form of) thinking.

I shall argue for an identification between moral thinking and critical thinking; but that form of thinking will (as modified) involve my acquiring an aversion to **your** suffering. Furthermore, that aversion will derive not from the rational agent's acknowledgment of logic, facts and the universalisation of moral judgment but from

Rejection of Hare's position on logical requirements.

the quality of the agent's motivation. The agent is such that knowledge of your suffering

> **does** yield a preference that the suffering be prevented or alleviated

and (unlike that of the cruel person)

> **does not** yield a preference that the suffering be ensured or heightened.

The inadequacy of Hare's appeal to 'moral' language.

Thus far I have not questioned Hare's claims with regard to universalisation, rationality and the epistemological premiss. But we might cast further doubt upon Hare's argument if we raise questions about the scope of the concepts to which Hare's analysis applies. Specifically: to which creatures does the 'ought' of morality (and hence universalisation) apply?

Hare says (1981 p.90) that he is happy to accept a scope which ensures inclusion not only of all people but also of other sentient beings. He adopts this position in deference to vegetarians who, he says, will wish to include other animals within the scope of morality.

But, surely, given his argument Hare should not defer to those vegetarians unless the scope which they desire is required by the logic of our moral language. If it were legitimate to defer to the wishes of the vegetarian then would it not be legitimate to defer to the wishes of, say, the racist who will wish to exclude other races from the scope of morality? The legitimacy of either response rests, given the argument, upon the nature of moral language. Hare claims that we all share a use of certain words and concepts. But is it the case that the vegetarian and the racist use those words and concepts in the same way; and, if they do not, is it clear that one or both of them is **misusing** those words and concepts?

Rejection of Hare's position on logical requirements.

Hare insists that the logic of our moral language does not permit us to make different moral judgments about cases which are identical in their universal properties (1981 p.115). The properties of a situation which are not 'universal' are, he says, the identity of those involved, the place, the time, and the 'actuality' (if that is a property). We cannot, logically cannot, make different moral judgments for two situations on the basis that one involves x and the other y, one is here and the other is there, one is on Tuesday and the other on Wednesday, one is actual and the other hypothetical.

We can, **logically can**, make different moral judgments on the basis of the species or race of those involved. However, our rationality then demands, given the argument, that we consider the consequences of acting according to those judgments and affirm, revise or reject those judgments **in the light of the preferences of all those involved.**

Presumably 'all those involved' includes all those having preferences - regardless of, say, species or race. The universalisability of moral judgments entails that any such judgment does not only apply to the actual situation involving this member of the species or race but also applies to the hypothetical situation in which **I** am that member. The nature of rationality and the epistemological premiss then, given the argument, ensure that my judgment reflects the preferences of that member.

But those who wish to exclude another species or race from the scope of morality may now reject this use of universability on the basis that judgments relating to members of that species or race are not **moral** judgments at all. We might (and I think Hare would - at least with regard to race) then argue that the logic of moral language does not permit such a move. But what force

Rejection of Hare's position on logical requirements.

would such an argument have? As Singer (1988 p.155) says the response might then be:

> "If you tell us that our concepts imply equal consideration for the preferences of animals, we shall simply adopt a new set of concepts, which implies universalisability up to, but not beyond, the boundary of our own species."

Hare (1981 p.18) admits this possibility and points out that "if we were to alter the meanings of our words, we should be altering the questions we were asking". He then goes on to insist that if we are going to ask new questions then we ought to be satisfied both that the new questions are important and that the old questions are unimportant.

However, as Singer (1988 p.156) points out:

> "If members of a society simply do not care about the welfare of outsiders, whether of another nationality, race, or species, they will easily accept that some appropriately restricted set of concepts captures everything important about the questions asked by the set of concepts Hare has analysed, and leaves out only some unimportant matters with which they do not wish to be bothered."

Hare must offer reasons why such a group of people should not adopt such a set of concepts. If he does not then, once again, the rationalist project fails - a rational agent can, say, be a racist.

It may be possible to argue that the welfare of such a group is reduced, or not improved, by their lack of concern for outsiders. Hare (1988 p.273) claims, in response to Singer, that the maltreatment of (certain types of) outsiders is not necessary for, or even conducive to, the happiness of those in such a group. But, firstly, this claim is not sufficient to ensure

Rejection of Hare's position on logical requirements.

equal concern across the boundary; it requires only that members of the group consider the consequences of maltreatment of the outsiders for the satisfaction of the preferences of those **in the group**; the preferences and degree of suffering of the outsiders has importance only insofar as it leads to undesirable consequences. Secondly, it is very doubtful whether such a claim would hold in **all** circumstances.

Singer (1988 p.157-8) offers a different argument. If we adopt a set of concepts which imply universalisability up to the boundary of our own nation, race or species then, he says, can we not be criticised for arbitrariness? "At whatever point universalisability stops, one can raise the question: 'Why stop there?' ... Only the boundary of sentience ... seems to avoid this kind of arbitrariness.". The response may now be that a 'closer' boundary is not at all arbitrary **if** it corresponds to the boundary of our concern. But then, Singer claims, such a response to the charge of arbitrariness has a considerable cost. Those responding in this way have no defence against those who say: 'I don't care for all those who are within **your** sphere of concern. I care only for a smaller group'. There is then "no logical stopping place short of individual egoism".

Singer concludes that, since we all have reason to defend our sphere of concern against those who do not share it, then we have reason to avoid arbitrariness not by drawing the boundary at the boundary of concern but by drawing it at the boundary of sentience - beyond which there are no preferences.

This argument rests upon the claim that the only boundaries which are not arbitrary are those which mark the boundary of concern or which mark the boundary of all preferences. If our rational desire to defend our sphere

Rejection of Hare's position on logical requirements.

of concern provides reason not to appeal to the first then we are left with the second.

It seems to me that this is merely to say that there is no way in which we can defend our own particular sphere of concern against those who do not share it (with which I can agree) **and**, having ceased to defend that concern, we must (to avoid arbitrariness) appeal instead to the 'relevant similarity' between those who have preferences as opposed to those who do not. But in what way is that similarity any more or less relevant than the similarity between, say, those who belong to a particular race? Setting the boundary at the boundary of concern is not arbitrary; setting the boundary according to some other feature (**any** other feature: sentience, or species, or race, or nation) is also not arbitrary. Neither means of setting the boundary will help in settling differences between rational agents.

I would argue that the only reasons we can offer for insisting upon a certain set of concepts (and the concerns which can be expressed by means of those concepts) are **moral** reasons. Given Hare's rationalist approach this would be to argue in a circle; but, perhaps, we should not adopt that approach.

Hare wishes to start from an analysis of language and end with a choice between amoralism and a certain form of Utilitarianism. In earlier sections I argued that his analysis does not yield that choice. In this section I have argued that, even if it did yield that choice, it would not determine **whose** preferences the 'Utilitarian' should consider.

We can add that neither would it adequately determine **which** preferences the 'Utilitarian' should consider. As Harsanyi (1988 p.90) points out: "even if we accepted Hare's argument at its face value, prescriptivity and

Rejection of Hare's position on logical requirements.

universalisability would be of very little help in deciding the specific form our utilitarianism should take". Should it disregard uninformed preferences, anti-social preferences, the preferences of the unborn? Sen (1980 p.80) adds to the list: past preferences which one no longer has, the preferences of the dead, preferences where one is not aware of their satisfaction or dissatisfaction.

Hare (1988 p.242) claims that such questions can all be answered. But I do not believe that those answers are adequate and (in the next chapter) I shall argue that this is so - at least with respect to 'anti-social' preferences. If any of the questions concerning **whose** preferences and **which** preferences cannot be answered on the basis (direct or indirect) of an appeal to Hare's analysis of the logic of moral language then that appeal is not adequate.

Hare's response to the central educational question.

In the last section I attempted to give support to the view that Hare's analysis (even if it yields Utilitarianism) will not yield a specific form of Utilitarianism. But now we can ask a broader question: even if that analysis did yield a choice between amoralism and a **specific** form of Utilitarianism, what would that show? As Brandt (1988 p.36) says: "there is a further problem of showing why anyone should be interested in whether one ought or ought not in that sense".

Brandt's exposition confuses this question with a different question: 'If I grant that I ought to do A then why should I act accordingly?'. Hare responds by pointing out that the prescriptivism which is part of his analysis provides the link between a sincere assent to

Rejection of Hare's position on logical requirements.

'ought' and an inclination to act. But this leaves the first question unanswered.

It it were the case that our use of moral language had the features which Hare describes, and if it were also the case that reflection upon those features revealed certain canons of moral thinking, then Hare would conclude that the thinking of rational agents **must** be governed by those canons when moral judgments are being made. But what is the force of that 'must'?

Suppose there are those who use moral language in the way Hare describes, but whose 'moral' thinking is **not** governed by those canons. If there are not many such people then either Hare's analysis of moral language is wrong or his efforts to make clear the implications of that use of language are unnecessary. Hare's efforts now reveal to those people that the form of thinking which they have engaged in, and which they thought was moral thinking, is not, given their use of language, moral thinking at all. **Must** such people, as rational, now adjust their mode of thinking?

Clearly they have a choice: if they wish to keep intact their use of language (the implications of which were unclear to them) then they **must** adjust their moral thinking; if they wish to keep intact their form of thinking (which may have been very clear to them) then they **must** adjust their use of language. The rational choice will, presumably, be the one which reflects the relative importance of the two aims. It may be the case that certain features of the way in which moral language has been used would prove (on reflection) to be less important than the way in which they have been accustomed to arrive at, and reach agreement upon, 'moral' judgments. If this were so then the rational course would be to alter the use of moral language.

Rejection of Hare's position on logical requirements.

Such considerations may lead us to doubt whether we should begin with an investigation of our use of moral language. It may be more fruitful to enquire into the nature of a particular way of thinking and the role which it plays in our lives. In particular, it may be best to focus upon variants of the questions given in chapter 2: 'Why engage in this particular form of thinking?' and 'Why educate ourselves and others to be inclined to engage in this particular form of thinking?'. As Hare points out, such questions remain central even if the particular form of thinking we choose to consider is determined by an analysis of language.

However, there is a crucial difference between the two approaches. If we set out to clarify and justify the role which a particular form of thinking has in our lives then we may well end up enquiring whether a **different** form of thinking might be **more easily** clarified or justified. If, on the other hand, we insist that an analysis of language reveals **the** form of 'moral' thinking then we will not stray from the task of attempting to clarify and justify that (and only that) form of thinking. The latter approach may lead us to ignore possible modifications to that form of thinking which - given the manner of our response to the two questions above - would be sensible.

For example, Hare asks how should we best educate our children. His discussion (1981 ch.11) contrasts two possibilities, educate our children to be:

1. disposed to act according to moral principles and able to think morally;

or

2. disposed to act according to prudential principles and able to think prudentially.

The choice is between morality and prudent self-interest.

Rejection of Hare's position on logical requirements.

Hare argues that an education aimed at 1. would "be best in the child's own interest" (1981 p.195) and, given this is so, he claims to have provided "an adequate defence of morality" (1981 p.191).

Now, firstly, I find Hare's argument unconvincing. His argument closely resembles that of, say, Foot (1958 final section) and Mackie (1977 p.191-2). The main points relate to the consequences of acting, and being seen to act, according to principles of self-interest and to the difficulties of concealing the fact that one is disposed to act in this way. But as Plato (Republic Book 1) points out those consequences and difficulties may well depend upon the strength and wit which one possesses.

Hare seems to claim that his argument applies even to those having a large measure of such strength and wit. He says, for example, that if it is alleged "that in the past people have amassed large fortunes in business careers which were far from unspotted, I reply that the money **did not on the whole bring them happiness**, and that with their talents they could have done better **for themselves** by making less money in a more socially beneficial career." (1981 p.196 - my emboldening). This seems to me to be wishful thinking. It would be nice to believe that "in the world as it is" (1981 p.194) good people on the whole do better **for themselves** than purely self-interested, unscrupulous or corrupt people, but I find it very difficult to convince myself that this is so.

Hare's answer to our earlier question 'Why educate our children to be inclined to think morally?' is in terms of the child's own interest. But now, secondly, if that is what matters and if that is what is involved in providing an 'adequate defence of morality' then why compare only two possibilities - morality and prudent self-interest? Hare compares and contrasts only these two possibilities

Rejection of Hare's position on logical requirements.

because he is concerned to defend **the** form of moral thinking - as revealed by his analysis of moral language.

Morality, for Hare, is about maximising satisfaction of **all** the informed preferences of **all** those involved in each situation. But perhaps the child's interest would be best served by an education which encouraged dispositions to act according to principles conducive to maximising satisfaction of **only some** of the preferences of **only some** of the people. For example, it may well be the case that the interests of a child born to the rich and powerful would be best served by an upbringing which ensured consideration in dealings with other rich and powerful people but the pursuit of prudent self-interest in dealings with others. A 'morality' which extended only to members of the child's own group may be the best 'morality' from the point of view of the interests of such a child.

I do not believe that 'morality' can be adequately defended by an appeal to the interests of those whom we educate in the world as it is. But my point here is that, even in terms of the aim of promoting the interests of the child, Hare's analysis has led him to consider a limited range of options. Hare does not consider other options because his analysis has provided very tight characterisations of moral, and prudential, thinking and his answers to questions relating to education and the child's interests are in terms of those characterisations.

If we are concerned to answer questions about why we do engage in this or that form of thinking, and what form of thinking we should engage in or educate others to engage in, then we should not be constrained by an analysis of our 'moral' language. Why should various features of the way in which we **happen to** use moral language have any special significance? Our use of moral language may not

Rejection of Hare's position on logical requirements.

exactly reflect the way we in fact think, and it may not provide any clue as to how we should think.

Now there are, of course, those who believe that moral language has a special significance because it relates to certain special 'facts'. A moral realist, for example, will say that answers to questions about the interests of our children, or of society in general, are simply not relevant, or not directly relevant, to questions in moral philosophy. For the realist, moral language has an 'extension' and it must relate to and be determined by moral facts. Thus 'analysis' of our use of moral language may play a central role and be seen to be an essential starting point.

To take an example from a different area of philosophy: the problem of the nature of causation. Here we might look closely at the features of paradigmatic examples of causation - striking a match, throwing a ball which breaks a window, and so on. We might also look at the ways in which we describe such examples and attempt, say, to discover the sorts of statements which we would see as warranting a description in terms of cause and effect. That is, we could engage in what Mackie (1974 p.ix) calls 'factual' analysis and 'conceptual' analysis. Whether or not a clear distinction can be made in this way, it is true to say that if we are realists about the world (and the 'role' of causation in that world) then we will maintain that the beliefs we have, and the meanings of our descriptive expressions, ought to reflect the way things are. The refinement of our concepts and the modification of our beliefs about the world will proceed hand in hand, but both will be constrained by the nature of the things to which the terms in our language refer. Here questions about why we should think and speak in a particular way, about cause and effect, would have a straightforward answer: because that is the way the world is (and that is the way causes and effects are). The aim

Rejection of Hare's position on logical requirements.

of philosophical analysis would be to bring conformity between the way we think and speak and the way things are.

If, similarly, we believe that there are moral 'facts' then analysis of our moral language would have a purpose: namely to refine and correct the meanings of our moral expressions so as to bring conformity with the nature of that to which it refers. If the use of language which resulted from such a process of analysis entailed constraints upon the way in which we reach moral judgments (as it presumably would) then that way of reaching judgments would have a special significance. Questions like 'why educate our children so that they reach judgments in this way?' would, again, have a straightforward answer: because that is how **moral** judgments are made. It would still make sense to ask whether we had any reason for bringing up our children to make moral judgments, but the point is that there would be a substantial difference between the two questions.

But if one rejects, as Hare does, any form of realism in morality then it is difficult to see how these questions can be separated in any significant way. For Hare, the relationship between a particular way in which we reach judgments and the fact that those judgments are 'moral' is simply a consequence of the way in which we happen to use language. To ask whether we should bring up our children to make moral judgments is just to ask whether we should bring them up to reach judgments in that way and, more importantly, that way has no special significance over and above its being (according to Hare) the way which we happen to have enshrined in a particular form of language.

Now it may be the case that a particular way of reaching judgments has become enshrined in our language because, as the realist claims, it relates to a special sort of

Rejection of Hare's position on logical requirements.

'fact'. But if we reject realism then we have to look elsewhere for an explanation of such features of our language. We might then claim that a particular way of reaching judgments has become enshrined in our language because it relates to certain of our aims and purposes and to distinctive features of human agency. But if this were the case then it would be sensible to ask not 'what way of reaching judgments would conform to our use of language?' but, rather, 'what way of reaching judgments would achieve those aims and purposes and reflect those features?'. It would be sensible to go straight to the 'main business' of investigating, firstly, some of our ways of reaching judgments about people, actions, and states of affairs and, secondly, the relationship between those ways of reaching judgments and the aims and purposes which we share.

Furthermore, such a line of investigation would have interest even if the realist were right. It may be that the realist with his analysis of 'moral' language (and Hare with his) will claim that the form of thinking which we describe at the end of the investigation has nothing to do with 'morality'. But if it turns out that this form of thinking does play a central role in our lives, and that it does promote some of the aims and purposes which we share, then it may not matter overmuch whether it also turns out to be 'moral' thinking.

Rejection of Hare's position on logical requirements.

CHAPTER 7.

Objections to Utilitarianism.

Recapitulation.
Consequentialism as indirectly self-defeating.
Malevolent preferences.

Recapitulation.

Hare's aim is to show that rational agents must reach their 'moral' judgments by means of a certain method. He also aims to show that if two agents fail to agree in their moral judgments, and are thus not disposed to act in the same way, then that must be because one (or both) lacks the knowledge which the method requires him to gain - the method is such that if it is fully undertaken then it will yield a unique judgment and the deliberator will be disposed to act accordingly. Unless I am an 'amoralist' (and according to Hare I have good reasons not to be) then my failure to want to act in the right way is always the result of my not being fully rational.

Hare is thus a moral rationalist. Not in the sense that he believes 'reason' alone can yield the answers to moral questions. But rather in the sense that he believes that if our reasoning makes use of the facts, and is in accordance with the logical requirements generated by [our] concepts, then that will be sufficient to settle moral questions. This is not, as Hare points out (1985 p.48), merely to claim that "we can rationally decide what to do"; it is to claim that rationality places constraints on the form of our practical reasoning and, in particular, these constraints relate to the logic of our moral language - logic requires that we

'universalize' our moral judgments. It is from this perspective that we can understand Hare's claimed affinity with Kant.

Hume (1888 p.414) says that our impulses do not arise from reason; reason merely discovers the means to the object of our impulse and thus 'directs' those impulses to their object. But, for Hare, reason not only directs our impulses it also demands that we share the impulses of others when making moral judgments. For Hare, as for Hume, reason may be in some sense "the slave of the passions" but, Hare believes, it is not merely the slave of **my** passion - it can demand (through the logic of our moral language) that I have the passions of others. In making this demand it does **give rise to** an impulse - the impulse to satisfy (as much as possible) the preferences of all concerned - and that impulse will be shared by other rational agents insofar as they make moral judgments. Each such rational agent will share the same impulses and reason will then direct those impulses to the same object.

Hare's archangel has "superhuman powers of thought, superhuman knowledge and no human weaknesses" (1981 p.44). Each archangel would therefore be able to scan all the properties of a situation, including the consequences of alternative actions, imaginatively identify with each person involved, and each would, by means of critical thinking, arrive at the same universal principle prescribing action for all situations similar to the one considered. Only those lacking the ability of the archangel could arrive at a different universal principle; the ability to form a judgment in the light of all the facts would be sufficient to ensure agreement and a disposition to act in the same way. If it could be shown that such a method of forming a judgment was (necessarily) appropriate to moral thinking then, as Hare says (1981 p.46), this would be "a highly rationalist

thesis". The links between facts, 'rationality', moral judgment and disposition to act would be firm.

If this were Hare's position, and if the arguments offered were sound, then the requirements for a moral education, of the type we are considering, would be clear: to ensure that educatees became as rational as possible and had those dispositions which reflected the principles they would adopt if they were fully rational.

However, I have already argued that critical thinking is not a logical requirement of moral thinking; so that even if, "at the end of their critical thinking, [archangels] will all say the same thing" (1981 p.46) and act accordingly, Hare has not produced an argument which allows us to conclude that the result of such thinking is a 'moral' judgment and that the resulting action (if performed successfully) would be 'morally right'. So that, I believe, Hare has not succeeded in establishing the link between facts, 'rationality' and moral judgment in a way required by a highly rationalist thesis.

As Hare points out (1981 p.190), there is a further gap, in his account, between factual beliefs and moral judgment. Someone may simply decline to make moral judgments. We need, at least, **the 'impulse' to engage in moral thinking.** This gap is, I believe, more important than the 'logical' gap which opens up if Hare's analysis of our moral language is incorrect - for this gap would be just as significant even if Hare's analysis were correct. The correctness of the analysis would simply mean that we could speak of the impulse to 'moral' thinking rather than, merely, the impulse to critical thinking. Hare offers "reasons of a non-moral sort" why 'amoralism' (a refusal to make judgments based on critical thinking) should not be chosen (as a future goal for ourselves or as an educational aim) but, I have argued, these are not convincing. The gap now seems to

be very large. We have a description of a form of thinking but apparently no good reasons for engaging in such thinking, nor for encouraging educatees to engage in such thinking, nor for believing that such thinking has anything to do with 'morality'.

It may well be, and I believe it is, the case that it is not possible to find grounds for a moral theory which establishes the link between rationality and morality as firmly as Hare or Kant would wish. We may, however, still have good reasons for combining some elements from both moral viewpoints in order to formulate a moral 'theory' which is acceptable to us - which entails a morality which we have reason to let into our lives and, especially, into the lives of those we educate.

Leaving aside the "highly rationalist thesis", Hare's work does give us a very clear description of a form of thinking which (if in this respect Hare and others are correct) can be shown to be capable of underpinning the ('intuitive') general moral principles which most of us would assent to. It also offers a means of resolving the inevitable conflicts between such principles. Hare's account, in Moral Thinking and earlier works, also attempts to make clear how the traditional attack upon this form of thinking, in terms of highly unusual cases, can be seen to miss its mark. The rebuttal of such attacks was seen by Hare, especially in earlier works (eg 1976 p.36), as "the main move" in his defence of "this sort of Utilitarianism".

Hare's description of the form of thinking involved in his version of Utilitarianism allows us to see, in a new light, the relationship between a way of arriving at moral judgments and our possession of certain moral 'intuitions' - our "spontaneous convictions, moderately reflective but not yet theorized" - as Williams describes them (1985 p.94). This relationship rests on the

implications which Hare's account of moral thinking has for the aims of moral education.

We are not archangels and therefore there will be many occasions on which we will not be capable of full critical thinking, we "will not have the time, or the information, or the self-mastery to avoid self-deception prompted by self-interest" (Hare 1976 p.32). We will, therefore, wish to educate our children (and ourselves) in such a way that we "implant" those general principles which will lead to actions in accord with critical thinking in "most situations that are actually encountered". Hare (1976 p.32) says "implant" because they will need to be "not rules of thumb, but principles which they will not be able to break without the greatest repugnance, and whose breach by others will arouse the greatest indignation".

If we address our critical thinking to highly unusual, or fantastic, cases then, of course, it will be a fairly easy matter to generate a conflict with such general principles because they are designed to be "in accord with critical thinking in most situations that are actually encountered". The morally well-educated person (as well as the intuitionist) would find that it would go "very much against the grain" to fail to act in accordance with those principles in order to act according to archangelic thinking (even his own).

This approach contrasts with that of Sidgwick. According to Sidgwick, the distinction (in the context of unusual cases) between what is right to do in theory and what one is disposed to do in practice is a distinction which determines two groups of people. The first group is capable of Utilitarian thinking in each situation and capable of determining which general principles ought to be adopted; the second group consists of those whose actions and thinking are guided only by those principles.

Whilst agreeing that Hare's "main move" may allow one to offer a more defensible form of Utilitarianism, I now wish to pursue a line of argument which will (if successful) lead to a substantial modification of that form of Utilitarianism. This is initially based upon two standard objections.

Consequentialism as indirectly self-defeating.

The first objection concerns the issues just raised. It relates to the way in which consequentialist theories in general (and Utilitarianism in particular) tend to be - to use expressions introduced by Parfit (1984) - indirectly 'self-defeating' and also, perhaps, 'self-effacing'.

In the context of a consequentialist theory along the lines of Hare's Utilitarianism, the maximisation of satisfaction of informed preferences in each situation is what makes outcomes better, critical thinking is the way in which one determines the best outcome in each situation, and one is disposed to act according to general prima-facie moral principles because they will result in the best outcome in "most situations that are actually encountered". Thus far the disposition to maximise preference satisfaction and the dispositions to act according to principles conducive to maximisation of preference satisfaction seem compatible and clearly directed towards the same end.

But (to adapt the argument of Parfit, and others, to this context) most of our preference satisfaction comes from having, and acting upon, certain strong desires - these "include the desires that are involved in loving certain other people, the desire to work well, and many of the strong desires on which we act when we are not working"

(Parfit 1984 p.27). If we were disposed to always try to do whatever would make the outcome as good as possible then "we would have to act against or even suppress most of these desires".

Parfit's point is about what would be required in order to have such a disposition. For example: according to a Utilitarian theory it may be morally better at this moment if I were to stop work and telephone my family; but, perhaps, I would be disposed to act in this way in this type of situation only if my desire to work were much weaker; and, if it were, then this might generally make the outcome worse. In this way, Parfit would claim, it is likely that such a disposition would enormously reduce the sum of preference satisfaction. The moral theory may be 'indirectly self-defeating': trying to achieve the aims given by the theory may mean that those aims will be worse achieved. A disposition to maximise preference satisfaction may presuppose a weakening of 'self-regarding' dispositions in a way which reduces overall preference satisfaction.

Hare would not, I think, disagree with this; and would certainly not disagree with Parfit's further point that if we were disposed to always (or often) try to determine, and to do, whatever would make the outcome as good as possible then we would be likely to deceive ourselves about the effects of our acts. According to any **consequentialist** theory we should have (and should educate others to have) those motives and dispositions which will result in the best consequences. It is likely then that most of us should not, according to the theory, be always disposed to engage in critical thinking, and some of us should be disposed never to engage in critical thinking. A disposition to maximise preference satisfaction may, through self-deception, tempt us to stray from principles and thus reduce overall preference satisfaction.

Parfit's point is that our 'self-regarding' dispositions and our dispositions to obey principles may need to be strong enough to ensure that we act accordingly even when we **know** that an alternative action would maximise preference satisfaction.

If we believe the moral theory in question, and if it is indirectly self-defeating then, as Parfit (1984 p.49) points out, "we shall sometimes knowingly act wrongly according to our own theory" but "we can believe these to be cases of blameless wrongdoing" because "we are acting on a set of motives that it would be wrong for us to cause ourselves" [and others] "to lose". However, he goes on to say (1984 p.40), it may then be the case that "we would not in fact continue to regard morality with sufficient seriousness" and "our desire to avoid wrongdoing might be undermined if we believed that other desires should often be stronger". If this were so, it might then be claimed that it would make the outcome better if we did not believe the moral theory. The theory "would tell us to believe, not itself, but some other theory"; it would be 'self-effacing'.

From an educational perspective, a consequentialist Utilitarian theory may thus be seen to require the educator to regard the developing dispositions, desires, beliefs and emotions of the educatee as simply instrumental and to aim that they should be divorced, to varying degrees, from the aims given by the moral theory. The extent of this separation, and the proportion of educatees to which it applied, would depend upon the possibility of development of abilities to ascertain the full facts of a situation and consequences of action, to avoid self-delusion, to step out of projects without reduction in commitment, to overcome repugnance on some occasions without losing it on others, and so on.

Many philosophers have pointed out these, and similar, features of Utilitarianism and have believed that the presence of such features renders the theory unacceptable. Williams (1985 p.108) argues that the theory requires us to have dispositions, feelings and judgments which are at odds with the theory and are purely instrumental, but that the agent cannot see them in this way - "there is thus a deeply uneasy gap or dislocation between the spirit of the theory itself and the spirit it supposedly justifies". Mackie (1977 p.130) argues that the theory is unrealistically demanding: we cannot expect people to have the happiness of all as their goal and "it is too much to expect that the efforts of all members [of a community] should be wholly directed towards promoting the well-being of all" - it is either a fantasy morality or (again) it has to sanction and recommend goals which are not those given by the theory.

Parfit does not believe that if a theory were indirectly self-defeating or partly self-effacing then that would, in itself, render the theory unacceptable. Sidgwick would certainly agree and makes the distinction between the two groups of people on the basis of such considerations. But whether we agree with that will depend, as Parfit says (1984 p.29), on our views as to the nature of morality and of the criteria for determining the best moral theory.

When Williams (1985 p.108) criticises the "deeply uneasy gap" between the spirit of the theory and the spirit it justifies, he claims that the latter does not merely involve strong dispositions to **act** in certain ways (eg. to tell the truth). Such dispositions will "do the job" (ie. ensure action in accordance with the principles) only if they are associated with dispositions "of feeling and judgment" and these dispositions "are expressed precisely in ascribing intrinsic and not instrumental value to such things as truth telling, loyalty, and so

on". Thus the **motives** which give rise to action do not relate to the outcomes which the theory claims have intrinsic value.

Hare responds to this type of objection by claiming that it **is** psychologically possible to take on board the two-level approach to moral thinking which is outlined in his theory. We can have strong dispositions (strong enough to 'do the job') even though we see those dispositions as purely instrumental. We can, Hare says (1981 p.52), take this attitude to our dispositions in just the same way as a good general can be strongly disposed to, say, concentrate his forces whilst seeing that disposition as good only because it is generally conducive to the overall aim of victory.

This may be an adequate response to the criticism above but it is not adequate as a response to the deeper objection which may lie behind it. It may be wrong to claim (as Williams does) that it is **always** the case that the required dispositions will have sufficient influence only if we see truth telling etc. as having intrinsic value; but we could plausibly claim that this is true of most (or many, or some) people. If it turns out that most (or many, or some) people are not able to be like Hare (or a good general) then the theory requires that we educate such people in a way which ensures that their **motives** for action are entirely divorced from the outcomes which the theory claims have intrinsic value.

The objection here is not just about whether it is, in fact, the case that we are required by the theory to educate significant numbers of people in this way. If it were then Hare's claim that we can (all?) be like the good general may be reassuring. The objection is that the theory requires us to consider such facts when deciding how to educate. It requires us to consider whether people would best achieve the outcome of

maximising preference satisfaction if they were educated so as to be motivated entirely by other considerations. To consider such facts is to see the moral worth of the individual as entirely a matter of how conducive each is to that end. It is to see the capacities, dispositions, beliefs, desires, emotions and motives of the individual entirely as means and as having no intrinsic moral worth.

I wish to investigate the possibility of elaborating a theory which (like Hare's) implies that there are two levels of moral thinking and that the right action in each situation is that which maximises the preference satisfaction of those involved, but which also implies that certain ways of responding to those preferences have **intrinsic** moral worth.

Malevolent preferences.

The second objection concerns the way in which Hare's moral theory "makes us give weight to bad desires (such as the desire of a sadist to torture his victim) solely in proportion to their intensity".

Hare (1976 p.30) responds to such objections by claiming that they are based upon intuitive principles which deal with cases likely to be encountered and that we are most unlikely to encounter a case in which utility will be maximised by letting the sadist have his way. This for three reasons: "the suffering of the victim will normally be more intense than the pleasure of the sadist"; "sadists can often be given substitute pleasures or even actually cured"; "the side-effects of allowing the sadist to have what he wants are enormous".

In a response to Harsanyi, Hare again emphasises the claim that there will always be, in actual cases, a better alternative then that which panders to the

preferences of sadists. Harsanyi (1988 p.96) claims that if, in a given society, the number of Nazis is large enough in relation to the number of Jews then we would, according to Hare's moral theory, "have to conclude that the social-utility maximizing policy will be to kill all Jews". Hare (1988 p.245-6) responds by insisting that in order to make the situation such that the conclusion would follow we would have to "adjust the case in a way bordering on fantasy"; that, in Germany as it was, in order to carry out the massacre "the whole apparatus of totalitarian dictatorship .. was a precondition, and that was certainly not optimific"; and that, in all actual cases, "there will be a better alternative policy .. namely to push our institutions in the direction of the abandonment of harmful pleasures and desires, and hope that those who now indulge in them will soon change their ways".

But the objection is not just about the possibility of it being, according to the moral theory, morally right to perform a sadistic act in certain circumstances. If it were then Hare's claim (which, I believe, rests on an over-optimistic view of the prevalence of sadistic inclinations and our ability to redirect them) that this is not at all likely to happen may be reassuring. The objection is about the fact that the theory requires us to **give weight** to such desires when deciding what is morally right. As Williams says (1985 p.87), the fact that "racists get some satisfaction out of the sufferings of Jews ... cannot be a consideration **at all**".

Harsanyi (1988 p.96) also claims that 'anti-social' preferences should be given zero weight and that to give them weight is, in fact, at odds with a Utilitarian moral theory. He claims that "a Utilitarian is presumably a Utilitarian out of benevolence to other people; and, being a benevolent person, he can no doubt rationally

refuse to cooperate with anybody's malevolent preferences".

If, however, the Utilitarian, **for whatever motive**, aims to maximise preference satisfaction and if that outcome is what, according to the theory, determines the morally right action in each situation then this will not do. If this is the nature of the moral theory and if, on some occasions, the 'best outcome' requires the satisfaction of malevolent desires then that is the right thing to do - if our benevolent motive stands in the way of our counting the malevolent preference then, according to the theory, it ought not to do so.

Harsanyi (1988 p.97-98) goes on to offer, what he regards as, a more fundamental argument. In effect he claims that the aim is not to maximise preference satisfaction but to maximise the satisfaction of **'personal'** preferences (as contrasted with 'external preferences'). He argues that not only should socially undesirable malevolent preferences be given zero weight but the same is true of socially desirable supportive preferences. This because "Utilitarian morality requires us to respect people's preferences about how they **themselves** ought to be treated .. it should not require us to respect their preferences about how **other** people ought to be treated". And because "the fundamental Utilitarian principle that our social utility function must give the **same weight** to every individual's interests" would be defeated if the preferences of those with many well-wishers were thereby given greater weight.

Dworkin (1977 p.105) also expresses a belief that external preferences would represent a threat to egalitarianism and claims that this represents a major difficulty for Utilitarianism which "owes much of its popularity to the assumption that it embodies the right of citizens to be treated as equals". He goes on to

claim that this is a difficulty which is not easily resolved since personal and external preferences are inextricably linked together.

In response to Dworkin I shall argue later that malevolent preferences can be discriminated. In response to Harsanyi I would suggest that the question whether the fundamental principles he identifies are part of "the very nature of Utilitarian ethics" is not very interesting. We could equally claim that it is the 'very nature of Utilitarian ethics' to aim for maximisation of **non-malevolent** preferences. The interesting question is whether a moral theory which incorporates such an aim is acceptable. In answering this, I think, the question of **motive** does become central. Perhaps we should look more closely at the claim that the Utilitarian is a Utilitarian out of benevolence.

As I said at the end of the previous section, I wish to investigate the possibility of elaborating a moral theory which retains a Utilitarian view of the rightness of actions but which implies that certain ways of responding to preferences have intrinsic moral worth. It may then be the case that those responses involve a rejection of malevolent preferences. Perhaps we can see primary moral worth as lying in our achieving certain forms of benevolence.

The route to that theory requires a consideration of 'second-order' preferences and their implications for Hare's account of critical thinking.

CHAPTER 8.

Preferences about preferences and ideal selves.

Second-order preferences.

Each of us has many desires - to have a rest, to eat an apple, to be better at our work - which we may call 'first-order desires; but we may also have desires about those first-order desires - that we should lack a desire to smoke cigarettes, that we should have a stronger desire to practise playing the piano, that we should have a weaker desire to retaliate when hurt - which we may call 'second-order' desires. Frankfurt (1971) makes this distinction and goes on to claim that the possession of such desires is a peculiar characteristic of humans and is a manifestation of our capacity for self-evaluation.

The notion that self-evaluation is a distinctive feature of human agency is explored again by Taylor. He considers a further distinction between "two broad kinds of evaluation of desire" (Taylor 1985a p.16). In the first, which he calls 'weak' evaluation, we are concerned primarily with outcomes; for example, considering which of two desired objects attracts us most, or which is the most convenient of two desired actions, or how to make different desires compossible, or how to get the most overall satisfaction. In the second, which he calls

'strong' evaluation, we are concerned with the quality of our motivation; for example, classifying desires and motives (as higher or lower, noble or base), or judging them as belonging to qualitatively different modes of life (fragmented or integrated, courageous or pusillanimous).

But weak evaluation may also be concerned with desires; as, for instance, when I want to lose my desire for cigarettes so that my health will improve. In such cases we are not making a "qualitative distinction of the worth of motivations" (1985a p.18). Where weak evaluation is concerned with desires that is only on the grounds that one desire (to smoke) is 'contingently incompatible' with a more desired alternative (to be healthy).

We may be tempted into redefining issues involving strong evaluation so that we see them as, instead, involving this sort of contingent incompatibility. For example, it may be that I wish to lose my desire for cigarettes because I believe that an addiction to nicotine is unworthy, base and degrading (Taylor's example is cream cakes). I may then be talked around to seeing this in terms of my desire for health and as a question of quantity of satisfaction (1985a p.22). Someone who had a 'reductionist' Utilitarian perspective, based upon the view that all that matters is the quantity of satisfaction of the desires we in fact have, would have to talk us around in some such way, or claim that our evaluation was groundless. Taylor wishes (as do I) to reject this reductionist Utilitarian perspective.

Taylor rejects that perspective because, he claims, it either leaves out of account a dimension which is essential to the notion of human agency; or because it implicitly appeals to such a dimension. Strong evaluation involves characterizing desires as higher or lower, more noble or base, etc. To characterize a desire

Preferences about preferences and ideal selves.

in this way "is to speak of it in terms of the kind of quality of life which it expresses and sustains. I eschew the cowardly act because I want to be a courageous and honourable being." (1985a p.25). The strong evaluator examines the different possible modes of being of an agent; he is not simply concerned with satisfaction of the desires he in fact has, he is also concerned to be a certain type of person. If the Utilitarian leaves this dimension of human agency out of account then he gives a hopelessly shallow account of what it is to be human.

Perhaps, also, "we might hold that the most hard-bitten Utilitarians are themselves moved by qualitative distinctions which remain unadmitted, that they admire the mode of life in which one calculates consciously and clairvoyantly as something higher" (1985a p.23). In this case there is an implicit appeal to a dimension of strong evaluation which is not acknowledged. Such a person would be suffering from an illusion as well as from shallowness.

However, Taylor does not consider the possibility that the Utilitarian might acknowledge the fact that as humans we may yearn to be other than we are, but yet insist (explicitly and in the language of strong evaluation) that nothing is more noble or worthy than to strive to be a person who endeavours to ascertain the consequences of his actions in order to act in the best interests of all concerned. In acknowledging such a dimension the Utilitarian would have to take (as we shall see) a very different approach to the evaluation of consequences but he would not be any less 'deep' than someone whose strong evaluations closely reflected Taylor's.

Taylor goes on to claim (and, again, with this I can agree) that those who make 'strong' evaluations are concerned not only with the satisfaction of those desires which they in fact have "but also with what kind of life,

Preferences about preferences and ideal selves.

what quality of agent they are to be". Furthermore, "our identity is defined by [such] fundamental evaluations .. shorn of these we would lose the very possibility of being an agent who evaluates .. we would break down as persons, be incapable of being persons in the full sense" (1985a p.34). A moral theory which took no account of the fact that we can and do strive to be other than we are, or which tried to insist that this should always be seen only in terms of the struggle to satisfy the desires we in fact have, would indeed be shallow.

The question then arises as to what sort of person we should strive to be. Taylor (1985a p.36-38) speaks of our struggle to give form to our sense of "what we hold important" and of "what is of decisive importance". He believes that such a struggle can reveal a self which is authentic. He claims that I can define an identity for myself that is **not trivial** only against a background of things which matter in a way which transcends the self (1991 p.40), that I can find genuine fulfilment only in something which has significance independently of me and my desires (1991 p.82). An authentic self, according to Taylor, arises out of a sense of such significance and is thereby able to achieve genuine, not merely personal fulfilment.

Taylor's view involves an evaluation not merely of my own preferences but of the preferences of all. It rests upon a contrast between those preferences I have and those preferences which **are important**, rather than upon a contrast between those preferences I have and those preferences which **are important to me.** But we do not need to appeal to universal authenticity in order to make space for the struggle to be other than we are. That struggle can be based upon a sense of what each of us, personally, hold to be important. We need to begin, I believe, by making a distinction between second-order preferences which involve evaluation of **personal**

Preferences about preferences and ideal selves.

preferences and those which involve evaluation of the preferences of everybody.

Personal second-order preferences.

Each of us may wish to be other than we are - more courageous, more cautious, more steadfast, more spontaneous, less malevolent, less scrupulous - and we may have such desires without having any desire that everybody should be that way. We have **personal** second-order preferences. We may wish that we lacked some of our preferences, or that some of our preferences were weaker or stronger than they, in fact, are. We can imagine ourselves with these altered preferences and prefer, in fact, to be that way.

We may prefer to be other than we are because we believe that would result in our achieving greater preference satisfaction overall - the smoker who sees the preference for cigarettes as an obstacle to good health. But we may also prefer to be other than we are simply because that is the way we are - the smoker who would prefer not to have a preference which was due to addiction and who would prefer not to have that preference even if smoking was not conducive to ill health. We may have second-order preferences which, like many of our first-order preferences, are not grounded in further reasons.

Such preferences may, of course, be irrational. Brandt (1979) uses 'rational' to refer to "actions, desires, or moral systems which survive maximal criticism by facts and logic". Hare (1981 p.215) adds that we might use 'irrational' to refer to judgments which would have become different had they been more exposed to facts and logic. If we take this line then we might say that a second-order preference is rational if we retain it having considered all the facts - including those which

Preferences about preferences and ideal selves.

relate to overall preference satisfaction (insofar as it is possible to discover them). But one may then still (after such consideration) prefer to have a preference even regardless of a likely lessening of overall preference satisfaction. Rationality does not require that we ground our second-order preferences in **further** reasons nor does it require reasons which relate **exclusively** to greater overall preference satisfaction. One just does prefer, say, to be a person who does not have a preference to flee at the first sign of danger rather than to be a person who has those preferences which are likely to result in greater (first-order) preference satisfaction.

[Choices between actions may, **perhaps**, be represented as choices between sets of consequences. To choose one set is to have a stronger preference for that act as a against the others, and is to have greater preference satisfaction if that act is performed. The reductionist Utilitarian perspective will aim to see choices between preferences in the same way - as between the sets of consequences of having alternative sets of preferences. To choose one set of consequences is to have a stronger preference for the possession of that set of preferences, and is to have greater preference satisfaction if that set of preferences is possessed. The position outlined above would then be incoherent. But this is to leave out of account that we may choose not just between sets of consequences but also between the sets of preferences themselves. The notion of a preference is not exhausted by the consequences of having that preference in the way that the notion of an action may, **perhaps**, be exhausted by the consequences of performing that action. To have or lack a preference is to **be** a certain sort of person, to be motivated in a certain way - eg. not driven by timidity.]

Preferences about preferences and ideal selves.

Possession of preferences is not something which we, as rational, must subject to a decision procedure involving an assessment of consequences and of overall satisfaction. To assess in this way is already to make an (implicit or explicit) appeal to a second-order preference - namely, that one should have those preferences which are most likely to give the greatest preference satisfaction overall. I may have such a preference but also have a preference that, say, my preferences should not include those relating to timidity - and the latter preference may be stronger than the former. We may here speak of a comparison of the strength of such preferences; but this is not to compare the strengths of my preferences for this or that set of consequences, but rather to compare strengths of my preference for being this or that sort of person.

Each of us may have a view as to the sort of preferences we prefer to have, the sort of person we wish to be. In this context it may be appropriate to speak, as Kierkegaard speaks (1843 Vol II p.263), of one's 'ideal self'. Such an ideal self is a goal towards which one may strive, it is "a picture in likeness to which [one] has to form [oneself]". We may have a 'personal ideal self' and we do not, as rational agents, have to justify that ideal by means of a determination of the consequences (for satisfaction of our first-order preferences) of becoming that ideal self. Such an ideal may be (as Taylor would claim) fundamental to our personal identity.

[One might also have a view as to the sort of preferences everyone should have and claim, as I shall, that such a 'universal' ideal self may be fundamental to morality.]

But given that one has a personal ideal self involving second-order preferences then in what way does that affect decisions about **actions**? Can we describe those

Preferences about preferences and ideal selves.

decisions in terms of strengths of preferences and awareness of consequences; and can we apply similiar criteria of rationality (in terms of degree of exposure to facts and logic) to such decisions as we have applied to decisions involving only first-order preferences?

Decisions involving second-order prefernces.

If one is striving towards a personal ideal self, if one does not now have the preferences which one's ideal self would have, then one will wish to deliberate and to act as would one's ideal self. However, those preferences, say, which one prefers not to have will be present and will sometimes have great strength.

Suppose I have a preference that A should happen (I smoke a cigarette), call it P(A), and a preference that I should not prefer A, call it P(notP(A)); and that, in general, P(notP(A)) is greater than P(A). But suppose that, in a particular situation S (I am having a drink and my favourite brand of cigarette is available), P(A) would be greater than P(notP(A)). It is certainly possible for me to intentionally not bring about A in S - for example, by avoiding S. But would that be rational? I avoid S because I know that in S P(A) is greater than P(notP(A)), I would therefore bring about A in S (assuming greater preferences do, by definition, outweigh lesser), I would therefore act as a result of having P(A), I do not want to do that, and so I avoid S. Is that rational?

Compare this with a conflict between two first-order preferences. Suppose I have P(A); A (watching television) always has further consequences not B (not studying); I have P(B); and, in general, P(B) is greater than P(A). But suppose that, in a particular situation S (an especially entertaining TV programme is broadcast, I

know about it, a TV is available), P(A) would be greater than P(B). Should I, as rational, avoid S? I know that in S P(A) is greater than P(B), I would therefore bring about A in S, I would therefore bring about not B, I do not want to do that, and so I avoid S. This surely is **not** rational.

If I **know** that, in S, P(A) is greater than P(B) then (using Hare's epistemological premiss) I **now** have preferences P(in S,A) and P(in S,B) and the former is greater than the latter. I therefore now prefer/prescribe 'in S,A' and, if this were the whole picture, then I would have no reason to avoid S. I will, in S, bring about A and, therefore, not B but this is what I now prefer and what I will prefer in S.

However, given such a conflict, there may well be reasons either for not having such a preference for S or for avoiding S. It may be that I do not **know** that, in S, P(A) would be greater than P(B); I believe this to be the case (and it is here assumed to be the case) but believe that my judgment is not wholly reliable; I do know that P(B) is, **in general**, greater than P(A); and I, therefore, base my preference for S upon that knowledge and prefer/prescribe 'in S,notA'. If this were the whole picture then I would not yet have a reason to avoid S. But it may be that, although I **now** prefer 'in S,notA', I believe that **in S** P(A) would be greater than P(B) - this because, for example, the possibility of immediately satisfying P(A) would blind me to the consequences of not satisfying P(B) and my preferences would be worse informed in the actual situation. I now have reason to avoid S; I prefer/prescribe 'notS'.

If, for simplicity, we leave out of account other possible consequences of avoiding S and also possibilities other than S and notS then I, rationally, do not avoid S if I believe that I can act and **will** act,

Preferences about preferences and ideal selves.

in S, in a way which will, overall, satisfy my best informed preferences. I, rationally, avoid S if I believe that I will, in S, act contrary to my best informed preferences.

If the conflict between a preference and an ideal self involving a second-order preference were entirely parallel **and I was fully informed (both now and in S)** then I would have no reason to prefer 'in S,notA' or to prefer 'notS'. In S P(A) is greater than P(notP(A)); I know that and it is therefore now the case that P(inS,A) is greater than P(inS,notP(A)); I therefore now prefer/prescribe 'inS,A'. I will in S bring about A; and I will therefore act in a way which will overall satisfy my best informed preferences.

Yet, surely, if I wish to lose my desire for cigarettes because I believe that an addiction to nicotine is unworthy, base and degrading then I will prefer that **'not** S' **precisely because** I know that in S P(I smoke a cigarette) is greater than P(notP(I smoke a cigarette)). I do not want my actions to be motivated by that preference, my actions would be so motivated in S, therefore I wish to avoid S. **But** this will not do for it is just to repeat that P(in S,notP(A)) gives me reason to avoid S (as would P(B) in the alternative example) and we have assumed that P(inS,A) **is greater.**

We could at this point simply state, with J.White (1990 p.30), that second-order preferences are 'more important' and 'count more' or, with Raz (1975 p.132) that they 'always prevail'. But why should they count more or prevail? White says that they count more because they regulate other desires. But that is just to say that second-order preferences count more because they are second-order preferences. Do they always count more; should very weak second-order preferences always prevail even when in conflict with very strong first-order

preferences? Can we not find a way of describing the conflict between second and first-order preferences, which like the description of conflicts between first-order preferences, appeals only to comparisons of strength and the satisfaction (or frustration) which results from alternative actions?

Fortunately we can point to a further preference which has not, thus far, been mentioned and which does, I believe, offer such a way of describing the manner in which we can rationally avoid S (and will if fully informed). If, given all the above, I bring about (or do not avoid) S then I do so because I believe that I can and will act, in S, in a way which will overall satisfy my preferences - **because** P(A) will be stronger and will be satisfied. **To bring about S for this reason** is to **now** fail to satisfy P(notP(A)). The preferences which in S affect my decision with regard to bringing about A are P(A) and P(notP(A)) but the preferences which now affect my decision with regard to bringing about S are P(inS,A), P(inS,notP(A)) **and** P(now,notP(A)).

Furthermore, at each step in which there is a possible action which would bring S closer I will (if fully informed) have a similar additional preference: P(now,notP(A)),P(inS',notP(A)),P(inS'',notP(A)),....., P(inS,notP(A)). In order to smoke a cigarette after dinner tonight (when the satisfaction would be very great) I must fail to throw away my cigarettes now and later (S'), I must fetch them after dinner (S''), etc.; and each such action or deliberate inaction, which is a mere step towards the overall satisfaction of P(inS,A) as against P(inS,notP(A)), is a failure to then satisfy P(notP(A)) and frustrates **that** preference. Each failure is a betrayal of my preference to become my ideal self, it is a failure to overcome my actual self.

Preferences about preferences and ideal selves.

[Of course it may be the case that by preventing S′ or S′′ or ... I will **not** succeed in preventing S - later steps in the sequence of situations in which there is a possible action which will bring S closer may still occur (friends with cigarettes may arrive). If this is so (if I know that they **will** occur despite my action) then keeping my cigarettes was not a necessary condition of S, it need not then be motivated by a preference to bring about S (in which I gain overall satisfaction of P(inS,A) as against P(inS,notP(A))), I do not therefore satisfy P(now,notP(A)) by throwing away the cigarettes. In order to overcome the fully informed preferences of my actual self and satisfy P(now,notP(A)) I will need to prevent a situation which is (together with my possible action) a sufficient **and** a necessary condition of S - I may need to prevent all contact with cigarettes until the moment of temptation (in which P(A) would be greater than P(notP(A))) has passed.]

To take a different example: suppose I have a desire to purchase sexual gratification and in order to do that I must leave my house, take out the car, drive to the town, cruise the streets, stop the car, roll down the window etc.; and further suppose that I wished that I lacked that desire; if the latter is the case then each step I take is a failure to satisfy P(now,notP(A)); each step is taken in order to satisfy P(A). If I, nevertheless, bring about S and A then the inability to stop and think, or the strength of P(A), must be great indeed. It is not sufficient that P(inS,A) is greater than P(inS,notP(A)). A second-order preference should not merely be balanced against a first-order preference in each situation in which the latter may be satisfied; the second-order preference should affect my decision whenever it is possible to act towards bringing about or preventing such a situation.

Preferences about preferences and ideal selves.

Furthermore, some of the conditions which are necessary to bringing about such a situation are cognitive: in S I reach out, pick up and light the cigarette because I know that it is here and because I know that P(A) is greater than P(notP(A)). I can prevent S by failing to have that knowledge. I may refuse to possess such knowledge, refuse to acknowledge that the cigarette is here or that P(A) is the greater preference. If such knowledge is necessary to S and possession of such knowledge is motivated by P(A) then to 'refuse to face facts' may be a means to the satisfaction of P(now,notP(A)). I gain and accept such knowledge because it is relevant to P(A) - "look the cigarette is here, you do want to smoke it more than you want not to want to smoke it". I rationally refuse to acknowledge the facts because their acknowledgement would be motivated by P(A). This may, when all other conditions are met, be the only way in which I can satisfy P(now,notP(A)).

The tempter helps me to bring about the opportunity to gain overall satisfaction from my unwanted desire and, having brought me thus far, bids me to consider the facts. From the perspective of the person I in fact am, I am irrational if I do not succumb. But from the perspective which includes my second-order preferences, I will want to resist at every step and, finally, I may close my eyes to facts which would be irrelevant or false if I were the person I wish to be. The rational agent who has an ideal self will strive to **imaginatively identify** with that self - to deliberate as if he had the preferences of that ideal self - and he may acknowledge in his deliberations only those facts which would be relevant to those preferences.

This analysis would equally apply to second-order preferences which are not based upon an ideal self - ie. which relate to overall satisfaction of first-order preferences. If, in general, P(B) is greater than P(A)

Preferences about preferences and ideal selves.

then possession, and knowledge, of P(notP(A)) might make it possible and rational to avoid all situations in which A was possible (and P(A) was strong enough to make temptation likely). Furthermore, avoiding all such situations might, in fact, result in my losing P(A) and thus remove the dissatisfaction which is now associated with situations in which P(B) is greater than P(A) and I do not bring about A. [This is a further reason for avoiding such situations which was not mentioned above.]

But, alas, it does not appear to be the case that we can 'adopt' a second-order preference simply because it would be rational to prefer to have such a preference - P(P(notP(A))) does not entail P(notP(A)). The acquisition of such a preference is not likely to result from deliberation, it is more likely to be the result of a fundamental change (revelatory or traumatic) in one's identity; as when the alcoholic suddenly sees himself for what he is - driven by addiction - and, seeing this, recoils from himself.

This may also be true of second-order preferences which **are** a feature of an ideal self. An ideal self may be the result of upbringing or education, or of traumatic change, or - as Taylor might claim (1985a p.42) - of a radical shift in identity which stems not from a mere radical 'choice' but from self-reflection which brings form to "those inchoate evaluations which are sensed to be essential to our identity".

Critical thinking and a personal ideal self.

The presence of personal second-order preferences as a central feature of (some of) our lives has clear implications for a moral theory which attempts to incorporate some commitment to critical thinking. If critical thinking involves imaginative identification

with the preferences of others then it involves identification with first-order **and** personal second-order preferences.

Let us assume that the person possessing a second-order preference is rational and has, therefore, considered what it would be like to have the preference he wishes to have (including consideration of the likely consequences for preference satisfaction), ie. the second-order preference is an 'informed' preference. If the person then retains that second-order preference (even despite a possible lessening of overall preference satisfaction) we may conclude that that preference is as important to him as the reductionist Utilitarian perspective is to one who is, or wishes to be, such that only maximisation of first-order preference satisfaction matters. Its importance lies in the fact that it is about being a certain type of person; it is not just about the decisions made on this or that occasion.

Thus, if there are, on certain occasions, desires (I want to hit him so much) and beliefs (he is over there) which from a first-order perspective would move that person to act in a way contrary to the way he would act if he were the person he wishes to be, then those desires and beliefs are obstacles to him. Insofar as I imaginatively identify with him they should be obstacles to me too. But I have a real practical advantage: by identifying with his ideal self (ie. with the first-order preferences he **wishes** to have) I can avoid those obstacles (the desire is absent and the belief is irrelevant). I do not have to overcome temptation in **my** deliberation. I can identify with the self he would be if he were to overcome his actual self.

[I may go further and refuse to acknowledge that he has not yet achieved his ideal self. I rationally refuse to help him 'face the facts': "no, that is not a cigarette",

Preferences about preferences and ideal selves.

"you do not want it, you would not enjoy it". In this way I can help him to raise obstacles to his actual self and to satisfy his preference to be other than he is.]

It is practically easier to imaginatively identify with the first-order preferences of someone's ideal self - rather than to identify with the first-order and second-order preferences of his actual self. Second-order preferences are a feature of our lives and the simplest way in which to take account of them when engaging in critical thinking is to **imaginatively identify with the personal ideal self of each person.**

Critical thinking, thus modified, does not merely require that we take account of the preferences of others for these or those consequences, it also requires that we take account of **their** views as to what preferences they prefer to have. It incorporates a tolerance and respect for the personal ideals of others. We can accomodate the fact that such views may be fundamental to our identity without abandoning an essentially Utilitarian perspective.

Taylor, however, would seem to equate having second-order preferences (involving strong evaluation) with having a view as to what preferences **others** should prefer to have.

Taylor says (1985b p.237): "some ways of living have a special status, they stand out above others"; to recognise the "higher value" of, say, integrity is an essential part of our .. having integrity"; the "aspiration to achieve [such a] good is also an aspiration to be motivated in a certain way"; such an aspiration involves a second-order motivation. If 'higher value' were to mean 'higher value **to me**' then this could be interpreted as referring to **personal** second-order preferences.

Preferences about preferences and ideal selves.

But Taylor also says (1985b p.237-8): "an essential part of achieving liberation is sensing the greatness of liberated humanity"; "ordinary goals, for instance for wealth and comfort, are goals that a person may have or not ..[but] it is in the nature of what I have called a higher goal that it is one we **should** have"; "for those who subscribe to integrity, the person who cares not a whit for it is morally insensitive, or lacks courage, or is morally coarse".

To subscribe to such an ideal is, for Taylor, to subscribe to it as an ideal for **all mankind.** It is to feel admiration for those who either strive for or achieve that ideal and contempt for those who do not (Taylor 1985b p.239). In resisting the reductionist Utilitarian perspective (according to which such ideals must be construed in terms of degree of attainment of first-order preference satisfaction) Taylor makes the very strong claim that such ideals must be construed in terms of ideals for humanity as a whole. Taylor may have other reasons for making that claim but, I have argued, we can resist the reductionist Utilitarian perspective without also insisting that ideals relating to, say, liberation or integrity (or - with Aristotle - courage, temperance, liberality, gentleness, wittiness, modesty, etc.) have to be such that we may legitimately impose them upon all.

This is not to deny the significance, or possible legitimacy, of ideals for humanity as a whole. Indeed I hope to found a version of Utilitarianism upon such an ideal. But that ideal will not require that we sweep aside all those personal ideals of others which differ from ours, rather it will underpin a demand for that critical thinking which incorporates a tolerance and respect for the personal ideals of others.

Preferences about preferences and ideal selves.

Universal second-order preferences.

I have thus far tried to show how one could use the method of critical thinking to arrive at judgments as to how to act in a given situation, and yet accommodate the fact that personal second-order preferences are an important feature in our lives. Critical thinking, as thus modified, represents the Utilitarian view that the preferences of each person matter, but it extends that view to incorporate not only preferences for what happens in the world but also preferences about the sort of person each of us prefers to become.

The fundamental idea behind this accomodation has been that judgments of action need not be made on the basis of what would maximise satisfaction of our **actual** preferences but, rather, can be made on the basis of what would maximise satisfaction of the preferences of our ideal selves.

But now it may be that a first-order preference which someone has (either as a preference of their ideal self or - if ideal and actual selves are the same - as a preference of their actual self) is a preference which I would prefer **them** not to have. I may wish **others** to be other than they are - more courageous, more cautious, more steadfast, more spontaneous, less malevolent, less scrupulous. I may prefer that all lacked malevolent preferences, or that no-one had an overwhelming desire to avoid danger, or that each had no pity. I may have **universal** second-order preferences.

[I may also, because of a first-order Utilitarian perspective, prefer that all had those preferences which would be most conducive to the maximisation of their preference satisfaction or to the maximisation of preference satisfaction of all - ie. I may have a 'reductionist' universal second-order preference.]

Preferences about preferences and ideal selves.

Taylor's concept of 'strong' evaluation (as resisting reduction to a first-order Utilitarian perspective) applies equally to personal and universal second-order preferences. But we are indeed more likely to have recourse to the **language** of strong evaluation in the context of universal second-order preferences - words like 'noble', 'base', 'worthy', 'unworthy' are generally used to imply a universality of judgment.

Unless we rule out such preferences (and I shall shortly look at some of the ways in which this may be attempted) then one's views as to what type of person each of us should strive to be **may** affect one's critical thinking. I have, thus far, merely described a form of thinking ('critical' thinking) which involves imaginative identification with the preferences of others, and have discussed the implications (for that form of thinking) of our taking account of second-order preferences. I have not yet offered any reasons why one should engage in such thinking and I have certainly not offered any reasons why someone engaging in such thinking should leave out of account their own universal second-order preferences.

Critical thinking would enable us to make judgments, as to how to act in a given situation, in the light of consequences and in the light of what others want; but my universal second-order preferences may form part of the process of reaching those judgments. Hare's archangel acquires the wants of each of us by imaginative identification, has a complete knowledge of consequences and reaches a decision; but such an archangel may (like Taylor) have universal second-order preferences. A willingness and ability to make judgments by means of critical thinking is not, in itself, inconsistent with such preferences.

Preferences about preferences and ideal selves.

If I prefer that others lack malevolence then I may (as rational) not count your desire to see others suffer; if I reject timidity in others then I may not give the same weight to your desire to avoid danger as you do; if I reject pity then I may not give weight to your desire to alleviate the suffering of others. Our judgments as to how to act in a given situation would then differ greatly. Having opened up our critical thinking to strong evaluations, and refused a reduction to a first-order Utilitarian perspective, we now appear to have no means of preventing the application to critical thinking of a whole range of universal second-order preferences.

Whether there is a means of resisting this application will depend upon our reasons for engaging in critical thinking. It may be that there are particular universal second-order preferences which would underpin a commitment to critical thinking and which would still enable us to shape a version of Utilitarianism that also addresses the objections discussed earlier.

Utilitarianism and a universal ideal self.

I now wish to investigate ways in which Utilitarianism may respond to **universal** second-order preferences. I shall begin by considering two different kinds of strategies - both of which aim to maintain the link with an essentially Utilitarian theory. The first strategy insists on a reduction to a first-order Utilitarian perspective; the second founds the moral theory on a particular set of universal second-order preferences.

The first approach could argue that a failure to 'reduce' universal second-order preferences was irrational (I have already argued that this is not the case), or that it is question-begging.

Preferences about preferences and ideal selves.

Hare (1981 p.179) argues that none of the preferences which we acquire by imaginative identification "has greater dignity or authority than another"; and he goes on to argue that we cannot therefore 'boost' some preferences on the grounds that they are 'moral convictions'. To do so would be "simply to refuse to think critically". The point here is that if we are using critical thinking in order to make a judgment as to what should be done in a particular situation then boosting a preference preempts that process. To boost a preference (so that "it has to prevail") is, effectively, to insist that what should be done is whatever satisfies that preference - the judgment is already made, the process of reaching that judgment is otiose, we have begged the question.

I have suggested ways in which having universal second-order preferences may lead us to discount or not give 'proper' weight to the preferences of others when we are engaged in critical thinking. Thus we might expect a similar argument with regard to the discounting of preferences: 'none has less dignity or authority', so we cannot discount some preferences.

But it is important to note that the discounting of a preference, unlike the boosting (in Hare's sense) of a preference, does not preempt the process of critical thinking. If one were to boost a preference so that it always overrode other preferences in one's deliberations then that would have the result that all actions which satisfied that preference would be deemed right. But if one were to discount a preference in one's deliberations then that would not have the result that all actions which satisfied that preference would be deemed wrong - all the **other** preferences would still have to be considered. 'Boosting' is incompatible with a commitment to critical thinking in a way in which discounting, or giving altered weight to, a preference is not.

Preferences about preferences and ideal selves.

The fact that one would discount someone's preference for hurting people would not mean that critical thinking must yield the result that actions which hurt people are always wrong, or even that actions which result from someone's desire to hurt people are always wrong. The consequences of such an action in a particular situation may be such as to satisfy other preferences. The sadistic dentist **may** (in some circumstances) be doing the right thing when he extracts a tooth without anaesthetic.

What is true is that the dentists's sadism, his desire to inflict pain, **is not relevant** to determining whether it is the right thing. Critical thinking is not preempted, judgments about actions are not ruled out in advance. It is just that the facts and preferences which determine the results of critical thinking are not **all** the facts and not all the preferences of those involved.

Equally, when we give more, or less, weight to a preference than would the person who had that preference then the judgment is **not** already made. We have not begged the question in the way in which boosting a preference so that it "had to prevail" would beg the question.

Some universal second-order preferences may, perhaps, involve boosting (in Hare's sense) one particular preference, or discounting all but one preference (which would have the same effect). But the universal second-order preferences which I shall be considering will not do either of these things. They will not, therefore, beg the question in a way which would render critical thinking irrelevant; they will simply introduce a further element into the process of decision-making which uses critical thinking.

Preferences about preferences and ideal selves.

I conclude that a failure to 'reduce' universal second-order preferences is neither irrational nor begs the question - even when it is a feature of a moral theory which implies that the making of certain types of moral judgments involves the use of critical thinking.

The second approach, in which the moral theory is founded upon a particular set of universal second-order preferences, could take two forms (where the aim is to maintain an essentially Utilitarian theory). The first insists on a 'reduction' not because that is rational or avoids begging the question but because that is to be part of the basis of the moral theory. The universal second-order preference which would form part of the basis of the theory would be to the effect that: each of us should strive to have just those preferences which would be most conducive to the maximisation of preference satisfaction of all. A less demanding version could involve a preference that: each of us should strive to have different preferences only when that would lead to greater preference satisfaction for all. Such a theory would be coherent but would take us no closer to answering the objections to Utilitarianism which I outlined in the previous chapter.

An alternative is to found our Utilitarian theory upon a universal second-order preference relating to benevolence: the 'benevolent archangel' as a universal ideal self.

Preferences about preferences and ideal selves.

CHAPTER 9.

Two types of Archangel.

The benevolent archangel and the malevolent archangel.
The benevolent archangel as ideal self.
Non-consequentialist 'Utilitarianism'.

The benevolent archangel and the malevolent archangel.

As we saw in an earlier chapter, Hare's 'epistemological' premiss is central to his project of appealing to logic and reason rather than to a shared sentiment of universal benevolence.

If we grant that premiss then it is the case that:
if I know a preference which x has for a situation S then I now **have** that same preference for a hypothetical situation HS - where HS is identical to S in all respects save that I am x.

This premiss relates equally to one's own preferences. Thus we have:
if
1. x knows(inS,xP(A))
then (given the epistemological premiss)
2. now xP(inS,A)
if
3. inS,xP(A)
then (given HS is identical to S save that y is x)
4. inHS,yP(A)
if
5. y knows(inHS,yP(A))
then (given the epistemological premiss)
6. now yP(inHS,A)

Two types of Archangel.

If y knows 3. then he also knows 4. and, therefore, 6. is true. This link between knowledge of the preferences of others and acquisition of those preferences is at the heart of Hare's position. As Hare makes clear (1981 p.99), the premiss does **not** involve a link between my knowledge of your preference, say, to avoid suffering and my preference that **you** do not suffer; rather the link is with my preferring that **I** would not suffer if I were you. It is such preferences as these which, according to Hare, do the work and lead to a prescription which takes account of the preferences of all. It is this use of the epistemological premiss which would ensure that (first-stage) universalisability is not trivial. If it were to achieve this then it would ensure that an **appeal to benevolence** was unnecessary - all that would be required is that people be willing and able to make moral judgments in the light of logic and the facts.

I have argued (in chapter 6) that this will not do: an appeal to the epistemological premiss in the context of a **totally hypothetical** situation is fruitless. If critical thinking is to be a form of thinking that, in some way, ensures judgment which responds to the preferences of others as a result of acquisition of those preferences then this cannot be through the acquisition of preferences for totally hypothetical situations. We shall have to take a more 'traditional' utilitarian approach and look towards a form of critical thinking which rests upon preferences for the **actual** situation acquired through 'benevolent' identification with others.

Let us begin then by clarifying what it means to say that 'x acquires a preference P(in S,A) through benevolent identification with y'. There are three requirements here: y has the preference P(in S,A); x knows that yP(in S,A); x has the preference P(in S,A) as a result of that

Two types of Archangel.

knowledge. The last requirement amounts to claiming that x's knowledge of y's preference is a necessary condition of x having that preference: x would not P(in S,A) were it not for x knowing that yP(in S,A). We may call a preference resulting from such identification a 'benevolent' preference.

Although benevolent identification, thus defined, occurs when such knowledge is a necessary condition of the preference acquired, that knowledge may not be a sufficient condition. If, for example, x acquires a preference of y's as a result of knowledge of y's preference **and** knowledge that y is a Frenchman (the latter also being in this case a necessary condition of acquisition of the preference) then we nevertheless have a case of benevolent identification - x benevolently identifies with y because y is a Frenchman. There may be many other different conditions which, in a given situation, are necessary conditions of x acquiring a preference through benevolent identification with y (y is a child, x is not under stress, it is Sunday, the preference relates to food, and so on) but benevolent identification has taken place whenever x's knowledge of yP(in S,A) is a necessary condition of xP(in S,A).

We may now adopt a similar approach to clarifying what might be meant by saying that 'x acquires a preference P(in S,A) through **malevolent** identification with y'. Thus: malevolent identification has taken place whenever x's knowledge of yP(in S,A) is a necessary condition of xP(in S,**not** A). We may call a preference resulting from such identification a 'malevolent' preference; and the above initial remarks will help us in clarifying what is to count as such a preference. If we intend to discover grounds for excluding such preferences (as I hope to do) then we must be clear about what is to count as a malevolent preference. As Sen and Williams (1982 p.9) point out when discussing Harsanyi's exclusion of 'anti-

social' preferences, we shall need to consider whether we are thus excluding "preferences the satisfaction of which will as a matter of fact exclude the satisfaction of others, as in competition, .. preferences which refer negatively to other preferences", preferences based upon envy, etc.

In a competition involving x and y, x may prefer that y did not win. If x's knowledge of y's desire to win is a necessary condition of x's preference then that preference is malevolent. If, however, x would have that preference even if x did not know that y wanted to win, or even if y did not in fact want to win, then that preference is not malevolent. Similarly, whether a preference based upon envy is malevolent will depend upon whether x is envious of y having what x wants (but cannot have) or whether x is envious of y having what y wants. In the former case, it may be that x will still prefer y not to have the thing in question even if y did not want it (or would not miss it if he did not have it); such a preference is not malevolent (as here defined). In the latter case, it may be that x prefers y not to have the thing in question because, say, y has so much more of what he wants than x. If, in this case, y's preference for that thing and x's knowledge of it is a necessary condition of x's preference that y not have it then x's preference is malevolent - x malevolently identifies with y because y has so much more of what he wants. As with benevolent identification there may be many different conditions which, in a given situation, are a necessary condition of x acquiring a preference through malevolent identification with y (y has more of what he wants, y is a Frenchman, x is under stress, it is Sunday, the preference relates to food, and so on).

Malevolent preferences do not simply refer negatively to other preferences, they involve a certain motive: I want you to not have what you want **because** you want it; I want

you to have what you do not want **because** you do not want it. I want you to be frustrated and to suffer.

Hare's archangel has complete knowledge of all the consequences of alternative actions and of the preferences of everybody. I shall define a **benevolent** archangel as one who has such knowledge and **acquires** all the preferences of all those involved in each situation as a result of knowledge of those preferences - he is wholly benevolent. Likewise, the **malevolent** archangel has such knowledge and acquires preferences which oppose the preferences of all those involved in each situation as a result of knowledge of those preferences - he is wholly malevolent.

Hare's archangel universalises his moral judgment over all those hypothetical situations in which he is x, he is y, etc., he thus (according to Hare) acquires the same preferences for those hypothetical situations as x, y, etc. have for the actual situation, he then balances those preferences (in the light of knowledge of consequences) in order to make the same judgment for the actual situation and for each of the hypothetical situations, and thereby makes a judgment (and is disposed to act) in a way which takes account of the preferences of others.

The benevolent archangel does not need to universalise judgments over hypothetical situations. He already shares the preferences of others for the **actual** situation. His benevolence ensures that he makes a judgment and is disposed to act in a way which would maximise the preference satisfaction of those involved in that situation. Likewise the malevolent archangel makes a judgment and is disposed to act in a way which maximises the frustration and suffering of those involved. Or rather, this would be the case if we could assume that the benevolent archangel was wholly non-

Two types of Archangel.

malevolent and that the malevolent archangel was wholly non-benevolent.

However, the benevolence of the benevolent archangel is not logically inconsistent with malevolence. It is logically possible, as a result of knowledge of y's preference P(in S,A), for x to acquire a preference P(in S,A) **and** a preference P(in S,not A). If, for example, x were benevolent to all children but malevolent to all those born in France then x would have such opposing preferences when considering a situation which involved a French child. But there may be a sense in which we can say that such opposing standpoints are not in the end equivalent to benevolence conjoined with malevolence. Since the two preferences are equal but opposing they must, rationally, result in a form of indifference (albeit under strain) - the result of benevolent identification and malevolent identification with one and the same preference is (rationally) no preference at all.

We could define benevolence in such a way as to require that the end result of knowledge of a preference is the acquisition of that preference and thereby ensure that the benevolent archangel is not only wholly benevolent but is thus also wholly non-malevolent. However, this may not be necessary once we focus upon the benevolent archangel as ideal self. To adopt as ideal self an archangel who was wholly benevolent but who was also wholly (or partly) malevolent would surely be bizarre. The end result of achieving such an ideal would be indifference to the preferences of all (or some). To adopt such an ideal rather than to directly adopt an ideal of indifference would require a preference for bringing into one's life a greater degree of psychological strain than one already 'enjoyed'.

If we can thus set aside the ambivalent (or schizophrenic) benevolent archangel as an ideal then the

Two types of Archangel.

benevolent archangel as ideal self is not only wholly benevolent but is also wholly non-malevolent. Likewise the malevolent archangel as ideal self is not only wholly malevolent but is also wholly non-benevolent. Both are images of perfection: complete knowledge entirely at the service of perfect sympathy or perfect spite.

[This is not to say that these two are the only possible rational ideals based upon benevolent or malevolent identification with the preferences of others. We can have such an ideal which is partial in the sense of extending only to some people, situations or preferences. We can also have such an ideal which is partial in the sense of not matching in strength the preferences which are the object of identification. In the former case we can imagine, for example, that those aspiring to malevolence (not being all-powerful) might wish to have associates who assisted in bringing about the maximum frustration and suffering but who were not themselves regarded malevolently. In the latter case we can describe a spectrum of benevolent identification such that responses to a preference of given strength could range from an acquired preference of the same strength to one with strength which was some small fraction of that strength. We might then describe malevolent identification in terms of acquiring the same preference but with negative strength and as thus continuous on a spectrum with benevolent identification - so that indifference (zero strength) lies in the middle. In what follows I shall be considering an ideal of benevolence which is not partial in either of these senses.]

The benevolent archangel as ideal self.

The benevolent archangel considered here has, as an essential characteristic, complete benevolence and it is that characteristic which I wish, for now, to focus upon.

Two types of Archangel.

This in order to determine the implications for our critical thinking of adopting such an archangel as an ideal. Furthermore, in the moral theory which I wish to elaborate, it will not only be the case that the judgment of a benevolent archangel determines **what we should do** but also, and more importantly, that the judging of a benevolent archangel offers **an ideal as to how we should be.**

If I wished to be a benevolent archangel (if this were my personal ideal self) then I would wish to be wholly benevolent, to lack all malevolence, to have knowledge of the consequences of alternative actions, and to judge accordingly. I would wish to have the abilities of an archangel and to be motivated by a sentiment of universal benevolence.

However, the attempt to deliberate as if this were so is **not** inconsistent with my counting **your** malevolence. Indeed my benevolence to you, and to all your preferences, **requires** that I count your malevolent preference. My benevolence towards the victim of your malevolence will ensure that I wish him not to suffer (just as he wishes not to suffer) but my benevolence towards you will ensure that I wish the suffering to take place (just as you wish it to take place). A reluctance to see the victim suffer is not the same as a reluctance to count your desire that the suffering should take place.

As a benevolent archangel **I** would lack malevolent preferences; but to have the benevolent archangel as a **personal** ideal self gives no reason to discount the malevolence of others. I would claim, contra Harsanyi, that to be motivated by benevolence does **not** give a reason to "refuse to cooperate with anybody's malevolent preferences" (Harsanyi 1988 p.96).

Two types of Archangel.

In order to have grounds for discounting the malevolence of others I must subscribe to the benevolent archangel as a **universal** ideal. Such an ideal, and the consequent desire that all lack malevolence, gives grounds for discounting the malevolent preferences of others. Let us then take the benevolent archangel to be a universal ideal self and use this as the starting point for our moral theory. The fundamental principle of the moral theory which I shall outline is a universal second-order preference that:

we ought all to be (more like) a benevolent archangel.

If I were to subscribe to this universal ideal self then I would wish **all** to be wholly benevolent, to lack all malevolence, to have knowledge of the consequences of alternative actions, and to judge accordingly. If I were to deliberate as if this were so then I would deliberate as if I were **a member of a community of benevolent archangels.** Such deliberation **is** inconsistent with my counting the malevolence of others. Each member of such a community would (when considering alternative actions) share the preferences of all others and would lack all malevolence. The benevolent archangel as a universal ideal self gives reason to discount the malevolence of others. My subscribing to that ideal gives reason to benevolently identify with all but the malevolent preferences of others.

[Note: Deliberations based upon benevolent identification may result in what we have called a 'malevolent' preference. I may, as a result of deliberation in the light of the universal ideal of the benevolent archangel, acquire a preference that x not achieve something which he wants and that because I know that x wants it. I may, for example, believe that by depriving x, who is a thief, of something which he wants he will be encouraged not to steal in future. In this example it is a necessary

Two types of Archangel.

condition of my preference that x wants the thing in question and my preference is therefore a 'malevolent' preference. But such a preference does not **count** in my deliberation, it is the **result** of my deliberation. A malevolent preference here results from benevolent identification with all concerned; I want x to suffer or be frustrated because I have considered all alternative actions and their consequences and have benevolently identified with the preferences of all concerned. But such a process of deliberation will take no account of malevolent preferences such as my desire that the thief should suffer simply because he is a thief or because he has taken something of value to you (towards whom I am benevolent).]

The benevolent archangel as universal ideal not only gives grounds for discounting malevolent preferences it also gives grounds for an approach which rests upon fully informed preferences. Each member of a community of benevolent archangels would know the consequences of alternative actions and would thus have what Hare calls 'perfectly prudent' preferences (and what Harsanyi call 'true' preferences). To deliberate in the light of this ideal is to deliberate as if this were so. Subscribing to that ideal gives reason to benevolently identify with the 'true' preferences of others.

The process of thus idealising the preferences of all agents through appeal to a universal ideal self can also be used to shed light upon the problem of 'double-counting' referred to by Dworkin (1977 p.103-6) and others. Suppose x is benevolent towards y; z is attempting to make a judgment which (by means of benevolent identification) takes account of the preferences of all; and xP(A), yP(B), zP(C). Given the benevolence of x towards y, then xP(B) just because yP(B). If z now makes a judgment he will acquire P(B) by means of benevolent identification with y and, again, by

Two types of Archangel.

means of benevolent identification with x; z will 'double-count' that preference and may, thereby, make a judgment which favours y. Such a judgment will have been influenced by the mere fact that y is fortunate enough to have a well-wisher, whereas x and z do not. This, Dworkin claims, cannot be right.

Hart (1979 p.108-110) claims that the preferences of 'disinterested' supporters **should** be included. He supports that claim by offering an example: if the issue is freedom for homosexual relationships, and if liberal heterosexuals prefer homosexuals to have that freedom, then not counting those preferences would be 'undercounting'. The views of supporters (and detractors) should be counted; and if, as a result, the judgment is wrong then that will be because those supporters (or detractors) are not willing or not able to listen to the issues - their preferences are not informed.

However, Hart's position relies on the assumption that the supportive preference is **not** 'merely' a benevolent preference. The liberal does not prefer freedom for homosexual relationships because it is what x prefers; he supports that preference because of the nature of what is preferred not because it is preferred by x. The fact that, as Hart says, the issues should be relevant makes this clear.

Dworkin's distinction (1977 p.104) between 'external' and 'personal' preferences is not helpful here. The liberal's preference is (presumably) an external preference (it does not relate to the liberal's "enjoyment of some goods or opportunities"); and Hart is right, I believe, to insist that it should be counted. At least, it is not clear that a moral theory which requires that we count such a preference is in need of modification.

Two types of Archangel.

However, once we distinguish between a supportive preference which derives from the nature of what is preferred and a supportive preference which derives from the identity of the person supported then it may well seem more difficult to defend a theory which implies that we should count the latter.

As Harsanyi says (1988 p.98), "the interests of persons with many well-wishers and friends would obtain much greater weight than the interests of persons without such supporters". Harsanyi claims that this is objectionable because it means that we do not "give the same weight to every individual's interests" and that contradicts a "fundamental utilitarian principle".

However, as I claimed in the previous chapter, the question whether such a principle is fundamental to 'utilitarianism' is not very interesting. The interesting question is whether moral theories which have amongst their consequences the ruling out of double-counting are acceptable. The moral theory which appeals to the benevolent archangel as universal ideal self has that consequence. Against the background of that ideal double-counting is objectionable because it involves our deliberating in the light of the **actual** preferences of others rather than in the light of those preferences we all would have if we all were to live up to that ideal. Each member of a community of benevolent archangels would share the preferences of **all** others and would not merely share the preferences of **some** others - there can be no double-counting in such a community.

[Or, more precisely, there can be no 'partial' double-counting (in which I, having benevolent preferences on behalf of all others, acquire your benevolent preferences on behalf of some others). There can be 'complete' double-counting (in which I, having benevolent

Two types of Archangel.

preferences on behalf of all others, acquire your benevolent preferences on behalf of all others) but this would be pointless since it would not result in a different judgment.]

If we subscribe to the universal ideal of the benevolent archangel then an attempt to deliberate as if that ideal were realised should involve neither double-counting nor the counting of malevolent preferences. If someone were to live up to that ideal then that person would prescribe, in each situation, that action which would maximise satisfaction of all the fully informed non-malevolent (and non-benevolent) preferences of the personal ideal selves of each of the people involved. Such a person would deliberate and be disposed to act in the way in which a utilitarian (who took account of personal ideal selves, who discounted malevolent preferences, and who did not 'double count' preferences) would wish to deliberate and act.

However, in thus idealising the preferences of those to whom we are benevolent we abandon a consequentialist position. We count the preferences of each individual's personal ideal self not because they are a means to greater overall preference satisfaction, but because we believe that all ought to be benevolent towards the (second-order as well as first-order) preferences of others. We count the 'true' preferences and discount the malevolent preferences of others not because we believe that satisfaction of 'manifest' and malevolent preferences always reduces overall preference satisfaction, but because we believe that all ought to be well-informed and non-malevolent. The principle which gives rise to these views underpins a fundamentally non-consequentialist position.

Two types of Archangel.

Non-consequentialist 'utilitarianism'.

A moral theory which is based upon the benevolent archangel as universal ideal is a theory which relates to those characteristics which determine our prescriptions rather than to the consequences of acting accordingly. If we start from the principle that 'we ought all to be (more like) a benevolent archangel' then we begin with a view which measures moral worth in terms of what we are rather than what we do. Such a theory is (according to distinctions made in chapter 3) a non-consequentialist theory but it is also a theory which yields a view of what would be the right action to perform in each situation.

As in Kant's theory the focus is upon a form of judgment which is intimately linked with action. Furthermore, moral worth resides not in the performance of the action but in the exercising of that form of judgment (and thus being disposed to act). If we appeal only to our principle then an action which conforms to such judgment but yet does not arise from such judgment has no moral worth (it has mere 'legality').

To respond to preferences in the way in which a benevolent archangel would respond is to have intrinsic moral worth. That response involves a rejection of malevolent preferences, a respect for the fully-informed first-order and second-order preferences of all, a view of what is the right action which is essentially utilitarian, and a disposition to act accordingly. But, alas, it also involves a requirement that in each situation we know the consequences of all alternative actions, and that we know and share the preferences of all involved. This we cannot (or can seldom) do.

If we, who cannot be benevolent archangels, nevertheless subscribe to the universal ideal of the benevolent

Two types of Archangel.

archangel then how ought we to strive to live up to that ideal? How ought we to educate ourselves and others in the light of that ideal?

The consequentialist utilitarian (when faced with the problem of our imperfection) can point out that, although weighing preferences and consequences in each situation is the only way of always ensuring right action, we can at least behave in accordance with principles which are generally conducive to that end. This response, however, leads us into the problems associated with a self-effacing theory - the possibility that the way in which some of us can best achieve that end is to be educated so as to believe that morality is not about that end but is rather about conforming to principles. We may thus be educated in such a way as to be unaware of the aims of the educator. The capacities, dispositions, beliefs, desires, emotions and motives of the educatee may be seen entirely as means and as having no intrinsic moral worth.

Our central question is now whether, starting from a non-consequentialist position based upon the universal ideal of the benevolent archangel, we can find a way of responding to the problem of our imperfection which is both coherent and does not have similar implications.

Two types of Archangel.

CHAPTER 10.

Morality and education in the light of our imperfection.

Hare's two levels of moral thinking.
The benevolent archangel as ideal for imperfect agents.
The role of cognitive humility.
Partiality to self.
Decisive preferences and general moral principles.

Hare's two levels of moral thinking.

I shall begin by looking again at Hare's two-level approach to utilitarianism. The right action (the action which ought to be performed) in a given situation is that which maximises preference satisfaction in that situation. We cannot (or can seldom) determine which action will have that consequence but we can ensure that our actions (and those of others) are, in general, likely to have that consequence. This we do by educating ourselves and others in a way which ensures the possession of "a set of dispositions, motivations, intuitions, prima facie principles (call them what we will) which will have this effect" (Hare 1981 p.46).

Those principles will need to be selected by ourselves or others. At the 'critical' level of thinking we select such principles for action (which may be very particular or very general); at the 'intuitive' level our action is guided by such principles.

We select moral principles by balancing "the size of the good and bad effects in cases which we consider against the probability or improbability of such cases occurring in our actual experience" (Hare 1981 p.48). The good and

bad effects are, presumably, balanced by determining the strengths of those preferences which would be satisfied by acting according to the principle and comparing these with the strengths of those preferences which would be satisfied by acting in an alternative way.

If, for the moment, we consider examples from the prudential field then the probabilities and preferences will presumably be balanced in much the same way as, say, Skyrms (1975 p.153-155) outlines the balancing of probabilities and values of consequences. The situation which he outlines is one in which: you are to guess whether someone (in his example the queen) is over 40 or not, if you guess correctly then you will be given 1000 dollars, if you guess that she is 40 or younger and she is over 40 then you will win nothing, if you guess that she is over 40 and she is 40 or younger then she will have your tongue cut out, you value your tongue at 1000000 dollars, and the probability that she is over 40 is (on the basis of all the evidence available to you) 0.9. In the model offered, a decision is then made by determining the 'expected values' of alternative courses of action.

	Consequence	Probability	Value of the consequence	Expected value
Guess over 40:				
	correct	0.9	1000	900
	too high	0.1	-1000000	-100000
	too low	0	0	0
Guess 40 or under:				
	correct	0.1	1000	100
	too high	0	-1000000	0
	too low	0.9	0	0

The expected values are determined by taking the product of the probability and the value of the consequence (or, we might say, the strength of the preference). As Skyrms

(1975 p.154) says, "by guessing that .. [she] is 40 or under, you have a smaller chance of winning money, but you eliminate the possibility of losing your tongue" and (we need to add) the preference for winning the money is quite strong but the preference for not losing your tongue is much stronger. If prudent choices are those which are based upon knowledge of the probabilities and values of consequences then the prudent choice will be to act in a way which maximises expected value ie. to guess that she is under 40.

In the context of establishing a **general** principle the assigned probabilities will reflect our estimate of the frequency of different consequences given a choice of action in that type of situation. They will be based upon, using Ayer's terminology (1972 p.27-28), 'statistical judgments' which apply to sets of persons or situations, and not upon 'judgments of credibility' which apply to individual persons or situations. For example, if members of the royal family (wholly disguised) often ask me to guess their ages, then the assigned probabilities which determine my derived preferences might, say, be based upon my knowledge of the proportion of royals who are over 40.

In such a model the choice of principle is based upon maximising expected value. This is the sort of model which Hare seems to recommend (1981 p.156): "the method to be employed is one which will select moral principles for use at the intuitive level .. on the score of their acceptance utility, ie. on the ground that they are the set of principles whose general acceptance .. will do the best, all told, for the interests" of all.

However, elsewhere Hare talks of selecting those moral principles which will yield actions having the "greatest possible conformity to" (1981 p.46), or "most nearly approximating to" (1981 p.50,61) those which would be

performed if we were able to use critical thinking all the time. It is not clear what this might mean. If we were capable of such thinking then, in the prudential example above, we would presumably act differently upon different occasions - sometimes guessing over 40 and sometimes under 40 depending upon what we knew the age to be. In what way would actions guided by principle (such that we always guessed under 40) 'approximate' to those guided by critical thinking? On some occasions they would be the same, on others they would be different.

Perhaps Hare intends to claim that moral principles selected by such a method will yield actions which are more likely (when compared with those selected by other methods) to be like those which would be performed if we were able to use critical thinking all the time. As Hare says (1981 p.137), if an "intuition is one which ought to be inculcated .. [then] the most likely way of doing the right thing .. will be to follow the intuition".

But, in the prudential example above, acting according to principle is **not** the most likely way of doing the 'right thing'. If the right thing to do in each situation is what one would do if one knew all the consequences then, in the example, the right thing to do will more often be to guess **over** 40. Following the principle, in this case, ensures that the probability of doing the right thing on each occasion is less than one would achieve if one guessed at random.

Other principles which involve a concern for the preferences of **others** will have a similar result. If, for example, I am considering preference satisfaction in order to decide whether I should, in general, put away sharp tools after use then I will, say, balance the inconvenience of putting them away against the probability of injury to my children. Consideration of **overall** preference satisfaction will yield the principle

'put away sharp tools' but (if the probability of injury is less than half) the balance of preference satisfaction on each occasion will more often be in favour of **not** putting away the tools.

We have here two models for the selection of moral principles. The first model is based upon maximising overall preference satisfaction (or acceptance utility, or expected value) over a range of situations. The second model is based upon maximising the number of situations in which preference satisfaction is maximised (ie. maximising the number of occasions on which we 'do right').

Others have, in a Utilitarian context, spoken as if principles or rules are to be arrived at by means of the second model. For example, Rawls (1955 p.18) says that according to the summary conception of rules (which applies to those rules not embedded in a practice) "One is pictured as estimating on what percentage of cases likely to arise a given rule may be relied upon to express the correct decision, that is, the decision that would be arrived at if one were to correctly apply the utilitarian principle case by case.". But rules not embedded in a practice can be arrived at in different ways: we can use our knowledge of a range of situations to ground rules which may be relied upon to maximise 'correct' decisions, or to ground rules which may be relied upon to maximise overall preference satisfaction.

However, to adopt a principle based upon maximising correct decisions over a range of situations is to ignore the strength of our preference for the action which maximises preference satisfaction in each situation. It is to ignore, for example, the fact that I have a very strong preference that the tools be put away in those situations where an injury will occur if they are left out and I have only a very weak preference that they be

left out in other situations. To ignore this fact is to treat the maximisation of preference satisfaction in each situation as if it always had the same value. To do this would be to adopt a theory in which moral rightness in each situation is measured by maximisation of preference satisfaction, but in which the aim is to maximise occurrences of morally right action.

We could attempt to ensure the rejection of such a theory by requiring acceptance of the proposition that we should use the same form of thinking when selecting a principle for a given range of situations as we would use in making a judgment for a particular situation. That is, we should adopt a form of thinking in which we acquire preferences for different types of case within the range (the frequency of situations of each case corresponding to the probability of occurrence) and form a judgment on the basis of the strengths (and frequencies) of those preferences in the same way as we would do on the basis of preferences relating to one situation.

This proposition is implicit in Hare's theory. It is in fact essential to his characterisation of the form of thinking used in making moral judgments for a **particular** situation. For that involves acquiring preferences for a **range** of hypothetical situations identical to the actual situation.

However, this proposition will not be sufficient. We could use just such a form of thinking in selecting principles but use it to select principles for those ranges of situations corresponding to the different types of case on which our judgments of probability are based. Thus, in the sharp tools example, we could select a principle for the range of situations in which an injury **will not** occur and select a different principle for the range of situations in which an injury **will** occur. We may then use our judgments of probability as a guide to

the **application** of a principle in each particular situation. If we then always apply the principle for that type of case which has the highest probability within the wider range of situations we will (once again) prescribe in a way which maximises correct decisions. In the example, we will always prescribe 'leave out the sharp tools'.

The rejection of a theory which was (in this way) based upon maximising right action would require the acceptance of the further proposition that we should select principles only for those ranges of situations which are such that we can identify particular instances. We cannot identify those situations in which injury will not occur but we can identify situations in which sharp tools have been used. This second proposition concerns the role of our judgments of probability. It insists that those judgments are used in the **selection** of the principle for action rather than in its **application**.

It seems to me that this second proposition cannot be derived from Hare's analysis of our moral language. That analysis points to a requirement that we universalise our moral prescriptions and in so doing take account of what those prescriptions mean through determining consequences and strengths of preferences for those consequences. It does not place any constraints upon the range of situations over which we universalise when selecting general principles.

Both propositions **are**, perhaps, implicit in a 'traditional' Utilitarian approach. That approach has as its starting point the value of maximising preference satisfaction. An action, principle, institution etc. is good insofar as it is conducive to that end. A principle which is selected on the basis of the two propositions will be conducive to that end.

Hare's 'rationalist thesis' is, however, an attempt to avoid such a starting point and to ground a moral theory in an analysis of the logic of our moral language. Hare attempts to show that the logic of that language generates a view of right action which requires that we ought to maximise preference satisfaction on each occasion. But the theory cannot (by its very nature) provide grounds for a 'traditional' response to the fact of our imperfection. The theory cannot, it seems to me, tell us whether we as imperfect ought to try to maximise preference satisfaction overall or whether we ought rather to try to maximise preference satisfaction as often as possible.

Aiming to maximise the frequency of right action is one way of approximating to consistent right action, and such an aim seems to be entirely compatible with Hare's theory. That aim will, however, sometimes generate very different principles to those generated by aiming to maximise preference satisfaction overall. Hare is, of course, able to interpret the theory (in the light of our imperfection) in either way but that interpretation cannot appeal to the logic of our moral language. We have here a further (and, I believe, significant) gap in Hare's "highly rationalist thesis".

The benevolent archangel as ideal for imperfect agents.

A moral theory based upon the ideal of the benevolent archangel attaches value to a form of thinking involving knowledge and acquisition of preferences (through benevolent identification) and knowledge of consequences. There may be different ways in which we could attempt to live up to that ideal; that is, different ways in which we could interpret the ideal in the light of our unavoidable imperfection.

Morality and education in the light of our imperfection.

We could, for example, try to know and acquire as many preferences as possible, and to know as many consequences as possible, in each situation that presents itself. That is, we could make our best attempt to live up to the ideal in **each** situation. But, in almost all situations, our knowledge of consequences and preferences will be woefully inadequate; and, even where it appears adequate, we may always have taken wrong account of (or left out of account) some crucial factor which would reverse our judgment. The first step in interpreting the ideal, in the context of our imperfection, is to reject this route on the basis that the knowledge we are able to acquire would seldom be adequate.

We could, alternatively, try to respond to the features of certain **types** of situation in a way which resembled the benevolent archangel's response to the features of each particular situation. That is, we could attempt to use a form of thinking based upon acquisition of preferences and knowledge of consequences relating to actions within a range of situations. But the discussion in the previous section has shown that there are different ways in which we could do this.

In the previous section two propositions were given stipulating that:

in selecting moral principles for a range of situations one should

1. consider a range of situations which is such that one can know whether a particular situation is within that range (one may thus include several types of case and therefore need to make use of judgments of probability);
2. consider the strengths (and frequencies based upon judgments of probability) of preferences for situations across that range in the same way as one considers those for one situation.

Morality and education in the light of our imperfection.

These two propositions ensure an approach to the **selection** of principles which takes account both of strengths of preferences and of probabilities. They are therefore consistent with an approach which aims to yield principles conducive to the overall maximisation of preference satisfaction. This is not, of course, an aim which is directly entailed by a theory founded upon the universal ideal of the benevolent archangel but it will be the case that I shall incorporate those two propositions within the interpretation offered of that ideal.

However, before offering that more detailed interpretation, I wish to clarify the relationship between the two propositions and the two models of selecting principles - aiming to maximise overall preference satisfaction and aiming to maximise 'right' action. Such clarification will, I hope, be of help when I attempt to justify the incorporation of the two propositions.

It is possible to pursue the aim of maximising overall preference satisfaction in a manner which goes against the two propositions. Those propositions require that we use knowledge of probabilities in the **selection** of principles. Suppose, contrary to that proposition, we select different principles for different types of case within a given range of situations and then use knowledge of probabilities as a guide to the **application** of those principles. We may then adopt one of the following strategies:

a. always apply the principle for that type of case which has the highest probability within the wider range of situations (as in the previous section);

b. attempt to apply the principle for each type of case with a frequency which approximates to the proportion given by the probability within the wider range of situations.

If we adopt the second strategy and achieve a good match then we may not only increase the occurrence of right action but we may also increase overall preference satisfaction. To adopt the second strategy and simply apply at random principles for each type of case would, presumably, be irrational. But it may be that, although I do not know the type of case in each situation, I can make a 'good guess' and, in this way, attempt to correctly match a high proportion of cases.

For example, I may believe that when I cannot hear children in the house then this is a situation in which injury will not occur if I leave out the sharp tools. I may then apply the principle 'leave out the tools' when I cannot hear children and otherwise apply the principle 'put away the tools'. If, as a result, cases in which children pick up the sharp tools which have been left out are extremely rare then I will have succeeded in increasing right action (as compared with always leaving out the tools) and I may also have succeeded in increasing overall preference satisfaction (as compared with always putting away the tools).

I shall call someone who adopts such a strategy a 'moral gambler'. A moral gambler is willing to use a guide to application of principles before he knows whether it is a reliable guide. If he is a good gambler then he will in fact succeed in increasing overall preference satisfaction. Given a consequentialist approach then the principles selected by a moral gambler are good principles if action guided by those principles **does** increase overall preference satisfaction. The principles will achieve this if it **turns out to be** the case that the gambler's method of applying those principles achieves a sufficiently good match.

But the point is that the moral gambler is willing to take a chance on this being so. **The moral gambler's approach to applying principles may outstrip his knowledge.**

Any approach in which knowledge of the probabilities (of different types of case within a range of situations) is used in the application, rather than the selection, of principles makes no use of our knowledge of the strength of preferences for the actions prescribed for each type of case. An approach of the sort which I have just described not only makes no use of that knowledge of preferences but also is not based upon knowledge of the type of case which a given situation represents (it is, at best, based upon successful speculation or guess-work).

The two propositions given earlier would ensure that knowledge of strengths of preferences for, and probabilities of, different types of case within a range of situations is used in the selection of principles. They would also ensure that application of such principles, in each situation, is based upon knowledge of the type of case which that situation represents. Our approach to the application of principles could not then outstrip our knowledge. For these reasons I shall incorporate those propositions into the interpretation of the universal ideal of the benevolent archangel in the context of our imperfection.

However, the moral gambler may reappear despite the constraints which this interpretation places upon our thinking. The moral gambler is willing to try out guides to the application of principles and see how they go. He may also be willing to try out different principles. That is, he may be willing to select and adopt principles for a range of situations even though he lacks sufficient knowledge of the strengths of preferences for, and

probabilities of, the different types of case within that range. **The moral gambler's approach to selecting principles may outstrip his knowledge.** I shall, therefore, incorporate a third proposition so that we now have:

> in selecting moral principles for a range of situations one should
> 1. consider a range of situations which is such that the knowledge one has of preferences and probabilities is sufficient as a basis for judgment;
> 2. consider a range of situations which is such that one can know whether a particular situation is within that range (one may thus include several types of case and therefore need to make use of judgments of probability);
> 3. consider the strengths (and frequencies based upon judgments of probability) of preferences for situations across that range in the same way as one considers those for one situation.

The ideal thus interpreted requires that moral judgments are based only upon **knowledge** of consequences, preferences and probabilities. Someone whose judgments outstrip such knowledge lacks moral worth even if those judgments lead to action which increases overall preference satisfaction. The moral gambler who selects principles, and uses guides to application of those principles, which turn out to be 'best' in consequentialist terms, nevertheless lacks moral worth. The moral gambler is not irrational; his judgments are based upon his knowledge but they outstrip that knowledge; he relies upon speculation or guess-work and is, for that reason, immoral.

If we are not constrained in a way which conforms to the three propositions then our judgment (when selecting or

applying principles) may outstrip our knowledge - we then lack (what I shall call) **'cognitive humility'**. The ideal thus interpreted requires that our moral judgment be constrained by such cognitive humility.

Before going further, I wish to emphasise that I am not claiming that the features of this interpretation are **entailed** by the adoption of that ideal. Just as the concept of right action which is embodied in Hare's theory does not (I have claimed) entail a particular account of a 'good principle' of action, so too the concept of an ideal agent which is embodied in the theory here outlined does not entail a particular account of the 'imperfectly good' agent.

The theory here outlined has offered a concept of an ideal agent and a (so far) partial interpretation of that ideal in the context of our imperfection. But the nature of the argument I shall offer for adoption of the ideal given by the theory will not be such as to require a relationship of entailment between the ideal and its interpretation. Hare's rationalist approach **does** require a relationship of entailment since the aim is to ground the theory in an appeal to logic and the facts. He cannot ground the theory in such an appeal and then elaborate the theory in a way which requires appeal to a new element - for example, to the value of overall maximisation of preference satisfaction.

The approach which I shall take to providing 'grounds' for the theory will be of a far less ambitious nature. Both the ideal and the interpretation will be argued for in the next chapter but I shall merely try to argue that the ideal **as interpreted** entails a morality which we have reason to let into our lives and, especially, into the lives of those we educate.

The role of cognitive humility.

So far the interpretation of the ideal of the benevolent archangel describes a form of thinking which will result in the selection of principles for action; that is, prescriptions for a range of situations which may be very wide or very narrow. I shall continue to refer to all such forms of thinking as 'critical thinking'. The ideal draws us towards thinking which would result in prescriptions for ranges which are less wide (ultimately to prescriptions for particular situations) but cognitive humility restrains us in such a way as to ensure that those ranges are not so narrow that our thinking outstrips our knowledge.

However, we are still at (what Hare would call) the 'critical' level in which we deliberate in order to select moral principles. We have not yet discussed the 'intuitive' level in which we act (without such deliberation) in accordance with such principles. For Hare, if an action is thus in accordance with principle then it has moral worth. In the theory here outlined, this cannot be sufficient and is not necessary. An action has moral worth if it **arises from** a form of thinking which in some way resembles that of the benevolent archangel. An action which is in accordance with principle but does not arise from such 'critical thinking' has mere 'legality'.

We might propose that an action has moral worth if:

> the action arises from a disposition to act according to principle **and** that disposition, in turn, arises from critical thinking.

This would be sufficient to ensure that we had a theory which was **not** self-effacing. Anyone whose actions thus had moral worth would not believe that action according to principle was morally right **simply as such**. He would

believe it was morally right because the prescription is given by a principle which results from critical thinking.

However, such a person may well be tempted to question the results of such thinking when faced by a particular situation. He may attempt to prescribe for a narrower range of situations than that covered by the principle and, in so doing, to act against the principle.

From a consequentialist perspective, we would need a disposition which is strong enough to prevent this because without such a disposition overall preference satisfaction may be reduced. From that perspective, the foundation of that disposition does not matter so long as it is effective. Williams claims that in order to be effective it must be based upon 'moral repugnance' - we will be (sufficiently) strongly disposed to act according to principle only if we believe that such actions are morally right (simply as such) and if we recoil from actions which are morally wrong. It is this claim which leads him to refer to the 'deeply uneasy gap' between theory and action which may be characteristic of a two-level consequentialist theory.

From the perspective of our theory we **ought** to be restrained by cognitive humility. We ought to be restrained in this way not because a lack of such restraint would reduce overall preference satisfaction but, rather, because such humility is intrinsic to the view of moral worth outlined by the theory.

If, as a result of critical thinking, I am (for example) disposed to always tell the truth then I may be inclined to use critical thinking in order to determine whether that prescription is appropriate in a particular situation (or less general type of situation). I may then ask myself, say, whether in this situation (or type

of situation) x would not mind being deceived, y would derive some large benefit which cannot be achieved another way, and no other harm would result. If I ask this then cognitive humility ought to ensure that I also ask 'Do I know?' and, if the answer is 'No', it ought also to ensure that I retreat to the prescription for the more general type of situation.

Cognitive humility is a feature which is central to our interpretation of the ideal of the benevolent archangel. We fail to have cognitive humility insofar as we permit our moral thinking to outstrip our knowledge. To fail to have cognitive humility is to fail to live up to the ideal for it is to engage in deliberations which are not based only upon knowledge - they are also based upon speculation or guess-work. The ideal requires that we recoil from this.

Partiality to self.

In the last section I said that we might propose the following criterion for the moral worth of an action:

> the action arises from a disposition to act according to principle **and** that disposition, in turn, arises from critical thinking.

But there is an alternative disposition which I wish to discuss at some length because, I believe, it may involve features which would significantly affect the 'strength' and persistence of the disposition. My initial approach to that alternative will be based upon an attempt to take some account of such human weaknesses as partiality to self.

Such weaknesses may mean that even if we were able to engage in perfect critical thinking (that is, thinking which was identical to that of a benevolent archangel and

was thus aimed at prescribing for one particular situation), we might nevertheless not act in a way which conformed to the results of such thinking. It may be that at the moment of such deliberation we could not but be disposed to act in such a way, but the moment of deliberation may not be the moment of action. The opportunity for action may not be immediate and/or the performance of the action may take time. In that time our weakness may assert itself.

Such a weakness may consist, in part, of our inability to reaffirm such a deliberation through time. We may be disposed to exercise our ability to deliberate in this way but we may also be disposed to deliberate in a way which, say, takes account only of our own preferences. We may be unable to sustain the former deliberation, and the resulting prescription, through the time required for completion of the action.

One response to such a weakness might be for us to attempt to sustain **some** of the features which gave rise to the prescription. In arriving at such a prescription we will have acquired (through benevolent identification) those significant positive preferences the satisfaction of which represents the advantages of acting according to the prescription and those significant negative preferences the satisfaction of which represents the disadvantages of acting otherwise. On the assumption that the prescription is correct, such preferences (or some subsets of them) will jointly outweigh those preferences which relate to the disadvantages of acting according to the prescription and the advantages of acting otherwise. I shall call such preferences the 'decisive' preferences in relation to the prescription for a particular situation.

For example, suppose that I am capable of perfect critical thinking and that my attempt to decide whether

we (myself and spouse) ought to invite an aged parent to share our home is based upon knowledge that: our children love being with him; he generally prefers his own company, is indifferent to our company but enjoys being with our children; we find his presence disruptive and his company irksome; we, nevertheless, worry about his living alone; he would dislike the rules and restrictions involved in alternative accomodation; and so on. Suppose the strengths of significant preferences are as follows:

	our home		parents own (distant)		alternative (closer)	
parent	company	1	independence	2	rules etc	-4
children	company	5	occasional visits	1	frequent visits	3
myself+spouse	disruption	-5	worry	-4	cost	-2
total		1		-1		-3

The result of a deliberation which involved acquiring all such preferences through benevolent identification would be the prescription 'invite the parent'. The decisive preferences are those relating to the (smallest set of) advantages of the prescription and the (smallest set of) disadvantages of each alternative such that, when we compare the prescription with each alternative in turn, these jointly outweigh all the disadvantages of the prescription and all the advantages of that alternative. Thus, the decisive preferences are those relating to company for our children if he were to come to our home, our worry if he were to stay in his own home, and his feelings if he were to go to alternative accommodation.

The suggested response to our weakness involves our attempting to maintain our benevolent identification with the decisive preferences, and to discount all other preferences, through the time required for completion of the action prescribed in a particular situation. Such a

response may enhance the ability to maintain a disposition to act according to the prescription which results from deliberating in the manner of our ideal agent. Thus, in the example, if we find that our preference for avoiding disruption is beginning to assert itself, and threatening to influence our action, then we should put it out of our mind and remind ourselves of just how our children would feel if he were to come to our home, and so on - we should focus on those preferences which our critical thinking showed to be decisive.

If we were disposed in this way to focus upon decisive preferences then we would be disposed to reaffirm our critical thinking without the necessity for a detailed repetition of that thinking. Such a disposition would underpin a disposition to act according to the results of critical thinking in a way which, I believe, would enhance our ability to maintain that disposition over time. More importantly (perhaps), the disposition to focus upon decisive preferences also provides a means of giving an alternative account of moral worth at the 'intuitive' level.

Decisive preferences and general moral principles.

Most of us are seldom capable of perfect critical thinking aimed at prescribing for one particular situation. Many people (especially the very young) are not capable of critical thinking even in the context of very general principles. Ought such people to behave in a way which merely conforms to principle?

According to the theory here outlined, actions which merely conform to principle have no moral worth (they have mere 'legality'). To tell the truth because one has been 'educated' to do so, or wishes to please the moral

educator, or fears punishment, or expects reward, or finds telling the truth pleasurable, or believes truth telling is right simply as such, is here (as for Kant) to lack moral worth. If our actions are to have moral worth they must stem from a form of thinking which in some way resembles that of the benevolent archangel - they must in some way stem from knowledge of the consequences of our actions and benevolent identification with the preferences of others.

Each person may have very different preferences. Each may have preferences which others lack and preferences which are stronger or weaker than the same preferences possessed by others. As Mackie, and others, might say: "different people have irresolvably different views of the good life" (Mackie 1977 p.169). But there are preferences which all (or most) share. Such preferences may be for those things which are prerequisites to the pursuit of our 'good life': to avoid injury, to keep what is ours, to not have false beliefs, to go where we choose. Other such preferences may be more fundamental: to be free from physical pain, to eat when hungry.

This is not to claim that all (or any) such preferences are a **necessary** part of human 'nature'. Nor is it to claim - as Foot claims (1958 p.11-12) - that they are wants which refer to benefits which "a man has reason to want if he wants anything" or to harms which are "necessarily something bad and therefore something which as such anyone always has reason to avoid". Such preferences may arise out of the social conditions and culture in which we find ourselves. But in each society or culture there will be a set of preferences which are in this way strong and pervasive.

Not only are such 'central' preferences shared by (nearly) all but also the opportunities for us to act in ways which have consequences for their satisfaction are

extremely extensive. In almost every situation in which we find ourselves we have the opportunity to lie to, or steal from, or hit, or kill, or coerce others and thus to affect the satisfaction of the central preferences of others. Furthermore, situations in which such actions as these would bring significant benefit, or avoid significant harm, to ourselves or others are, comparatively, extremely rare.

Thus, across the widest possible range of situations, central preferences will be **decisive** in relation to prescriptions such as 'do not hit'. We do not need knowledge of the probabilities of different types of case within that range to determine that this is so; knowledge of the extensiveness of our opportunities and the strength and pervasiveness of the preference is sufficient.

If such preferences are in this way decisive then we ought to focus upon them in the way that was outlined in the previous section. That is, we ought to strive to maintain our benevolent identification with those preferences, and to discount all other preferences, in those situations where we may be tempted to act against prescriptions such as 'do not hit'. If we were disposed in this way to focus upon decisive preferences then we would be disposed to reaffirm our critical thinking without the necessity for a detailed repetition of that thinking. Such a disposition would underpin a disposition to act according to the results of critical thinking in all such situations.

Even if we are not capable of critical thinking we may well be capable of focussing on such preferences. We may know that hitting will cause physical pain to the other person and we may be able (since we also have that preference) to benevolently identify with the preference of that person not to suffer physical pain. If our

actions stem from such a focus then they stem from knowledge of decisive consequences and benevolent identification with the decisive preferences of others.

Thus I propose the following criterion for the moral worth of an action:

> the action arises from benevolent identification with the decisive preferences of those involved.

If I do not hit you because I want you not to suffer physical pain (just as you want not to suffer) then my action has moral worth. If I do not lie to you because I want you not to have false beliefs (just as you want not to have false beliefs) then my action has moral worth.

Dispositions to act in such ways are dispositions to be motivated by the features of a situation which are, in fact, morally decisive in relation to those actions. Such dispositions may arise from critical thinking and, as we saw in the previous section, that would be a rational response to our weakness. Thus the disposition to focus in this way upon certain preferences will be a characteristic of the imperfect moral agent at every level of critical thinking (from the most general to the most particular). What I am proposing here is that it also be a characteristic of the imperfect moral agent at the intuitive level - the level at which no critical thinking has taken place.

In the context of our imperfection the ideal that

> **we ought all to be (more like) a benevolent archangel**

will thus require that we be disposed to:

> **focus upon decisive preferences;**
> **engage in critical thinking;**
> **be restrained by cognitive humility.**

For those capable of critical thinking, the disposition to focus upon decisive preferences can be and ought to be

informed and underpinned by, or arise out of, critical thinking constrained by cognitive humility. For those not (yet) capable of critical thinking the aim of moral education must be to develop an ability and inclination to benevolently identify with the (decisive) preferences of others. The aim of moral education cannot simply be the development of a disposition to act in accordance with the critical thinking of others. Actions ought not to be merely 'right' they ought to have moral worth. If the actions of those educated have no moral worth then moral education has not begun.

CHAPTER 11.

An educational approach to moral theory.

Summary.

Hare's moral theory has the following features:

the informed preferences of each individual are to be given a weight in our moral deliberation which corresponds to their strength;

moral judgment involves a comparison of alternative actions which is based upon knowledge of preferences and consequences;

the morally right action in each situation is that which would maximise the preference satisfaction of all concerned;

moral judgment aimed at selecting general principles for action will yield principles which would ensure overall maximisation of preference satisfaction;

in determining general principles moral judgment will yield principles for action which conform to moral 'intuition';

a sound moral education will result in a reluctance to act according to moral judgment which does not thus conform to moral intuition.

That theory is founded upon an appeal to the 'logic' of our moral language.

An educational approach to moral theory.

A theory which is derived from the universal ideal of the benevolent archangel shares all the features listed above but two of those features acquire a new significance.

Firstly, a respect for the informed preferences of others is extended to include the second-order preferences (the personal ideals) of others. Such preferences need not be construed in terms of degree of attainment of personal first-order preference satisfaction. The moral theory outlined takes account of the fact that we can and do strive to be other than we are but does not insist that this should be seen in terms of the struggle to satisfy the preferences we now have - it does not insist upon a reductionist utilitarian perspective in the context of personal ideals. Benevolence demands a concern for the preference satisfaction of others but it does not demand that our concern for ourselves should only be a concern for the satisfaction of our preferences. The utilitarian theory offered here is not based upon the value of preference satisfaction in the way that the utilitarianism of (for example) Mill is based upon the value of happiness.

Secondly, if we are to have moral worth then a reluctance to act against general moral principles cannot be based upon a habit successfully instilled, or upon a 'knowledge' of right and wrong, or upon a desire to please our educators. It must be based upon cognitive humility and a disposition to focus upon decisive preferences. A moral education which develops the moral worth of the agent will ensure that whenever we are inclined, in a given situation, to question the application of such principles that inclination will be inhibited by the focus upon decisive preferences and that questioning will be constrained by cognitive humility.

The ideal (as interpreted in the previous chapter) entails several further features which stem from the

direct commitment to benevolence and the direct rejection of malevolence.

I claimed (in chapter 7) that Hare's **indirect** approach to the rejection of malevolence rests upon an over-optimistic view of the prevalence of sadistic inclinations and our ability to redirect them. It seems to me that, just as a direct commitment to benevolence is a feature of our 'intuitive' moral views, so also is the direct rejection of malevolence. Such a view may be justified, in part, by a recognition that Hare's view **is** over-optimistic; we recognise that, just as we can all be moved by the joy of others, so too we can **all** desire the suffering and frustration of others. Human history not only inspires through its examples of pure benevolence but also repels through its endless succession of examples of pure malevolence between individuals, groups, societies and races.

But it is not only the case that Hare's view may be over-optimistic it is also, and more importantly, the case that it fails to respond to the fundamental difference between bringing about an outcome in order to satisfy my preference and bringing about an outcome in order to satisfy or frustrate a preference which I know to be yours. It is a feature of human nature that we do not benefit or harm others only in order to bring about an outcome which we already prefer - we have pure benevolence and pure malevolence.

It is this difference which Schopenhauer made use of in claiming that there are three basic human motives: self-interest, benevolence and malevolence. To a greater or lesser degree, we are all motivated in these ways. The direct rejection of malevolence stems from a recognition that malevolent preferences are **not** simply self-interested preferences the satisfaction of which unfortunately reduces the preference satisfaction of

others. As Schopenhauer (1851 p.136-139) says, "everyone bears within him something altogether morally bad" and the worst and most distinctive trait of wickedness that we possess is that which leads us to take pleasure in the frustration and suffering of others, to "torment another simply for the sake of tormenting". Malevolence and benevolence correspond to aspects of our nature which are fundamentally different to that of self-interest.

The theory here outlined gives rise to the claim that the moral worth of an agent is related to the way in which he is motivated by the preferences of others - it is centrally about becoming benevolent and ceasing to be malevolent. Moral worth involves (at most) judging as if we were members of a community of benevolent archangels and being, thereby, motivated to act; and (at least) judging on the basis of benevolence towards those preferences which a more complete judgment would reveal to be decisive. Thus, according to the theory, the moral worth of an individual cannot be seen in terms of how conducive that individual is to some end - whether that be the maximisation of overall preference satisfaction or the absence of lies.

If this were all then it would still be possible to claim that, although the moral worth **of an agent** is determined in this way, nevertheless morality is centrally about the 'legality' (the moral worth) of our actions. Those having moral worth would (as a result of their benevolence) wish to ensure that others act rightly (either in a particular situation or in general) and they might consider that the best way of achieving that is to ensure that the behaviour of others conforms to general moral principles (by whatever means is most effective).

But the theory here outlined is based upon the universal ideal of the benevolent archangel. The theory consists only of that ideal and its interpretation in the context

of our imperfection. That theory assigns no moral worth whatsoever to actions in respect of their legality and requires that **all** ought to be more like the benevolent archangel. To ensure only that the actions of others has legality is to entirely ignore **their** (actual or potential) moral worth. My desire that you should do right has moral worth if it stems from my benevolence; but your doing right has moral worth only if it stems from your benevolence.

The theory permits us to say, with Kant, that nothing "can be called good without qualification except a good will" and that "a good will is good not because of what it performs or effects, not by its aptness for the attainment of some proposed end, but simply in virtue of the volition". But the theory also permits us to say (again, perhaps, with Kant) that not only ought we not to treat others merely as a means to our own satisfaction but also that we ought not to treat others merely as a means to right (ie. to the satisfaction of all). As moral agents each of us is, in this sense, an end in himself.

Moral education, therefore, cannot simply aim to instil a sense of right and wrong in terms of a disposition to obey such principles as 'do not lie'. It must be concerned with benevolence and non-malevolence as the motivation for our action and as the inspiration for our lives. A capacity and disposition to determine the consequences of our actions and to benevolently identify with the preferences of others is the beginning and end of moral education.

There cannot be, therefore, a 'deeply uneasy gap' between the spirit of the theory and the spirit it justifies. Truth-telling, loyalty, and so on do, in a sense, have instrumental and not intrinsic value. But they are instrumental in satisfying the preferences of others (not

to be deceived for example) and our desire to satisfy such preferences stems from critical thinking restrained by cognitive humility. At all levels it is our knowledge of the consequences of our actions and our benevolence towards the preferences of others which ought to motivate us to act.

The theory here outlined is, however, not founded upon an appeal to logic. It is founded upon the universal ideal of the benevolent archangel and the appeal to that ideal has not yet been argued.

Human nature, moral intuitions and decision procedures.

I have argued that the theory here presented can accommodate certain fundamental features of human nature.

The first of those is the capacity which we have for self-evaluation. We are capable not only of evaluating the outcomes of our actions but also of evaluating the quality of our own motivation. To leave this dimension of human agency out of account is to give a hopelessly shallow account of what it is to be human. Such a dimension gives rise to personal and universal ideals of self. The theory incorporates a respect for personal ideals of self and is founded upon an appeal to a universal ideal.

The second is our propensity for benevolence and malevolence. We do not merely pursue our own interest; we are all inclined, to varying degrees, to engage in acts of gratuitous malevolence and benevolence. This, one might also claim, is a peculiar characteristic of human agents. This feature is linked with the first in that both require imaginative identification - the first with the preferences of an ideal self, the second with the preferences of others. Both are linked with prudence

insofar as that involves the capacity to imaginatively identify with the preferences of one's future self. The capacity to identify with, and to be motivated (in various ways) by, preferences other than those we now have has a central role in the theory.

These features also provide a means by which we can make intelligible the experience which Kant claims is central to morality. That is: it is possible to choose to act in a way which would lead to something we wanted to avoid more than we wanted anything else. Our recognition of this possibility may be seen as a recognition of the fact that we all can deliberate as if we had preferences other than those we now have. Furthermore, that we can do this is not just a feature of our human nature it is, according to the theory, the feature without which we could not have moral worth.

It is also the case that the theory (as interpreted) responds to our imperfection in a way which links us (all too imperfect) to the ideal of perfection by means of a common motivation. Benevolence, non-malevolence, and knowledge of consequences is, whatever the level of ability, the source of actions which have moral worth.

Finally, the theory incorporates or underpins many of our moral intuitions. Firstly, it underpins a view of others which requires that we treat them as ends and not as means. Secondly, it involves a direct commitment to benevolence and a direct rejection of malevolence. Thirdly, it incorporates a form of thinking which would give rise to moral principles largely in accordance with moral intuition.

The last feature is one which the theory has in common with other forms of utilitarianism but the first three stem from the approach taken here. In requiring a moral education which does not result in an 'uneasy gap'

between theory and action, and in directly rejecting malevolence, the theory (I would claim) conforms even more closely with our moral intuitions than does a consequentialist utilitarianism.

The non-consequentialist nature of the theory gives rise to the requirement that we engage in a form of thinking, rather than that we act according to the results of such thinking. But the form of thinking which is required by the theory is essentially one that corresponds to a preference-based approach to consequentialist utilitarianism. Such thinking provides a means not only of resolving those moral conflicts which result from moral intuition but also, as Sidgwick points out (1874 Book 4, Chapter 2), of systematising the exceptions to, and clarifying the vagueness in, our intuitive moral principles.

Utilitarian theories offer a method of judgment which can be applied to selecting very general principles and to prescribing for one particular situation. Use of the method in these ways can, itself, lead to apparent conflict - between the prescription given directly for the particular situation and the prescription yielded by the principle. In Hare's theory that conflict is removed by distinguishing between a good man and a right action. The good man is disposed to act according to those principles which are most likely to result in an overall maximisation of preference satisfaction (the generality of those principles will depend upon the sophistication and moral self-discipline of the individual concerned - see Hare 1976 p.34). In the theory presented here the good man uses the method to consider a range of situations which is such that his knowledge of preferences and consequences is sufficient for judgment; his focus upon decisive preferences is the result of critical thinking restrained by cognitive humility. In both theories, the right action is that which would be

prescribed as a result of applying the method for the particular situation.

The principles which a good man would obey will be the same in either theory - each would be a principle which would maximise preference satisfaction over the range of situations to which the principle applies and which the agent would be disposed to obey. Whichever approach is taken, the same method of judgment is used to select principles, make prescriptions, resolve conflicts, clarify and systematise our intuitions. All of these are subject to one and the same criterion of decision.

Thus (I am claiming) the theory offered here can accommodate certain fundamental features of human nature, can incorporate or underpin many of our moral intuitions, and can offer a uniform criterion of decision. But, even if all of these were granted, would that provide any reason for our adopting such a theory?

The benevolent archangel as an educational ideal.

I have argued against Hare's highly rationalist thesis and, more briefly, against Kant's equally rationalist thesis. I shall now assume that we cannot ground a moral theory in reason; that is, a rational agent may not only be immoral or amoral but may also subscribe to one of a variety of different sets of moral views. But is it at least a requirement of reason that if we have moral views then we should have a **theory** of our morality, we should attempt to systematise those moral views in some way? Williams argues (1985 Chapter 6) that there is no such requirement.

If we start from our moral intuitions then we may attempt to construct a moral theory by, say, representing those intuitions in a set of stateable principles, making

explicit the relationships between those principles, deriving them from some small set of fundamental ideas, and resolving conflicts between them by means of some kind of decision procedure. In constructing such a theory we may modify the theory and, to some extent, the intuitions until they roughly fit one another.

Williams calls this approach 'rationalistic' (1985 p.100) and points out that some may feel that such a rationalistic approach "simply follows from being rational". But, he claims, rational reflection need not draw us towards theory and systematisation and "it is quite wrong to think that the only alternative to ethical theory is to refuse reflection and to remain in unreflective prejudice" (1985 p.112).

With this I can agree. The rational agent **need not** attempt to systematise his moral intuitions and he certainly need not attempt to provide some fundamental idea or a single decision procedure. As Sen and Williams say (1982 introduction):

> "a large question is being begged .. if one assumes that the agent is required in rationality to subject all .. decisions to one criterion of decision, and it is still being begged if one assumes that rationality requires that any other criteria of decision must themselves be justified by one over-riding principle".

However, if we grant that rationality does not **require** a moral theory of that sort that does **not** mean, of course, that good reasons cannot be given for attempting to provide one. It may be the case that the reasons we give for providing a moral theory, or for selecting a particular moral theory, cannot appeal to some notion of rationality; but we may still make a rational choice between having this theory or that theory or no theory. It may still be the case that there are good reasons for

representing our moral intuitions in a set of stateable principles, making explicit the relationships between those principles, deriving them from some small set of fundamental ideas, and resolving conflicts between them by means of some kind of decision procedure.

It is at this point, I believe, that an educational perspective is most important. From that perspective the primary aim is to promote a particular set of moral views. It may be the case that some sets of moral views are more easily promoted than others. It will clearly be the case that moral views which closely fit the moral intuitions of the educatees and educators (parents, teachers and others) will be more easily promoted than those which do not. It may further be the case that moral views which not only underpin and illuminate our moral intuitions but also are systematised and founded upon some fundamental idea will be more easily promoted than those which do not. If this is the case then it provides a very good reason for striving to formulate a moral theory of that sort.

The discovery that moral intuitions can not only be clarified by critical reflection but can also be derived from and illuminated by a fundamental idea may inspire and profoundly influence development. It may do so in a way which an unsystematised set of moral views will not. After reflecting upon our moral intuitions it may be perfectly rational to offer a moral view which, say, incorporates a number of values or principles which are, to various degrees and in various ways, incommensurable with one another; but such a view may be less easy to promote. As philosophers we may be reluctant to accept any moral view which, in any way, goes beyond intuition; but as educationalists this may not be the most important issue.

An educational perspective can provide further reasons for adopting such a theory. These relate to the nature of the fundamental idea and the way in which that idea indicates a certain view of moral development. If the view of moral development is one which can be elaborated into a clear and practical educational programme then we will be all the more inclined to adopt it.

According to the theory here outlined, moral education would begin with a focus upon those central preferences which are decisive for the widest possible range of situations. We will encourage a disposition to act in a certain way (not to hit others) which is such that **it results from** knowledge of consequences (it will hurt him) and benevolent identification with the decisive preference (he does not want to be hurt). We will also foster the ability to determine the consequences of actions and to determine and identify with the preferences of others - in any situation or type of situation. Such an ability can then be used in considering situations in which preferences have to be compared and balanced. As the ability to thus think critically develops we will hope that the focus upon decisive preferences will be informed by, or arise out of, such thinking constrained by cognitive humility.

This is a very brief and inadequate description but it is sufficient to make clear the way in which the initial disposition to identify with the preferences of others and consider consequences will provide a thread which runs through moral development. From an educational perspective we will want to ask, for example, whether such a disposition can appear at an appropriate stage in overall development.

Such a disposition may manifest itself in a reaction to the distress of others and action in order to prevent or alleviate that distress. Research would seem to indicate

that such behaviour occurs at a very early stage of development. Bottery (1990 Chapter 7) points to such research in order to challenge a view which he attributes to Piaget. According to that view, moral development presupposes cognitive development and children are, for a considerable part of their early life, cognitively egocentric - they are simply incapable of taking another person's point of view. Bottery calls this the cognitive-developmental model and he contrasts this with a model in which empathy and cognition develop together. For example, in one experiment children of 18 to 24 months responded to a show of distress by, say, offering a doll. The child may learn that other responses are more appropriate but, it is claimed, the experiment shows that very young children can see how others feel and so be motivated to do something.

In The Emergence of Morality in Young Children (Kagan 1987) several writers offer evidence of responses by very young children which appear to be motivated by knowledge of the preferences and feelings of others. Jarrett (1991 p.38) maintains that "the prerequisites for moral behaviour, such as empathy, sensitivity to others' distress, and being able to understand what kind of help is needed, have their origins very early along"

It may thus be the case that very young children will act to prevent or alleviate distress just because they know that is what the other person is feeling. It might also be the case that such motivation can occur at a stage which is **at least as early as** alternative motivations - for example, alleviating distress because the child knows that others will approve, or a reward will follow, or it is right to do so. If that were so then the view of moral development outlined above might begin to correspond with views resulting from research into overall development.

An educational perspective may thus provide a means of choosing between particular moral theories (or of choosing no theory at all). I shall not attempt to argue in any detail that the theory here outlined is (on these grounds) worthy of choice for that can only be determined empirically. I **would** argue that the theory underpins moral views which closely match intuition **and** provides an ideal which may both inspire and humble - the ideal of the benevolent archangel may both inspire us to be other than we are and provide us with the humility which stems from knowing what we are not.

Morality and the limits of philosophy.

My approach involves an appeal to empirical claims in justifying adoption of a moral theory. Those empirical claims relate to education. If we (teachers, parents and others) have a set of moral views such that we believe that all ought, say, "to keep their word, to tell the truth, to refrain from physical and mental maleficence, to be helpful to people in distress or need, and to be tolerant and fair" (to quote J.White 1990 p.36) then we will wish to educate so that all are disposed to keep their word etc.. If systematisation of our moral views helps in that aim then, I am claiming, that provides a good reason to strive for such systemisation. If systemisation in terms of a fundamental ideal helps in that aim then that provides a reason to strive for systemisation involving such an ideal. If the systemisation of our moral views yields a view of moral development which corresponds with aspects of overall development then that provides further justification.

From a realist perspective such an approach would appear bizarre. We do not choose to teach, say, Einstein's theory of relativity because we believe that more children would understand physical aspects of phenomena

in terms of that theory than would understand physical aspects of phenomena in terms of an alternative theory (or no theory at all). We teach such a theory because we believe that it gives the best account of such phenomena, because we believe that (in some sense) it is correct. If it were not correct then children who learned such a theory would **misunderstand** such phenomena.

However, I have assumed that moral realism is false. A moral theory relates to motivations and dispositions which can be systematised in various ways. The universal ideal of the benevolent archangel is not offered as a 'correct' way of understanding such motivations and dispositions (nor of understanding the moral 'facts' which a realist would see as underlying them) but, rather, it is offered as a way (perhaps the best way) of inspiring, strengthening and underpinning **those** motivations and dispositions.

The choice between theories (or of no theory) is thus made on educational grounds. However, it might be claimed that the push to theory is merely a push into disagreement and away from consensus. We may agree upon examples of moral motives, dispositions and behaviours but, as J.White claims (1990 Chapter 3), once we begin to ask what makes them 'moral' - once we attempt to systematise and theorise - then all kinds of divisions appear.

This may well be true; but it may also be true to say that such disagreement arises because of the nature of the criteria which moral theorists have applied in selecting a favoured theory. If we see a moral theory as essentially incorporating a rational justification for morality (or as revealing the impossibility of such a justification) then we will select that theory which incorporates the 'best' such justification. Given such an approach, a major aspect (**the** major aspect) of any

moral theory will consist of an attempt to provide such a justification. The direction in which that attempt takes the theorist will determine the nature of the systematisation of our moral views which is then offered.

Even where there is broad agreement upon what would constitute such a rational justification, divisions will rapidly appear. For example, as MacIntyre claims (1981 Chapter 5), those involved in the 'Enlightenment project' all agreed that the key premisses in such a justification "would characterise some feature or features of human nature; and the rules of morality would then be explained and justified as being those rules which a being possessing just such a human nature could be expected to accept". But whereas Hume looked to "characteristics of the passions", Kant looked to "the universal and categorical character of certain rules of reason".

MacIntyre claims that all such routes to justification are doomed because there is an 'ineradicable discrepancy' between the shared conception of moral rules and the conception of human nature - this is ineradicable since the conception of moral rules derives historically from a fundamental contrast between man as he is and man as he could be if he realised his 'true end'. According to MacIntyre, we must, therefore, seek a basis for morality in some notion of the good life for man (or the ends of human life).

I would agree with MacIntyre that the enlightenment project fails. We cannot appeal to features of human nature (man is such) in order to justify morality (man ought to be such). However, I would not agree with the claim that "the whole point of ethics - both as a theoretical and a practical discipline - is to enable man to pass from his present state to his true end" (MacIntyre 1981 p.52) for there is (I believe) no 'true' end for man. This is not to say that we cannot form

notions of a good life, or of a good man, and go on to show how these relate to our moral views. But, I am claiming, we should not be asking which such notion is **true** but rather which such notion **would best inspire, strengthen and underpin those motivations and dispositions which we see as constituting morality.**

J.White (1990 p.46) recommends that, when considering 'moral' education, we should turn away from moral theory and concentrate upon various "types of altruistic behaviour, reactions or attitudes towards others"; this because it is a fact that we cannot come to any agreement over which moral theory to select. If, however, we adopted the above criterion for selecting a moral theory then, perhaps, we might reach agreement more easily.

CHAPTER 12.

A community of (imperfect) benevolent archangels.

Aims for education and aims for moral education.
Moral relativism and moral education.
Educating for benevolence, non-malevolence and humility.
Love, humility and assessment.
Preferability of a community of benevolent archangels.

Aims for education and aims for moral education.

In both The Aims of Education Restated and Education and the Good Life, J.White offers a systematisation of, and a rationale for, our wider educational aims in terms of a notion of the well-being of an autonomous pupil. This contrasts with his approach to aims for moral education where he recommends that we move away from an attempt to offer a systematisation and rationale based upon moral theory and, instead, look towards the consensus which exists with regard to a range of 'altruistic' dispositions. His approach to education as a whole is, in this way, the reverse of his approach to moral education.

In the Introduction, I claimed that the attempt to systematise our wider educational aims in terms of an underlying rationale quickly leads to disagreement - we reach differing bed-rock commitments at an early stage in the justification process. I suggested that we turn away from such educational theory and adopt an approach based upon practical consensus. Thus my approach to education as a whole is also the reverse of my approach to moral education - but in both cases I am recommending that we

move in the opposite direction to that which J.White recommends.

We could adopt an approach to aims for education as a whole which was similar to that which I have recommended for moral theory; that is, we could take our shared views about the aims of education and attempt to systematise them in terms of some fundamental idea. As in the case of moral theory, such an approach need not involve an attempt to provide a 'rationale' for that systematisation of aims - that is, it need not attempt to provide a more ultimate justification for those aims. But, firstly, do we have shared views about the aims of education; and, secondly, what would be the benefits of such a systematisation as compared with piecemeal consensus?

There **is** substantial agreement over moral views. The disagreement comes when we try to systematise and theorise. The agreement over aims for education is, I believe, much less substantial. This is not just a matter of disagreement over priorities, interpretation and attempts to systematise; the disagreement comes earlier. This can be seen most clearly if we consider dispositions. We might agree that if someone has no disposition to help people in distress then their moral education has failed. But can we agree that if someone has no disposition to pursue (a possible) excellence in some field then their education has failed? The same question can be repeated in respect of any of the aims which are "currently at large in the world of education". Ought educated people to be **disposed to use** their knowledge and understanding, to **exercise** their autonomy, to **take pleasure in** art and culture?

If, on the other hand, we consider capacities (knowledge, skills and understanding) there will be substantial agreement over aims. But what would be the benefits of a systematisation of those aims in terms of some

fundamental idea? We cannot inspire or strengthen a capacity in the way in which we can inspire a disposition to use that capacity (or to use it in a particular way, or to use it in the service of some particular end).

Agreement over dispositions is less easy because we, perhaps, shrink from imposing a way of life or from imposing an ideal of self. We can agree over capacities because such capacities provide the means to pursue one's own view of the good life and one's own ideal self. Our reluctance to insist that an educated person should, say, take pleasure in art and culture stems from a respect for the individual's view of that life and self. The universal ideal of the benevolent archangel, as interpreted here, enshrines such a respect. However, it does (in common with all other moral views) impose an ideal, viz: **all** ought to be (more like) a benevolent archangel.

If we cannot (as I have claimed) provide any 'ultimate rational justification' for that ideal then do we have any right to impose it by means of moral education?

Moral relativism and moral education.

Before considering the issue of 'imposing' moral views upon others, I would like to consider the related issue of whether our own moral views would or should survive the realisation that they have no 'ultimate justification'. As Blackburn (1985 p.9-11) says, those who believe that our moral views would not and, perhaps, should not survive such a realisation have much in common with "those thinkers who felt that if there were no God or after-life then it would be rational to ignore the claims of morality whenever self-interest suggested it".

A community of (imperfect) benevolent archangels.

Morality may indeed lose some of its hold upon those people who cease to believe in God; and, so too, morality may lose some of its hold upon those who cease to believe that moral commitments have "real, objective truth values certified by an independent reality". But, as Blackburn points out, it will do so only if such people believe that "things do not matter unless they matter to God, or throughout infinity, or to a world conceived apart from any particular set of concerns and desires, or whatever". It will do so only if they have, what Blackburn calls, a 'defective sensibility'.

Morality matters to us because we **do** approve of certain types of action or motivation (and recoil from others) or because we **do** approve of the consequences of those actions or motivations. However, this does not bring us any closer to a 'justification' of morality; and those with objectivist or rationalist leanings will find it unsatisfactory because the 'we' may not be all of us. Those with such leanings will hanker after a means of ensuring that the 'we' is at least all those who are able to rationally decide what is right or wrong - whether that be by means of establishing the moral facts, or by means of a particular type of prescriptive moral thinking which is rationally required. But, I have argued, it is likely that such people will be disappointed.

Now Mackie (1977) claims not that we all have such hankerings but, rather, that we all do feel that the demands of morality are, in some sense, independent of us and our motivations - whereas in truth they are not and we are, therefore, in error. Furthermore, as Williams (1985a) suggests, this feeling "can be plausibly explained by supposing that ethical constraints and objectives have to be internalised in such a way that they can serve to control and redirect potentially destructive and uncooperative desires, and that they can do this, or do it most effectively, only if they do not

present themselves as one motivation or desire among others .. [if, that is, they] present themselves as something given". If this were true (and Mackie's error theory were true) then, as Williams says, it would be difficult to claim that moral conviction need not be upset by a realisation that the demands of morality lack any such independence.

I have claimed that our moral views may rest upon, and be interpreted in terms of, a universal ideal. If our moral views are related in this way to a universal ideal then our moral conviction (our inclination to respond to those views) will be relative to our commitment to that ideal. From this perspective our moral views lack the sort of independence which Mackie describes. If we feel that our moral views do have, or should have, such independence then it may be that our commitment to those views would be weakened by an acceptance of a lack of that independence. But if we believe, as I do, that our moral views cannot have such an independence then our commitment to those views need not be weakened at all. Mackie's claim that we (all) believe that our moral views have such independence is false; and the claim that our moral conviction may be weakened by an acceptance of moral relativism is only true of (some of) those who feel that our moral views do have, or should have, such independence.

It may be that some perspectives upon our moral views incorporate, what Blackburn calls, a defective sensibility. Those who have such a perspective may not only find it difficult to accept moral relativism (and to cease to hanker for objectivism or rationalism); but may also find that an acceptance of moral relativism brings with it a weakening of moral commitment. But, as Williams claims (1985a p.213), some moral perspectives are better adapted to being seen for what they are. I would claim that a perspective upon our moral views which

rests upon an ideal of benevolence, non-malevolence, understanding and humility may ensure that our moral commitment survives the realisation that moral relativism is our only option.

Having adopted a relativist position we may now consider the issue of whether that relativism should bring with it a reluctance to impose those moral views upon others by means of education. But first we perhaps need to clarify our use of the expression 'relativist'.

According to Krausz and Meiland (1982 Introduction) the view that there is no criterion which would reveal a particular set of moral beliefs to be **the** true or correct set of moral beliefs may lead us in one of two directions. The first involves concluding that there is no truth (or that the truth cannot be known) and this they refer to as moral scepticism. The second involves concluding that there are many truths (the truth may be different for, say, different societies) and this they refer to as moral relativism. In making this distinction they draw a parallel with empirical beliefs.

However, throughout this thesis I have been concerned, like Hare, to describe a form of thinking which will yield **prescriptive** moral judgments. So a distinction which runs parallel to that made for **descriptive** empirical judgments may not be helpful. My aim, and Hare's, is "to find a system of moral reasoning which we can use when faced with moral questions" (Hare 1981 p.214). Hare claims that there is only one such system which is rational. I would claim that there are many: Hare's critical thinking would be one, deriving prescriptions from a limited set of general principles would be another, and deliberating as a member of a community of (imperfect) benevolent archangels would be a third.

My position, and Hare's, is thoroughly 'sceptical' with regard to the claim that moral beliefs can be true of moral 'facts' but it is not sceptical with regard to the claim that there are rational methods of determining prescriptive moral judgments. Hare's position is rationalist: there is only one such method and that method is employed by all rational agents. My position is relativist: there are many such methods and each such method is employed by all rational agents who share a fundamental commitment. That commitment may be to a form of thinking, or to a set of principles, or to an ideal.

I would claim that a commitment to the ideal of the benevolent archangel can yield a commitment to a set of moral principles and moral views which most of us share. But neither the commitment to the ideal nor the commitment to the principles and views has any ultimate justification. Some might then argue: if there are alternative commitments or moral systems and none of them has any ultimate justification, then each is as good as the other; and, if that is so, then we have no right to impose **our** moral system upon those whom we educate.

There are several key expressions here: 'right', 'impose' and 'as good as'. Firstly, I would agree with Chamberlin (1989 p.32) who claims that nothing extra is gained "when we say 'I have a right to do this' or 'You have no right to do that' beyond what is expressed by 'You ought not to stop me doing this' and 'You ought not to do that'". Claims about rights are basically statements about how people ought to be treated and how others ought to treat them. Secondly, to say 'impose' is, here, simply to convey a rather unacceptable manner of influencing and educating. So I will recast the conclusion of the above argument:

> (given one moral system is as good as another) we ought not to educate others according to **our** moral system.

A community of (imperfect) benevolent archangels.

In a different context this would be the same conclusion as that which says that one ought to be tolerant of, and not interfere with, the actions and views of people in another society where those actions and views stem from the moral system which prevails in that society. We ought to be tolerant because their moral system is just as good as ours.

But, as Harrison (1976 p.240) points out, the phrase 'as good as' is highly ambiguous - it "can be taken in a moral or non-moral sense". In the moral sense it might mean that the actions and views of those in another society are morally as good as our own; we could then have no moral reason for interfering or failing to be tolerant. But this sense simply builds the conclusion into the premiss. In the non-moral sense it might mean that each moral system was equally consistent and coherent (or rational). But the fact that one system of morality (or immorality or amorality) may be no more and no less rational than another does not entail that those committed to one system ought to be tolerant of the actions and views of those committed to another.

The degree of tolerance one has towards the actions and views of others will be determined by one's **moral** judgment not by one's judgment as to the rationality of their moral (or immoral or amoral) views. If our moral judgments did not determine our actions - if they did not determine the way in which we interact with, seek to influence or educate others - then they would not be moral judgments. Insofar as our actions and views derive from a moral system we will, and we ought to, influence and educate others accordingly.

This is not to argue against tolerance or to advocate intolerance - it is merely to argue that tolerance is not to be derived from moral relativism. Nor is it to say

that we ought to seek to influence the views of others in an authoritarian way, or by means of coercion and confrontation. The manner in which we seek to influence and educate others will be determined by a variety of factors, and the most important of those will be our moral views.

Educating for benevolence, non-malevolence and humility.

The moral system here outlined will have two facets; one deriving from the application of the universal ideal to ourselves and the other deriving from its application to others. Insofar as we apply the ideal to ourselves we will seek to engage in critical thinking restrained by cognitive humility. In doing so it may be that we will acquire a preference to act in ways which involve preventing, interfering with and controlling the behaviour of others. However, there are many factors which may lead us to prefer an alternative action. Harrison (1976 p.242) lists some of those factors but I will modify and elaborate the list in the light of the moral theory offered here:

> in most cases an action by a person which satisfies that person's own preferences will be the right action;
>
> in many cases interfering with an action will create ill-will and further consequences sufficient to outweigh the benefits of interference;
>
> in some cases the right action will not involve interfering with or preventing the actions of others but will involve compensating for, or preventing, some of the consequences of that action;
>
> in many cases our moral judgment may not be correct and the judgment of the other person will result in right action;
>
> in some cases interference and control may inhibit moral development.

A community of (imperfect) benevolent archangels.

The first four of these factors give reason to refrain from interference which is wilful, trivial, unnecessary, or lacking in humility. However, the extent to which they influence our interaction with others will, clearly, depend upon the abilities of those with whom we are dealing. An educator is likely to be dealing with those whose preferences are often not well-informed, the on-going relationship with the educatee may allow ill-will to be countered, and the judgment of the educator is more likely to be correct. In all cases it will, nevertheless, be possible for the educator to try to not only interfere with and control behaviour but also to make clear the way in which that interference and control stems from critical thinking. This possibility will also play a part in determining the extent to which the final factor in the list influences interaction.

That final factor (not included in Harrison's list) brings us to the other facet of the moral system. In applying the ideal to others we will seek to encourage **their** focus upon decisive preferences, critical thinking and cognitive humility. The way in which we respond to the behaviour of others will not only be influenced by the application of the ideal to ourselves (in the ways outlined above) but will also be guided by this further aim. It is likely that the most powerful factor in this respect will be the opportunities which are provided for witnessing and experiencing the benevolence, non-malevolence and cognitive humility of others.

It is clear that interference with, or control of, the behaviour of others may sometimes contribute nothing towards moral development. This will be true whether the control is direct or whether it is by means of, say, reward and punishment. Such control may also be counter-productive. This is clearly the case, for example, of punishment which is vindictive, relentless or

humiliating. Whilst such punishment may (in some cases) produce a long-term inhibition from certain types of behaviour, the person experiencing such punishment is witnessing and experiencing, at best, a marked lack of benevolence or, at worst, clear malevolence.

But control may be counter-productive in less obvious ways. Docking (1987 p.111-112), when considering the advantages and disadvantages of techniques of behaviour modification, points out that just as providing extrinsic rewards for a task such as drawing may undermine a child's wish to engage in the activity for its own sake, so too behaviour modification programmes may create problems of motivation. "The crucial question", (Docking suggests) "is this: Does the approach not only seem to get the child acting more acceptably but also help him to view his behaviour and those of others in a different light?".

Peters (1974 p.151-2), when considering the development of a disposition to act according to rules or principles, points out that a desire to strengthen such dispositions may lead us to discourage any questioning of the validity of the rules and thereby to inhibit development of an ability to see the reasons for the appropriateness of the rules. What we ought to seek to develop, according to Peters, is an awareness of those features of a course of conduct "which constitute a non-artificial reason for .. decision and judgment, as distinct from extrinsic associations provided by praise and blame, reward and punishment, and so on".

This is not, of course, to argue against employment of a whole range of means of controlling or influencing behaviour. But it is to argue against losing sight of the central aim of moral development. It is to point out that if we lose sight of that aim then our interactions with others, however successful and laudable they may be

in terms of controlling and influencing behaviour, will contribute nothing to that development and may actually impede that development. In terms of the moral theory here outlined, it is to emphasise that:

a) judgments as to appropriate means of control and influence ought to derive from **our** benevolence, non-malevolence and cognitive humility;

but

b) moral development of others rests not upon the successful control and influence of behaviour but upon the development of **their** benevolence, non-malevolence and cognitive humility.

The aim of moral education is to influence educatees in such a way that they will be motivated by knowledge of consequences and preferences; to ensure that for them consequences matter, preferences matter, but malevolent preferences count for nothing. Such an education will require not only opportunities to gain knowledge of consequences and preferences, and to be motivated by them, but also opportunities to witness and experience such motivation in others. The educator will need to respond to, and create, opportunities in which such motivation can be made explicit.

There has been a great deal of research into the development of benevolent identification (called 'empathy' by Bottery and others). As mentioned in the previous chapter, such research points to evidence that empathy is present at very early stages of development, and evidence that the emergence of a cognitive grasp of alternative viewpoints is not a separate and earlier stage of development. Bottery (1990 Ch.7) also outlines some of the suggestions which have been made with regard to the various modes and phases which might be involved in the development of empathy with cognition, the ways in which it may be evoked, and the techniques which might be employed to help development. I have neither the space

nor the expertise to explore these suggestions here. What is clear is that unless we create appropriate opportunities for educatees to witness, experience and engage in actions motivated by knowledge of consequences and awareness of the preferences of others then moral development is unlikely to take place.

However, what is also clear is that development of knowledge of consequences and awareness of preferences is not sufficient. Awareness of preferences may result in **malevolent** identification and knowledge of consequences may then serve that malevolence. Bottery (1990 p.67) says: "Techniques must .. be developed stimulating the empathic abilities and especially the emotional type. Then not only will children understand a situation as another views it, but will also see how that person feels it as well, and so be motivated to do something about it". But to see how a person is, say, feeling distress may, alas, result in a desire to see that distress continue. If empathy is the ability to gain knowledge and understanding of the feelings and preferences of another then it can result in malevolent identification. If empathy is **sharing** the feelings and preferences of another then it is benevolent identification but it goes beyond mere knowledge and understanding.

Awareness, knowledge and understanding of the feelings and preferences of others is not sufficient for benevolent identification. The educator must encourage such benevolence and discourage malevolence. There is much psychological evidence to indicate that the fostering of one's own self-esteem is a crucial factor in the development of benevolence towards others. One might speculate that it is also the case that experiencing the benevolence of others towards oneself is a crucial factor in the development of benevolence towards others **and** in the development of that self-esteem.

Such development might also bring with it a rejection of malevolence. However, all of us (even the most benevolent) are prone to malevolence. When, for example, someone has caused our suffering it is, at the very least, difficult not to gain satisfaction when they in turn suffer. The educator will need to encourage an ability to recognise and guard against the many ways in which malevolence may manifest itself. Just as the development of benevolence is likely to require opportunities to experience, witness and be involved in actions explicitly motivated by benevolence; so too the development of non-malevolence is likely to require a similar exposure to actions and responses involving an explicit rejection of malevolence. One might again speculate that the most crucial factor may be experiencing the fact that malevolence towards oneself arouses the indignation of others.

The emphasis upon benevolent identification leads to an approach to moral education which is very similar to that of Wilson (1967). He identifies several components which are necessary to a consideration of moral problems: counting other people's feelings and interests as of equal validity with our own; awareness and insight into one's own and other people's feelings; knowledge of what is likely to occur if one acts on one's feelings in this or that way; formulation of, and commitment to, a set of rules relating to other people's interests; and the ability to act, to live up to, those rules and principles.

I too have argued that just this sort of caring and concern lies at the heart of morality. But I would also emphasise that recognition and rejection of evil (recoiling from malevolence) is as central to morality as that caring and concern (the inclination to benevolence). Educators will need (as Noddings says - 1989) to pay attention and, at some stage, to draw attention to the

cruelty, the torture, the hatred for other individuals or groups (marked out by race, gender, nationality or religion) which we see around us now and throughout human history.

Equally important, however, is learning to be restrained by cognitive humility. In making motivation explicit the educator must make clear the knowledge, and the limitations of knowledge, which are involved. The benevolence, concern and love which we have for one another, **and** our responses to (and observations upon) the benevolence and malevolence we detect in others, must contain humility.

We must learn that it is often the case that we do not know all the preferences of others and we do not know all the consequences of our actions. It is our awareness of the limitations of such knowledge which ought to dispose us to focus upon decisive preferences and to be guided by rules and principles of action. If our knowledge were greater then perhaps we could love our neighbours as we love ourselves but we must be ready to acknowledge our imperfection.

We must also learn that we do not have any great insight into the knowledge of preferences and consequences which underlies, or fails to underlie, the actions of others. If our insight into others were deeper then perhaps we could judge our neighbours as we judge ourselves but, again, we must be ready to acknowledge our imperfection.

Love, humility and assessment.

Benevolent identification with the preferences of others brings with it a caring and concern for others. Our morality requires such concern. But morality does not require that we share all the concerns of others, nor

that their joys and sorrows are always our joys and sorrows. To always seek knowledge of the preferences, joys and sorrows of another; to share all the joys and sorrows of another; to always seek opportunities in which we may be able to bring joy to, and to prevent or alleviate the sorrow of, another - these are the characteristics of love not of morality.

If we were benevolent archangels then, perhaps, we would have such love for all. As a benevolent archangel we would know, and share, all the preferences of everybody in each situation; we would, therefore, share all their joys and sorrows. We would also know all the consequences of all possible actions in each situation; we would, therefore, know when and how it would be possible to bring joy to, and prevent the sorrow of, others. Such a being could not 'take seriously' the separateness of persons. Such a being would, of course, know that we are each different persons but he would love us all equally and thus always treat all of our equal interests and preferences as of equal weight (see Hare 1990 p.257). But we are not such beings. The scope and extent of our concern is determined not only by the ideal of the benevolent archangel but also by our imperfection.

We take seriously the separateness of persons when we have cognitive humility. We ought to have such humility even in our dealings with those for whom we have the closest attachment. We know less than we think of the preferences of our loved ones; and we know very little of all but the decisive preferences of others. All too frequently we also have a thoroughly imperfect grasp of when and how we might bring joy to, or alleviate the sorrow of, others. In our dealings with others, and in our close attachments, we can only aspire to love.

In most situations we ought, therefore, to be guided and motivated by our knowledge of the decisive preferences of

others and our knowledge of the ways in which our actions may affect the satisfaction of such preferences. However, it may be that in some situations we can engage in perfect critical thinking. If this is so then at such moments we too would not, and ought not to, take seriously the separateness of persons - it is as if we loved each equally.

Murdoch (1970) claims that "we need a moral philosophy in which the concept of love, so rarely mentioned by philosophers, can once again be made central". She says that, although love is the source of our greatest errors, "it is the energy and passion of the soul in its search for Good .. its existence is the unmistakeable sign that we are spiritual creatures, attracted by excellence and made for the Good" (Murdoch 1970 p.103). I would claim, perhaps with Murdoch, that through sharing and experiencing the love a person can have for another we have glimmerings of our perfectibility. I would also claim that the ideal of the benevolent archangel is an ideal of love and that morality is, in a sense, love written small (the love of imperfect beings). However, Murdoch's notion of love is very different to the notion which is encapsulated in that ideal.

Murdoch holds that in loving others we do not merely overcome our selfish concerns but we also suppress our own self will. That will - "the avaricious tentacles of the self" - is, for Murdoch, the source of evil and of blindness. Human will is relentlessly concerned with looking after itself and with fabricating a veil with which it conceals the world (1970 p.78-79). Through pureness and meekness we may suppress that will and thus overcome our evil and achieve clarity of vision. Murdoch draws a parallel with art and science: with clarity of vision comes appreciation of beauty, knowledge of truth, and compassion for others. The search for Good involves

striving for a **selfless** attention to art, nature and others.

Murdoch's thoroughly pessimistic view of human will has much in common with Schopenhauer, but her view is tied in with arguments for moral realism. Morality and goodness require attention to reality, an ability to perceive what is true, and that, she claims, is automatically at the same time a suppression of self.

With true vision comes right conduct: "the more the separateness and differentness of other people is realised, and the fact that another man has needs and wishes as demanding as one's own, the harder it becomes to treat a person as a thing" (Murdoch 1970 p.66). To see justly and clearly requires that we turn away from self; to see thus is to love and to be thereby both liberated from fantasy and motivated to act.

But, I have claimed, the malevolent archangel sees clearly, and knows the needs and wishes of others. Clarity only brings right conduct if it is motivated by love. Love does not require the suppression of self-will and a detachment from selfish concerns. Love is the **extension** of those concerns, it is to have the concerns of others as one's own and **alongside** one's own.

For Murdoch humility is the suppression of self - the absence of those avaricious tentacles. The humble man sees himself as nothing and thus sees other things as they are (1970 p.104).

I would claim that the role of humility in morality is to enable us to see not that we are nothing but that we are unavoidably imperfect. We can seldom truly love, we can seldom have the concerns of others as our own. We, therefore, recoil from the illusion of love; submit ourselves to moral principles as motivated by the central

needs and decisive preferences of others; and thus we do indeed suppress our own self-will. But we do so only because we recognise our imperfection and recognise that we cannot meet the demands which love would place upon us. We cannot love others as we love ourselves.

Such humility limits not only the demands we place upon ourselves but also those which we place upon others. Furthermore, that humility ought to constrain the judgments we make of the moral worth of others and of their actions. We may make our judgment of the rightness or wrongness of the actions of others, but in order to judge moral worth we must judge their knowledge and benevolence - that is, the quality of their motivation. Our humility ought to lead us to hesitate when judging and assessing the morality, and the moral progress, of those whom we educate.

Many aspects of educational development are difficult to assess. As Bottery (1990 p.123/4) points out, assessment of, for example, an appreciation of drama may require evaluation of achievement of objectives which are varied, vague, complex and unpredictable. Furthermore, as he also points out, assessment of, say, a readiness to cooperate with others may require not only observation of but also interaction with those educated, and may require that both extend over a considerable period of time and over a range of activities.

It is clear that assessment of moral development shares such difficulties. Moreover, if moral development is centrally about motivation, and not merely about achievement of behavioural objectives, then there will be another layer of difficulty. Do those who tell the truth do so because they hope for reward or fear punishment? Do those who fail to tell the truth (on an occasion when the educator believes that the truth ought to be told) do so because of a disregard for the preferences of those

deceived, or a lack of knowledge of preferences and consequences, or a greater knowledge than that of the educator?

The educator's desire, and the desire of others, to know what progress is being made may lead to a demand for forms of assessment and evaluation which, whilst being more easily achieved, have little or nothing to do with assessment of moral development. This may then lead to an inadvertent failure to prioritise achievement of moral development. But the difficulties of assessment may also lead directly to the rejection of such a priority. The educator may decide to prioritise only those objectives which are such that achievement is readily evaluated and demonstrated.

A clarification and systematisation of our moral views in terms of a fundamental universal ideal may help us to avoid inappropriate (or over-ambitious) forms of assessment and may inspire us to resist abandonment of those objectives which our moral views require. If we have the cognitive humility which comes from recognition of the ways in which we cannot live up to that ideal then we know that we cannot judge others as we judge ourselves. If we have a moral outlook which is inspired by that ideal then we know that we ought to foster the moral development of those whom we educate just as we ought to foster such development in ourselves.

The priority which our own moral development has for us will determine the priority which we give to the moral development of those whom we educate. If our moral development has a high priority for us then we will strive for morality in ourselves and in others. Finally it is, perhaps, through witnessing and sharing **our** struggle for morality that the morality of others is best fostered.

Preferability of a community of benevolent archangels.

Although I have argued that there is no 'ultimate justification' for any moral system, I have thus far assumed that there is a large measure of agreement over moral views with regard to behaviour, dispositions and motivations; and I have argued that, given that agreement, we have reason to systematise those moral views in terms of a universal ideal. However, I may have over-estimated the extent of agreement over moral views and, even more likely, I may not be able to convince others to adopt that perspective upon those moral views. I would like to end by offering a few considerations and speculations which might convince some educationalists to adopt that perspective.

Most of us (I speculate) would prefer a life guided by the ideal of the benevolent archangel. If we could attain a clear view of the nature of such a life then most of us would prefer it. Mill (1863 p.31) claims that those who have a care and concern for others regard those feelings as ones which it would not be well for them to be without; that view of those feelings is, for Mill, the "ultimate sanction of the greatest happiness morality". Perhaps we can claim that to clearly see ourselves as guided by informed benevolence, and as lacking in malevolence, would be (for most of us) to prefer to be thus.

But most of us would also prefer a life in which our preference satisfaction was maximised. A life guided by the ideal is likely to have a cost on the scale of my overall preference satisfaction. In a world in which selfishness, unscrupulousness, corruption and evil are, to say the least, not unknown the cost may be great. One living such a life may often sacrifice self-interest out

A community of (imperfect) benevolent archangels.

of consideration for others but may seldom receive such consideration themselves, may be abused and mistreated by others who confidently expect not to be abused and mistreated in return, and so on. In short, the decrease in that person's own overall preference satisfaction may be sufficient to outweigh the preference for such a life.

This may not be true for some. On the one hand there may be those who are able to attain a clear view of the nature of such a life and yet would not find such a life preferable even if there were no cost. On the other hand there may be those (saints) for whom the preferability of such a life is so great that it would outweigh any cost. Most of us (I speculate) lie somewhere in between. Most of us would find such a life preferable even if there was likely to be some cost in terms of our own overall preference satisfaction. If we were able to attain a clear view of such a life then a choice of that life would not require (contra Hare and others) a conviction that it would be (or would be likely to be) in our own self-interest.

Most of us would prefer to live in a community largely consisting of people whose lives were guided by the ideal. Whether or not our own life were guided by that ideal, we would prefer to live in such a community because it would be conducive to the satisfaction of our own preferences. But if our life were guided by that ideal then we would have further reason to prefer that the life of each member of the community be guided by the ideal. Firstly, we would benevolently identify with the informed preferences of others for such a life. Secondly, we would benevolently identify with the preferences of others and thus prefer a community which was conducive to the satisfaction of those preferences.

Most would prefer a life guided by the ideal, and most would prefer to live in a community largely consisting of

people whose lives were guided by the ideal. Those who have the first preference have a stronger and deeper reason for having the second. Once we acknowledge that it would not be well to live a life which did not involve a struggle inspired by that ideal then we have a compelling reason to strive for a community which is also inspired by that ideal. We strive for that community by educating ourselves and others according to that ideal. We do so because we glimpse the preferability of life as an (imperfect) benevolent archangel in a community of (imperfect) benevolent archangels.

BIBLIOGRAPHY

Anscombe G		1958	'Modern Moral Philosophy'	**in**
Hudson WD	(ed)	1969	The Is-Ought Question	London,Macmillan
Ayer AJ		1972	Probability and Evidence	London,Macmillan
Bentham J		1789	Introd. to the Principles of Morals and Legislation	London,Athlone Press
Bentham J		1834	Deontology	Oxford,Oxford University Press
Blackburn S		1971	'Moral Realism'	**in**
Casey J	(ed)	1971	Morality and Moral Reasoning	London,Methuen
Blackburn S		1985	'Errors and the Phenomenology of Value'	**in**
Honderich T	(ed)	1985	Morality and Objectivity	London,Routledge & Kegan Paul
Bottery M		1990	The Morality of the School	London,Cassell
Bradley FH		1876	Ethical Studies	Oxford,Oxford University Press
Brandt RB		1979	A Theory of the Good and the Right	Oxford,Oxford University Press
Brandt RB		1988	'Act-Utilitarianism and Metaethics'	**in**
Seanor & Fotion	(ed)	1988	Hare and Critics	Oxford,Clarendon Press
Caldwell & Spinks		1988	The Self-Managing School	London,Falmer Press
Chamberlin R.		1989	Free Children and Democratic Schools	London,Falmer Press
Clark C		1987	'Why Teachers Need Philosophy'	unpublished research paper
Dainton Committee		1968	Flow of Science & Technology Candidates into H.E.	D.E.S.
Dearden RF		1968	The Philosophy of Primary Education	London,Routledge & Kegan Paul

Docking JW		1987	Control and Discipline in Schools	London,Harper and Row
Dworkin R		1977	'The Double-Counting Objection'	**in**
Glover J	(ed)	1990	Utilitarianism and Its Critics	New York,Macmillan
Foot P		1958	'Moral Beliefs'	**in**
Foot P	(ed)	1967	Theories of Ethics	Oxford,Oxford University Press
Frankfurt H		1971	'Freedom of the Will and the Concept of a Person'	Journal of Philosophy,67(1),5-20
Glover J	(ed)	1990	Utilitarianism and its Critics	New York,Macmillan
Hampshire S		1978	Public and Private Morality	Cambridge,Cambridge Univ'ty Press
Hare RM		1952	The Language of Morals	London,Oxford University Press
Hare RM		1963	Freedom and Reason	Oxford,Oxford University Press
Hare RM		1976	'Ethical Theory and Utilitarianism'	**in**
Sen & Williams	(ed)	1982	Utilitarianism and Beyond	Cambridge,Cambridge Univ'ty Press
Hare RM		1981	Moral Thinking	Oxford,Oxford University Press
Hare RM		1985	'Ontology in Ethics'	**in**
Honderich T	(ed)	1985	Morality and Objectivity	London,Routledge & Kegan Paul
Hare RM		1988	'Comments'	**in**
Seanor & Fotion	(ed)	1988	Hare and Critics	Oxford,Clarendon Press
Harrison G		1976	'Relativism and Tolerance'	**in**
Krausz & Meiland(ed)		1982	Relativism - Cognitive and Moral	Indiana,Univ.of Notre Dame Press
Harsanyi JC		1988	'Problems with Act-Utilitarianism & Malevolence'	in
Seanor & Fotion	(ed)	1988	Hare and Critics	Oxford,Clarendon Press

Hart HLA		1979	'Comments on the Double-Counting Objection'	**in**
Glover J	(ed)	1990	Utilitarianism and Its Critics	New York,Macmillan
Honderich T	(ed)	1985	Morality and Objectivity	London,Routledge & Kegan Paul
Hudson WD		1970	Modern Moral Philosophy	London,Macmillan Press
Hume D		1888	A Treatise of Human Nature	Oxford,Oxford University Press
Jarrett J		1991	The Teaching of Values	London,Routledge & Kegan Paul
Jenkins HO		1991	Getting it Right	Oxford,Basil Blackwell
Kagan J	(ed)	1987	The Emergence of Morality in Young Children	London,Univ'ty of Chicago Press
Kant I		1781	Critique of Pure Reason	**translated by**
Smith NK		1929		London,Macmillan
Kant I		1785	Fundamental Principles of the Metaphysics of Morals	**translated by**
Abbott TK		1949		Indianapolis,Liberal Arts Press
Kant I		1788	Critique of Practical Reason	**translated by**
Beck LW		1956		Indianapolis,Bobbs Merrill
Kierkegaard S		1843	Either / Or	**translated by**
Lowrie W		1959		New York,Anchor Books
Korner S		1955	Kant	Harmondsworth,Penguin Books
Korner S		1960	The Philosophy of Mathematics	London,Hutchinson
Krausz & Meiland(ed)		1982	Relativism - Cognitive and Moral	Indiana,Univ.of Notre Dame Press
MacIntyre A		1967	A Short History of Ethics	London,Routledge & Kegan Paul

MacIntyre A		1981	After Virtue	London,Duckworth
Mackie J		1974	The Cement of the Universe	Oxford,Oxford University Press
Mackie J		1977	Ethics: Inventing Right and Wrong	Harmondsworth,Penguin Books
Melden AI		1961	Free Action	London,Routledge & Kegan Paul
Mill JS		1863	Utilitarianism	London,Dent & Sons
Moore GE		1966	Ethics	Oxford,Oxford University Press
Murdoch I		1970	The Sovereignty of Good	London,Routledge & Kegan Paul
Nagel T		1982	'The Excessive Demands of Impartiality'	London Review of Books,82/83
Nagel T		1986	The View from Nowhere	Oxford,Oxford University Press
Noddings N		1989	Women and Evil	Los Angeles,California Univ Press
Parfit D		1984	Reasons and Persons	Oxford,Oxford University Press
Passmore J		1980	The Philosophy of Teaching	London,Duckworth
Peters RS		1972	'General Editor's Note'	**in**
Hirst & Peters	(ed)		Education and the Development of Reason	London,Routledge & Kegan Paul
Peters RS		1974	'Moral Development and Moral Learning'	**in**
Peters RS		1981	Moral Development and Moral Education	London,George Allen & Unwin
Platts M		1979	Ways of Meaning	London,Routledge & Kegan Paul
Putnam H		1975	Mathematics, Matter and Method	London,Cambridge University Press

Rawls J		1955	'Two Concepts of Rules'	**in**
Foot P	(ed)	1967	Theories of Ethics	Oxford,Oxford University Press
Rawls J		1971	A Theory of Justice	Oxford,Oxford University Press
Raz J		1975	'Reasons for Action, Decisions and Norms'	**in**
Raz J	(ed)	1978	Practical Reasoning	Oxford,Oxford University Press
Russell B		1919	Introduction to Mathematical Philosophy	London,George Allen & Unwin
S.M.P.		1963	'Director's Report'	Secondary Maths Project,62/63
Scheffler I		1973	Reason and Teaching	London,Routledge & Kegan Paul
Scheffler S		1982	The Rejection of Consequentialism	Oxford,Clarendon Press
Scheffler S	(ed)	1988	Consequentialism and its Critics	Oxford,Oxford University Press
Schopenhauer A		1851	Essays and Aphorisms	**translated by**
Hollingdale RJ		1970		Harmondsworth,Penguin Books
Seanor & Fotion	(ed)	1988	Hare and Critics	Oxford,Clarendon Press
Sen & Williams	(ed)	1982	Utilitarianism and Beyond	Cambridge,Cambridge Univ'ty Press
Sen A		1980	'Plural Utility'	**in**
Glover J	(ed)	1990	Utilitarianism and Its Critics	New York,Macmillan
Sidgwick H		1874	The Methods of Ethics	London,Macmillan
Singer P		1988	'Reasoning towards Utilitarianism'	**in**
Seanor & Fotion	(ed)	1988	Hare and Critics	Oxford,Clarendon Press
Skyrms B		1975	Choice and Chance	California,Dickenson

Smart & Williams		1973	Utilitarianism For and Against	Cambridge,Cambridge Univ'ty Press
Sullivan R		1989	Immanuel Kant's Moral Theory	Cambridge,Cambridge Univ'ty Press
Taylor C		1985a	Human Agency and Language	Cambridge,Cambridge Univ'ty Press
Taylor C		1985b	Philosophy and the Human Sciences	Cambridge,Cambridge Univ'ty Press
Taylor C		1991	The Ethics of Authenticity	London,Harvard University Press
Walker R		1978	Kant	London,Routledge & Kegan Paul
White J		1982	The Aims of Education Restated	London,Routledge & Kegan Paul
White J		1990	Education and the Good Life	London,Kogan Page
Williams B		1985	Ethics and the Limits of Philosophy	London,Fontana
Williams B		1985a	'Ethics and the Fabric of the World'	**in**
Honderich T	(ed)	1985	Morality and Objectivity	London,Routledge & Kegan Paul
Williams B		1988	'The Structure of Hare's Theory'	**in**
Seanor & Fotion	(ed)	1988	Hare and Critics	Oxford,Clarendon Press
Wilson J (et al)		1967	Introduction to Moral Education	Harmondsworth,Penguin Books