THE FALLIBILIST THEORY OF VALUE AND ITS APPLICATION TO DECISION MAKING

A thesis for the degree of Doctor of Philosophy

April 1979

THE UNIVERSITY OF ASTON IN BIRMINGHAM

David George COLLINGRIDGE

THE FALLIBILIST THEORY OF VALUE AND ITS APPLICATION TO DECISION MAKING

A thesis for the degree of Doctor of Philosophy

David George COLLINGRIDGE

April 1979

SUMMARY

The aim is to develop a fully general theory of value, showing particularly how value judgements may be tested, using Popper's theory of scientific method as a model, and to explore the the application of the general theory to decision making.

Justificationist accounts of value are rejected (Part I). Justification requires some fundamental value claims to terminate the regress which finding support for a value judgement generates, but these cannot exist. In addition, all but very weak sets of value judgements have factual consequences and so cannot be regarded as justified since they always stand in jeopardy of falsification through these consequences. Recent attempts to overcome these problems are reviewed and none found adequate.

A fallibilist account of value is then developed (Part II). Value judgements are to be tested by exposing them to criticism which is potential falsification from factual sentences. There is, therefore, a premium on value judgements of high universality and precision. Methodological rules are necessary to ensure that no value judgement can escape criticism. Success for a value judgement is the survival of criticism and may be measured by degree of corroboration.

In Part III the methodology is extended to cover decision making. Standard theories of decision making are criticised because they incorporate a false view of individual values; fail to satisfactorily connect individual and social values; and require great quantities of factual input not generally available in real world decision situations. The fallibilist theory of decision making can overcome all these problems. Decisions are rational, not if they optimise some objective function, but if they are submitted to critical assessment by the methodology developed earlier. Since any decision may prove wrong, there is a premium on decisions which may be reversed easily and measures for reversibility are developed.

ethics; decision making; flexibility; Popper

CONTENTS

OUTLINE

Part I - The Failure of Authority

Traditional views of values have attempted to show how values may be justified. Such justification is not possible.

Chapter 1 - The Failure of Authority

Justification of a value judgement is only possible if the regress of reasons generated by any attempt at justification can terminate. Thus justification is traditionally seen as stemming from some authority. Such termination is not possible.

Chapter 2 - Attempts at Justification

A survey of the main responses made to the problem of chapter 1 over the last c. 50 years. None of them is satisfactory.

Chapter 3 - Factual Error and Justification

All but very weak sets of value judgements have factual consequences and so stand in jeopardy from falsification, so they cannot be justified (see chapter 5).

Part II - Criticism and Value

Develops a fallibilist theory of value modelled on Popper's theory of science.

Chapter 4 - Popper's Theory of Science

A brief outline of the features important for the development of the fallibilist theory of value.

Chapter 5 - The Autonomy of Value

Value judgements may have factual consequences, counter to Hume's Rule and the autonomy of values, and so may be falsified by factual discoveries.

Chapter 6 - A Fallibilist Theory of Value

Value judgements are to be tested by exposing them to criticism = potential falsification from factual sentences. There is, therefore, a premium on highly universal, precise value judgements. Methodological rules are necessary to ensure that no value judgement can escape criticism. Success for a value judgement = surviving criticism and this may be measured by degree of corroboration.

Chapter 7 - Case Studies

Case studies applying the findings of chapter 6 to (1) A debate on corporal punishment; (2) Schumacher's Buddhist economics; (3) the debate between the U.S. EPA and Ethyl Corporation.

Part III - Criticism and Decision

Contemporary views of decision making are criticised, and the fallibilist theory of value developed into an account of how decisions should be taken.

Chapter 8 - Contemporary Decision Theory I - Individual and Social Values

Contemporary approaches to the problem of basing social upon individual values are criticised, especially the work of Harsanyi, Rawls and welfare economists.

Chapter 9 - Contemporary Decision Theory II - Fact and Forecast

Contemporary decision theories require so much knowledge to be available before they can assist in decision making that they are useless for real decision problems.

Chapter 10 - Towards A Fallibilist Theory of Decision Making

Making a decision is a special case of adopting a value judgement, so the theory of chapter 6 applies to decision making. Decisions are peculiar, however, because their revision imposes costs. For this reason easily reversible decisions are to be favoured. A measure for reversibility (flexibility) is developed and examples given of its application.

Chapter 11 - Contemporary vs Fallibilist Theories of Decision Making

Advantages of the fallibilist theory of decision making over contemporary ones include: much more modest requirements for factual information; no dependence on subjective probabilities; much less dependence on predicting future events; a much better account of the relationship between individual and social values is possible; and a closer match between theory and decision making in practice.

PART I

THE FAILURE OF AUTHORITY

CHAPTER ONE - THE FAILURE OF AUTHORITY

Most philosophical theories of value so far developed have been justificationist theories. They have, that is, attempted to show how value judgements of some sort can be justified, grounded or established. It is one of the central claims of this work that all such theories are fundamentally mistaken, and that no value judgements of any kind can be justified. Instead of searching for ways of justifying value judgements, we should seek to develop a fallibilist theory of of evaluation which shows how such judgements may be criticised and assessed by their resistance to criticism. In what follows I shall outline such a theory of value, basing my account squarely on Popper's theory of scientific method. Realising the inadequacy of justificationist views of science, with their dependence on protocol sentences and the like, and inductive logic, Popper developed a theory of science according to which scientific claims cannot be justified, but can be tested by exposure to criticism. His account of science, therefore, provides a model for the fallibilist theory of value to be discussed below.

Facts and values must be brought together in the making of a decision, so that if value judgements cannot be justified, then neither can decisions be justified. If decisions are to be open to rational assessment at all, then this assessment can only

be based on criticism. My motivation for developing a fallibilist theory of value is that it can be expected to throw light on the assessment of decisions through criticism. This will become clear towards the end of this work, but this central interest in decisions must be remembered in earlier places, and especially in the next chapter.

In the first part, I hope to establish my sceptical claim that no value judgement can be justified. This will occupy three chapters. The present one contains the most general argument for scepticism, whilst the next will expand on this by looking at the failure of recent attempts to show how value judgements may be justified. The third chapter relies on a result to be established later - that value judgements may have factual consequences - to urge scepticism about all but trivial sets of values.

The sceptical arguments of this chapter depend in no way upon a special characterization of value judgements. If my arguments are correct, then whatever analysis or description is offered of value judgements, no such judgements can be justified. Similarly, the arguments in no way depend upon a particular view of justificatio Whatever account is given of the exact nature of justification, scepticism about the possibility of justifying value judgements is inevitable. This enables me to couch the sceptical arguments of this chapter in a very general way. Although this reveals the power and scope of the arguments, it does so at the cost of making them

sound remote and artificial. In remedy, I shall give a number of illustrations to show how the arguments count against justificationist accounts of value which have been given in the past. This will be done extensively in the next chapter, although one or two illustration will be given here to fill out the skeleton of the general arguments.[1]

1. Fundamental Reasons

The central element in the arguments which I wish to deploy against the possibility of value judgements being justified is what I call a regress of reasons. Imagine that a value judgement of some kind, call it J, is to be justified. Reasons for regarding J as justified must be brought forward. As far as the arguments go, there is no need to specify what kinds of things may be used to justify a judgement such as J; it suffices that reasons of some sort are necessary if J is to be justified. What kinds of things may justify value judgements, and, in particular, whether these must be other value judgements, factual sentences, commands or other imperatives need not, therefore, detain us. Neither need we show concern for the exact nature of the relationship of justification. Reasons are needed for the justification of a value judgement; once this much is admitted, we have enough. Suppose R_1 is proposed as a reason which justifies J. Whatever account is given of justification, it must be admitted that R_1 justifies J only if R_1 is itself justified. If R_1 is not justified, we have no reason for

accepting it, and hence no reason for thinking $\underline{J}$ justified. Suppose, therefore, that $\underline{R}_2$ is produced as a reason justifying $\underline{R}_1$. As before, this is only possible if $\underline{R}_2$ is itself justified, which will lead to the invoking of a further reason, $\underline{R}_3$, and so on. The regress ----$\underline{R}_3$, $\underline{R}_2$, $\underline{R}_1$, $\underline{J}$ I shall call a regress of reasons.

The next step is to show that $\underline{J}$ can only be justified if there is some non-infinite regress of reasons leading to $\underline{J}$. Imagine that the regress of reasons above is infinite, so that the need to provide reasons never stops. In that case any attempt to justify $\underline{J}$ must be conditional; i.e. all we can say is something like 'if $\underline{R}_4$ is justified, then $\underline{J}$ is justified'. This, however, does nothing but raise the question of the justification of $\underline{R}_4$. If $\underline{R}_4$ is justified, then so too is $\underline{J}$, but if $\underline{R}_4$ is not justified, we have no reason to think $\underline{J}$ justified. There can be no more assurance about $\underline{J}$ than there is about $\underline{R}_4$, but once the question of $\underline{R}_4$'s justification is raised, we are back into the regress. In other words, the sort of conditional justification which an infinite regress of reasons permits is no justification at all.

If $\underline{J}$ is to be justified, it follows that there must be some regress of reasons leading to $\underline{J}$ which terminates at some point so that it is non-infinite.[2] If justification of $\underline{J}$ is possible, there must exist some reason, $\underline{R}_F$, which can be justified without appeal to further reasons. Such a reason I shall call a <u>fundamental</u> reason, avoiding the confusion which might result from the use of such words as self-evident, <u>a priori</u>, certain, self-certifying and so on. The

word fundamental is appropriate here. A value judgement is justifiable only if there exists some fundamental reason which justifies it, though perhaps through a chain of intermediate reasons. It is natural, therefore, to see such fundamental reasons as providing a basis for values, or as having authority in evaluative questions, or as a source for proper values.[3]

This much will be familiar to most people having a nodding acquaintance with ethics and, as we shall see in the following chapter, the search for fundamental reasons has been one of the main driving forces in the development of ethical theory, just as the search for a certain source of knowledge has been central to epistemology. The force of the present chapter is that the search for such reasons in ethics has been a search for the impossible. Before seeing why this is so, it will be useful to list the assumption I have made thus far. To me there seem three:

A1 B is justified only if there is some reason supporting B.

A2 If A is a reason supporting B, then A is justified.

A3 No judgement B can be justified simply by a conditional justification of the form 'if A is justified then B is justified'.

These seem reasonable assumptions on any view of justification, so that the fine points of these views need not concern us.

2. The First Sceptical Argument

Imagine that a value judgement $\underline{J}$ is held to be justified because there exists a regress of reasons leading to $\underline{J}$ which terminates at a fundamental reason $\underline{R}_F$. $\underline{R}_F$, being fundamental, is justified, but not by an appeal to further reasons. To press our sceptical intent we need only ask why $\underline{R}_F$ was chosen as fundamental rather than any other reason. This is not an idle question, because chosing some other reason as fundamental might lead to quite a different value judgement being regarded as justified, perhaps one contrary to $\underline{J}$. In reply, it will be argued that $\underline{R}_F$ is fundamental because it has some special property which we may call $\underline{P}$. The next question, of course, is why it is that reasons with the property $\underline{P}$ and not some other property are fundamental. Seeking an answer to this question continues the regress of reasons which $\underline{R}_F$ was supposed to stop. A <u>reason</u> has been given for thinking $\underline{R}_F$ fundamental - that $\underline{R}_F$ has $\underline{P}$, and a further reason - that all reasons with property $\underline{P}$ are fundamental - is obviously needed, but providing reasons for $\underline{R}_F$ in this way means that $\underline{R}_F$ cannot terminate the regress of reasons. The need for these reasons shows that $\underline{R}_F$ is not fundamental after all. Obviously, any attempt to stop the regress at some point beyond $\underline{R}_F$ will run into exactly the same difficulty, so the regress of reasons is an infinite one. Since the argument is stated perfectly generally, all regresses of reasons leading to $\underline{J}$ will be infinite and so $\underline{J}$ will be incapable of justification, as will all value judgements.

The point I am trying to make is a very simple one, as the length of the argument shows. Philosophers have generally agreed that if a value judgement is to be justified then there must exist a regress of reasons leading to the judgement which terminates somewhere. Many philosophers have, therefore, searched for a set of fundamental reasons which can terminate regresses leading to value judgements. The selection of fundamental reasons cannot of course, be arbitrary - reasons must be given why one set of reasons and not another is fundamental. But this leads to contradiction. On the one hand, there are to be fundamental reasons which are justified without appeal to further reasons, and, on the other hand, it is necessary that reasons be given why these reasons, and not others, are fundamental.

It may be helpful to consider a brief example here, whose artificiality I hope may be excused on the ground that more realistic examples will be dealt with in the next chapter. Consider the view that all fundamental reasons which can lead to value judgements are commands of God, so that all such judgements rest ultimately upon God's commands. The relationship which exists between these commands and the value judgements they justify may be taken as entailment. Suppose the value judgement J is entailed by 'do not cause unnecessary suffering' which is supposed to be fundamental. The regress of reasons leading to J, in other words, is supposed to terminate at the command 'do not cause unnecessary suffering'. If we now ask why this command is singled out as fundamental we will be told that it is a command of God. Then if we ask why commands of God are fundamental

we might be told that this follows from God's goodness. We may then, of course, ask how we know that God is good, and so on. The proponent of this view of value has tried to justify J by the finite regress;

avoid unnecessary suffering, ..., ..., ..., J .

But the terminus of the regress cannot be arbitrary or else we could equally argue

cause unnecessary suffering, ..., ..., ..., not-J.

Reasons must be provided for chosing the first terminus rather than the second and so we now have;

..., ..., God is good, All God's commands are fundamental reasons, God commands us to avoid unnecessary suffering, avoid unnecessary suffering, ..., ..., J

But this regress does not terminate at a fundamental reason. We can ask for reasons for thinking that God is good and thus continue the regress indefinitely. Thus the attempt to justify J by appealing to a fundamental reason fails. Reasons are needed for regarding 'avoid unnecessary suffering' as fundamental, and giving

these reasons means that the regress does not terminate after all.

It is of some interest that such a simple point, but one with such damaging consequences, has gone largely unnoticed both in ethics and in epistemology. Part of the reason lies in the shift of topic which occurs at a vital point in the regress. Consider the last regress above. From 'avoid unnecessary suffering' to the right and up to J, all entries in the regress are value judgements. To the left, however, they concern theology, not values. If our principal concern is to show how value judgements may be justified and if some of the sentences in the theological part of the regress seem to provide justification for value judgements, then there is a tendency to avoid the nasty issue of how these theological sentences are to be justified. But, of course, if they are not justified, they can provide no justification for value judgements. We shall see this time and time again in the next chapter where the sceptical arguments developed here will be applied to recent justificationist theories of value. Philosophical views of value are proposed in order to show how value judgements are to be justified, but then the justification of a particular value judgement depends upon whether the philosophical theory is itself justified, which, because of the problems discussed in this chapter, it never is.

The Second Sceptical Argument

The above argument concerns the impossibility of terminating a regress of evidence because of the need to justify the claim that

some set of sentences are fundamental. The second argument is independent of the first, which can be shown by assuming that the difficulties raised above are unreal and that it can be shown that some kinds of sentence, say those with property P, are fundamental. If this is so, then a reason, R_F, can be known to be fundamental and justified without support from further reasons if it can be shown to possess the property P. But, of course, showing that R_F has P is something which demands reasons, and in demanding reasons, the regress of reasons which R_F is supposed to terminate begins all over again. R_F does not, therefore, terminate any regress of reasons, even if our earlier sceptical arguments are ignored. No regress of reasons leading to a value judgement, therefore, terminates and so no value judgement is justifiable.

We may revive the example about God's commands being fundamental in order to illustrate this second sceptical argument. Let us grant that there are no difficulties in justifying the claim that all God's commands are fundamental reasons. If it can be shown that the imperative 'avoid unnecessary suffering' is a God given command, then it can be used to terminate the regress of evidence leading to J, but reasons are needed for believing that the imperative is so commanded. It might be held, for example, that God has revealed this command in the Bible, but then we may ask how this is known and so on and so on. The regress of reasons leading to J obviously cannot halt at 'avoid unnecessary suffering'. It must, in fact, be infinite

as must any regress leading to $\underline{J}$. $\underline{J}$ is, therefore, incapable of justification.

To draw the threads together a little, consider a final example. Consider the theory that all value judgements are ultimately justified by value judgements known directly through intuition. A value judgement so intuited will be a fundamental reason, since no further reasons are needed for its justification - intuition alone suffices for this. Any value judgement entailed by a fundamental reason of this kind will, of course, be justified. Suppose $\underline{J}'$ is held to be entailed by the intuited value judgement $\underline{R}_I$, so that $\underline{J}'$ is justified. The first sceptical argument asks how it is known that all intuited value judgements are fundamental, and providing reasons which justify this claim continues the regress which $\underline{R}_I$ was supposed to terminate. This apart, the second sceptical argument queries how it is known that $\underline{R}_I$ is intuited. As before, this requires reasons, and the giving of reasons continues the regress which $\underline{R}_I$ was supposed to halt.

CHAPTER ONE - FOOTNOTES

1. The doctrine that some regresses of evidence terminate at fundamental sentences so that justification of some sentences is possible is termed foundationalism. I here discuss ethical versions of foundationalism, but it dominates epistemology and the philosophy of mathematics as well. The doctrine has received considerable critical scrutiny recently.
See W. Alston, Has Foundationalism Been Defeated?, Philosphical Studies, 29, 1976, 145-155; L. Bonjour, Can Empirical Knowledge Have a Foundation?, American Philosphical Quarterly, 15, 1978, 1-15; J.W. Cornman, Foundational vs Non-Foundational Theories of Empirical Justification, American Philosophical Quarterly, 14, 1977, 287-297; R. Foley, Inferential Justification and the Infinite Regress, American Philosophical Quarterly, 15, 1978, 311-319; J. Margolis, Skepticism, Foundationalism and Pragmatism, American Philosphical Quarterly, 14, 1977, 119-127; N. Nathan, What Vitiates an Infinite Regress of Justification?, Analysis, 37, 1977, 116-126, and F. Wills, Induction and Justification, Bell, 1974.
For the history of the arguments deployed here see; R. Popkin, The History of Scepticism from Erasmus to Descartes, Dover, 1965.

and P. Feyerabend, On the Improvement of the Sciences and Arts....., Boston Studies in the Philosophy of Science, III, Dordrecht, 1967, and Classical Empiricism in R. Butts and J. Davis (eds.) The Methodological Heritage of Newton, Blackwell, 1970.

2. J could, of course, be equally well a factual or a mathematical claim, so that the arguments here may be generalized to cover all judgements and not just value judgements. In this case, the target is the whole doctrine of foundationalism, and not just its ethical varieties.

3. If J is replaced with a factual or mathematical claim, we have the traditional problems of finding a suitable knowledge base for science and for mathematics.

CHAPTER 2

ATTEMPTS AT JUSTIFICATION

This chapter will consider some recent attempts to show how value judgements may be justified. It is clearly impossible to deal with all attempts, even those made within the past 30 years or so, and so I have chosen those attempts which may be most illuminating given the problem posed by the previous chapter. I shall, of course, be critical of all attempts to show how justification for value judgements is possible and my principal weapon will be the two sceptical arguments developed earlier. In most cases, however, I shall also employ independent arguments to re-inforce the sceptical attack. Here it must be remembered that our principal interest is in the making of decisions, as outlined in the introduction. I shall frequently object to one of the attempts that we will consider, that even if it is successful, it shows how value judgements may be justified only at the cost of severing the link between such judgements and human action. The attempt, even if successful, is then of no interest to us, although, of course, it might be valuable to those whose interests are different from our own.

The aim of this chapter, then, is to illustrate the sceptical arguments developed earlier and, by adducing additional arguments against attempts at justification, to unearth the obstacles which lie in the way of justificationist theories of value. Before beginning to consider these attempts it will, however, be useful to briefly consider other sources of recent scepticism about value.

continued.....

1. Value Judgements Cannot Be Justified

Several writers have urged that no value judgements can be justified, though none has done so on the basis of the kind of sceptical argument developed in the previous chapter. Instead, this scepticism about values has been presented as a consequence of some philosophical theory or a particular analysis of evaluative language. After developing his ethical theory, Moore was forced to admit that there was little chance of ever discovering what our duty and obligations are. Value judgements presented a difficulty for proponents of logical positivism, since essential to their theory was the claim that only sentences verifiable (or, later, confirmable) by sensory experience could have meaning. Since a judgement like 'wealth is evil' is not related to any body of sensory experience in this way, it would appear to be without meaning. Grasping the bull firmly by the horns, Ayer boldly declared that this was indeed the case; value judgements have a function, for they express the feelings of the speaker and they arouse similar feelings in the hearer, tending to stir him into action, but they have no meaning. It follows that value judgements cannot be justified or falsified, any more than frowns, smiles and sighs. If I make one judgement and another makes a judgement which appears to contradict mine, the contradiction is illusory:

> For in saying that a certain type of action is right or wrong, I am not making any factual statement, not even a statement about my own state of mind. I am merely expressing certain moral sentiments. And the man who is ostensibly contradicting me is merely expressing his moral sentiments. So that there is plainly no sense in asking which of us is in the right. [1]

continued

The emotive theory of value, as this view came to be known, was greatly advanced by C. L. Stevenson, although he severed the theory's dependence on logical positivism and based it, instead, on an analysis of evaluative language.[2] Despite these changes, Stevenson maintained that the principle kind of meaning possessed by evaluative terms was not the descriptive which interested the logical positivists, but emotive meaning.[3] Such language has the function of affecting the attitudes and actions of the hearer or reader. Stevenson suggests two patterns of analysis which succeed in capturing this function of evaluative language. Leaving aside refinements, the first pattern reads such anowals as 'this is good' as 'I approve of this; do so as well'. The second pattern allows evaluative terms to have descriptive meaning as well as their primary emotive meaning. A phrase such as 'this is good' is given the meaning 'this has descriptive properties X, Y, Z' where 'good' still retains its emotive force. In seeking to persuade another to favour objects with properties X, Y and Z, the definition is adopted, and the use of the word 'good' conveys and encourages the favouring of such objects, Stevenson calls such definitions 'persuasive definitions'. Both patterns of analysis may be used to understand any real argument over evaluation, the choice between them resting on grounds of simplicity and convenience. In this way Stevenson claimed to develop a theory which could account for all features of evaluative discourse.

We need not consider the details of Stevenson's theory, since strictly it is beyond the terms of reference of the present chapter, maintaining as it does, that evaluative judgements cannot be justified. It was,

continued.....

though, a theory of great influence, and added impetus to the linguistic direction which ethics was taking, some of whose fruits we shall come across below.[4]

2. Justification is Possible Even Though Regresses of Reasons are Infinite

We saw in the previous chapter that all the difficulties we encounter when seeking to justify a particular claim, be it of any sort, arise from the need to terminate the regress of reasons which any attempt to justify the claim generates. If B is proposed as a reason for accepting A, then we may immediately ask for reasons for accepting B, and then for reasons for accepting these reasons, and so on. I argued that unless this regress terminates somewhere, no justification for A can be found. I also argued that it cannot terminate, so that no justification for A exists. Many writers on ethics have, however, questioned the first of these two claims. They have held that value judgements can be justified, even though the regresses of reasons which their justification leads to do not terminate.

A popular refrain of such writers is the triviality of the problem of the regress. Singer, for instance, tells us that the demand for a justification of a set of value claims all together;

> .....is contradictory, if not perverse, and the way in which it is put would effectively preclude any possibility of satisfying it. For to demand this is to demand a moral reason for accepting anything as a moral reason, and this is a self-contradictory demand.[5]

continued.....

Of course, once we are allowed to accept a value judgement even though we know of no justification for it, the problems of the previous chapter simply evaporate. Any regress of reasons generated by trying to justify a particular value claim can be halted anywhere, the value judgement at this point being accepted in spite of there being no justification for it. The problem with this response is that it opens the floodgates. If it is legitimate to adopt a value claim, despite any justification for it, on the grounds that 'we cannot justify all our values', then reason has no role to play in the making of value judgements. Suppose Singer's advice is followed, and C is adopted even though it lacks any justification. Let C entail evaluations B and A. Can we say that B and A are justified? B and A are justified only if there is some reason for accepting them, and this reason is supposed to be provided by C. But, of course, there can be no more reason for accepting B and A, if we confine ourselves to the examples, than there is for accepting C. If there are reasons for accepting C, there are reasons for accepting B and A; if there are no reasons for accepting C, there are no reasons for accepting B and A. C has, however, been chosen arbitrarily, in the complete absence of reasons. Hence, C fails to justify B and A. C justifies B and A only if there are reasons for accepting C, i.e. only if C is justified. In other words, we must justify all our value claims together, or not at all.

A similar, but much more sophisticated version of Singer's argument is proposed by Becker.[6] Becker appreciates the significance of the sceptic's use of the infinite regress argument and feels dissatisfied with the standard rebuttals offered by more optimistic writers on ethics.

continued.....

His own answer is to shift the burden of proof to the sceptic; 'Once the sceptic, rather than the moralist, is under interrogation, the possibility of a rational basis for morality does not seem so remote.'. Becker therefore proposes several 'presumptive value criteria', i.e. criteria for sorting values for which there is no proof or justification from other criteria, but which may be presumed correct in the absence of reasons to the contrary. For such criteria, the task of the sceptic is to find such reasons; if he can find none, then our acceptance of the criteria is justified. What enables us to treat the criteria in this special way is the fact that they are 'features of our lives which emerge without our having chosen them, and which do so in every normally-formed human being.....'. [8] Becker produces three presumptive value criteria, purposiveness, personalness and the aesthetic nature of life, but to see how they operate we need only consider the first of these.

Human beings who are normally formed are purposive, i.e. they seek satisfaction or fulfilment of purposes. They also try to satisfy as many purposes as possible, and where a choice between purposes is unavoidable, they prefer those which are pervasive through time and productive of greater satisfaction. Becker then argues;

> that the direction and priorities operative from the given purposiveness of our lives can reasonably be allowed to function as a value criterion as long as no reason to the contrary can be justified..... When there is no reasoned objection to an existing set of priorities, there is no

continued.....

further 'need' for justification of those priorities; and if those priorities were not themselves chosen, but simply happened, there is nothing 'arbitrary' about them. [9]

There are two reasons for thinking Becker's account inadequate.

(i) Becker's account runs foul of the sceptical arguments discussed in the previous chapter. Becker relies upon the assertion:

> If X is a value criterion built in to all normal human beings and if no argument against X is known, then X is justified. (1)

The sceptic can be just as troublesome as before. For instance, the sceptic may ask what is the justification for (1) ? As far as I can see, Becker offers no justification for (1), but whatever justification is offered, a regress of reasons will be generated. As I tried to show in the previous chapter, attempting to terminate this regress at some fundamental reasons is futile. Moreover, repeating Becker's move and proposing some presumptive principle, one which we may accept even though it is unjustified, in the absence of counter-arguments, will similarly fail. Such a move will depend upon some criterion for a principle being presumptive, and asking for justification of this criterion will set the regress rolling again. Hence, (1) can never be

continued.....

justified. This is, of course, an application of the first sceptical argument.

Even if we allow the justification of (1), it can be used to justify adopting some value criterion only if this criterion can be shown to be built in to all normal beings. Thus, for example, Becker's adoption of the criterion of purposiveness depends upon;

> Purposiveness is a value claim built in to all normal human beings. (2)

The second sceptical argument operates on (2). Becker tries to justify (2) by appealing to the way in which children are reared, but whatever justification is offered a regress of reasons will follow. We can, for instance, ask how we know how children are reared and so on. This regress will not terminate at some fundamental reason, as I have tried to show earlier, nor will appeal to a presumptive principle be of any help, as I have explained above. In other words (2) is not justifiable. Thus, even if (1) were justified, we would not be able to employ it to justify the adoption of a value criterion.

(ii) A judgement of any kind may only be said to be justified if there is some reason for thinking it correct, by whatever standards of correctness are applicable to such judgements. Becker's claim is that there are value judgements for which there are no reasons, but which are,

continued

nevertheless, justified because no countervailing reasons are known. My objection is that where there are no reasons supporting a value judgement, we cannot regard the judgement as justified. Becker's account confuses two things which it is most important to distinguish - justifying a value judgement and justifying the adoption of a value judgement. First of all, let us be quite clear that these are two distinct issues. To avoid controversy, consider Popper's account of scientific method to be discussed in chapter 4. According to Popper, no scientific theory can be justified, so that the only way to assess a theory is to criticise it by subjecting it to experimental tests which may falsify it. For this reason, theories which are easily tested are to be preferred to those which are difficult to test. In this way, a scientist may justify his adoption of a highly testable theory rather than a theory of low testability. The theory he adopts is not justified, nor can it ever be justified, but his adoption of the theory can be justified by the need to expose theories to experimental test.

The same may hold for value judgements. As Becker concedes to the sceptic, no reasons can be given for value judgements, since the regress of reasons generated has no terminus. Hence no value judgement can be justified. If we wish, nevertheless, to assess value judgements by holding them open to criticism, we must adopt some judgements as, at

least provisionally, agreed upon. If all judgements are open to question at the same time, reason abandons the scene. All value judgements may be questionable, since none are justified, though not all can be questioned at once. If some value judgements are to be provisionally accepted, what more reasonable to adopt than those for which there is no known counterargument and which are 'built in' to every normal person. Such judgements will not be justified, although their adoption may be justified by the need to have some (provisionally) agreed set of value judgements to enable the criticism of other judgements to proceed.

This view seems to rescue something of Becker's account and it will be considered in much greater detail in chapter 6. A little more must be said here, however, about the distinction between a value judgement being justified and the adoption of a value judgement being justified. Consider a scientific theory for which there is no justification, but whose adoption is justifed because it is highly testable. If the theory entails a sentence, this sentence is not shown thereby to be justified. There is no more reason to think the sentence true than to think the theory true, which is none. At most, the adoption of the sentence might be justified. The same considerations apply to value judgements. The fact that the adoption of a value judgement is justified lends no credence whatsoever to consequences of the judgement. This is only possible when the value judgement itself is justified.

A second point to be noted is the difference in attitude which is appropriate towards a justified claim and a claim which is not justified

continued.....

but whose adoption is justified. If a claim is not justified, then its adoption must be provisional and we must always be prepared to abandon the claim should some counterargument be found in the future. The only rational attitude to such a claim, even though its adoption is justified, is a critical one. It is not enough to admit that it will be rejected once a counterargument is produced - we must conscientiously seek for counterarguments. The claim must be held tentatively, and always be open to rejection, and it must always be remembered that any consequences drawn from the claim are open to the same strictures. If the claims in question are value judgements, then the first question of the philosophy of value, or ethics, is how value judgements can be criticised. Involved in this, as we shall see, are finding ways of outlawing manœuvres which prevent us from recognizing counterarguments, and encouraging the development of rival ethical views, for without such rivalry there can be no real criticism. On all these points Becker is silent. He seems quite unaware of the radical changes in our attitudes towards values which his concession to scepticism entails.

A similar blindness vitiates the work of Singer, who has been mentioned before and whose position will be more fully discussed below. Here it is enough to point out that he sees what he calls the 'generalization principle' as central to ethics. This states that 'what is right (or wrong) for one person must be right (or wrong) for any similar person in similar circumstances'. Singer admits that no evaluative principle can be known to be true, since all justification is relative to some assumed values. Nevertheless, he claims to have provided a defence of the generalization principle;

> .....what is relevant here is not so much a justification in the sense of a demonstration, or a deduction from self-evident premises, as a defence. This involves the elimination of misunderstanding, the meeting of difficulties, and the answering of objection, and this is the procedure I have in fact followed.

Despite this confession, Singer seems to regard his defence as tantamount to a justification. Thus, reviewing his discussion of the principle, he states that he has tried:

> .....to clarify it, to show how it can be applied, to meet various objections to it, and to show how it is involved in all moral reasoning. This being so, I cannot think that there is anything else that is needed. I do not think it impossible that I have failed to do what I attempted, but if I have not, then any lingering doubts about the validity of this principle would be unreasonable. [11]

I must admit to a great deal of sympathy for Singer here; at least he is aware of the limitations of justifying philosophical theories. We must, however, be very strict about the gulf between justification and defence. There is no way at all in which a defence of the kind described by Singer can amount to justification. If all objections are

continued.....

successfully answered, this says nothing against the possibility of a new and damning objection being discovered tomorrow or the day after. The only rational attitude towards an unjustified principle such as Singer's is to treat it as open to the maximum criticism possible and to hold it tentatively, not dogmatically, attitudes which are quite unnecessary to adopt towards principles which we know to be justified. Moreover, we should search for criticisms with as much diligence as we can muster, and it must always be remembered that since the principle is not justified, then it confers no justification upon its consequences, towards which the same attitude is, therefore, appropriate. Like Becker, Singer sees nothing of this.

Yet a third example of this blindness is provided by Rawls. Rawls attempts to construct an elaborate argument, in the social contract tradition, for principles of justice. What is of interest here is not so much the detail of his arguments as his remarks on the justification of his principles. Like Becker and Singer he admits the impossibility of fundamental value judgements and to some extent recognises the sceptical consequences of this impossibility. Justification is not a matter of deduction from self-evident first principles;

> ..... instead (it) ....is a matter of the mutual support of many considerations, of everything fitting together into one coherent view.[12]

> Justification rests upon the entire conception and how it fits in with and organizes our considered judgements in reflective equilibrium.[13]

continued.....

By 'reflective equilibrium', Rawls means the Socratic element in our evaluations.

> ...we may want to change our present considered judgements once their regulative principles are brought to light. And we may want to do this even though these principles are a perfect fit. A knowledge of these principles may suggest further reflections that lead us to revise our judgements. [14]

When there is no motivation to change a judgement in this way, it is in reflective equilibrium.The task of philosophical ethics is to subsume judgements at equilibrium under some set of organizing principles. Once it is admitted that value judgements, even those under reflective equilibrium, are not justified, they can confer no justification upon theories which subsume them. If Rawls' principles of justice can subsume many value judgements under equilibrium, as he claims they can, this provides us, therefore, with no reason for thinking the principles true. Further developments may always call them into question.

Simiarly, we have no reason to think something true because it is a consequence of Rawls' principles. The principles are not justified, but it may be that we are justified in adopting them - assuming that they are, after all, our best guess about how to subsume our equilibrium value judgements. But if we are so justified, then the principles must be held open to criticism, and we must search for counterarguments as strenuously as possible. As we shall see, a

continued.....

consequence of this search for criticism is that we should favour theories with novel, unexpected consequences so that a theory like Rawls', which accommodates only what we already know, is of no interest to us.

3. Evaluative Foundations for Value Judgements

Perhaps the most popular response to the sceptical arguments developed earlier is to insist that a value judgement can be justified because the regress of reason involved terminates at some value judgement which is a fundamental reason. On this view, all value judgements are founded upon some set of value judgements which are somehow justified without appeal to any further value judgements. The aim of the present section is to show that there are no fundamental value judgements, and, therefore, that this response to the sceptic is a failure. For convenience, I shall break responses of this kind into three mutually exclusive and exhaustive groups. If there are such things as fundamental value judgements, then there are just three possibilities; either (a) they are self-evident, or (b) they are arbitrary, or, on the middle view,(c) they are neither self-evident nor arbitrary, but they can be argued for rationally. For each group I will first show how the sceptical arguments of the previous chapter apply, and then bring forward independent grounds for thinking then inadequate.

(a) Fundamental Value Judgements are Self-Evident

This has been a very popular response to the sceptic. Ross, for example, tells us that as we progress along a regress of reasons;

> The ultimate propositions at which we arrive seem not to express mere brute facts, but facts which are self-evidently necessary ..... [15]

For instance by a self-evident necessity, we have 'the general prima-facie duty of producing the maximum good'.[16] According to Prall, 'all value is intuited ...', value judgements giving 'an ex post facto account of these intuited values'.[17] In a similar vein, Garnett states that value is a sensum like colour; 'As immediately given (values) have that final reality which belongs to everything else with which we have direct acquaintance'.[18] For him, moral value is 'immediately felt, as a quality, in the experience of conscience.'[19] Ross' fellow intuitionist, Moore, held that all propositions about the value of objects;

> ....rest in the end upon some proposition which must be simply accepted or rejected, which cannot be logically deduced from any other proposition. This result..... may be otherwise expressed by saying that the fundamental principles of Ethics must be self-evident ..... The expression 'self-evident' means properly that the proposition so called is evident or true, by itself alone; that it is not an inference from some proposition other than itself.[20]

An equally blunt statement is made by Prichard, also of the intuitionist school.

> This apprehension (that something is our duty) is immediate, in precisely the same sense in which a mathematical apprehension is immediate .....Both apprehensions are immediate in the sense that in both, insight into the nature of the subject directly leads us to recognise its possession of the predicate; and it is only stating this fact from the other side to say that in both cases the fact apprehended is self-evident. [21]

C. A. Campbell held that all moral theories had to find a place for intuition. Individual acts could be held obligatory because they are intuited to be so, or else because their obligatoriness follows from some moral principle itself known intuitively. In either case, the evaluation of the act rests fairly and squarely upon intuition. [22]

I first want to show how the doctrine of self-evident, intuitively known value judgements falls foul of the sceptical arguments of the first chapter. A central claim of all the theories mentioned here is one of the form;

> All value judgements recognised by intuition under conditions C are correct ...(3)

For example, Ross holds that condition C is that the agent has 'reached sufficient mental maturity, and ... given sufficient attention to the proposition'. [23] For Moore, great care must be taken

continued.....

to isolate the object whose value is to be determined by intuition, and condition C will include a clause that the object is properly isolated.[24] But whatever version of (3) is proposed, it may properly be asked how it is known that the proferred version is true. Whatever the details about the nature of the supposed intuition, we may follow Kemp and ask why it is that this intuition can be relied upon in the stated circumstances.[25] As I argued in the previous chapter, asking such a question will lead to a regress of reasons which will not terminate. Whatever reasons Ross, Moore and their friends bring forward in favour of their particular version of (3), we may properly ask for reasons for accepting these reasons, and so on ad infinitum. This is, of course, an example of the first sceptical argument.

Putting these difficulties aside for a moment, let it be granted that some principle of form (3) is known to be true. Before it can be used to generate value judgements, some sentence of the following kind must be justified.

> Value judgement J is recognised by intuition under conditions C ....(4)

The second sceptical argument now presses down on the intuitionist. As I tried to explain earlier, asking for reasons for accepting some sentence of the form (4) leads to a regress of reasons which does not terminate. It follows that the sentence cannot be justified. It cannot, therefore, be conjoined with some principle of form (3) to justify a particular value judgement.

continued.....

We may now turn to additional arguments against the doctrine of a self-evident moral base which are independent of the highly general arguments of the previous chapter. There are two such arguments. The first begins by observing that whether or not a value judgement is intuited under the right conditions is a psychological matter. This being so, there is no way of arguing rationally when there is disagreement about these values. If one person claims that a principle is self-evident to him, whilst a second claims the same about some contrary principle, there is no possibility of rational argument. But if this is so, why should the first person rely on his own intuitions rather than those of the second? He has, after all, no reason, let alone any guarantee, that his opinions on the matter are more likely to be correct than those of another. As Ayer points out;

> ...it is notorious that what seems intuitively certain to one person may seem doubtful, or even false, to another. So that unless it is possible to provide some criterion by which one may decide between conflicting intuitions, a mere appeal to intuition is worthless as a test of a proposition's validity.[26]

The second difficulty concerns the doctrine's inability to provide reasons for acting. This is best seen by considering the example of Moore's version of intuitionism. Moore holds that goodness is an

indefinable, and hence simple, quality such that intuition can justify the acceptance of singular statements about the intrinsic goodness of objects.[27] What his theory seems unable to answer is the question, given that this is what goodness is, why should we act in order to produce good, rather than bad, consequences? This may sound a peculiar question, until Moore's special views about goodness are remembered. Why should we, for example, favour an object which has this property over one which has not? If Moore's theory is adopted, therefore, we sever the link between values and action. This alone seems damning enough, but it also means that Moore's views are no longer of interest to us, since as stated at the beginning of the chapter, our first concern is the question 'how shall we act?'.

So far, I have discussed the attempt to find an unquestionable evaluative basis for value judgements only with reference to intuitionist theories of value. In recent years such theories have become much less popular, but the search for the unquestionable base survives, although the logic of moral language replaces the earlier feelings of certitude. A particularly bold variant of this approach is offered by Warnock, who applies the so-called 'paradigm case argument' to establish the existence of evaluative knowledge. Typically, the paradigm case argument points out that some word is taught by the identification of some object; thus, for example, we are taught the meaning of 'white' by being shown snow, so that it cannot be denied that the word correctly applies to the object. According to Warnock;

> ....if the phrase 'morally wrong' is not absolutely meaningless; if it is possible to say, in elucidation of what it means, what sorts of things rank semantically

continued....

> as morally wrong; then there are some things, such as (doing just as you please)....., from which that appellation could not be withheld by anyone not unaware of the meaning of the expression, or not deliberately misusing it. [28]

Unfortunately, this manoeuvre runs into all the trouble encountered by the more simple-minded intuitionists. The obvious questions are whether the paradigm case argument is ever valid, and then whether it is valid in this particular application. [29] Even if a positive answer is given to both questions, this can only be because a certain view of natural language and its learning has been adopted, and this view will be as incapable of justification as the intuitionists' theories of value. Hence, Warnock cannot use the paradigm case argument to justify any value judgement, since the validity of the argument cannot be justified. We shall encounter the same problem later when discussing Singer's generalization argument.

A final problem is that Warnock tells us nothing as to why we should act morally. If knowing that it is morally wrong to beat up old ladies is on a par with knowing that snow is white, how can the moral wrongness of hitting old women be any more relevant to doing such an act than the whiteness of snow is to whether we ought to throw snowballs?

A different approach is attempted by other contemporary writers on ethics. They urge that the whole edifice of our evaluative discourse crumbles if some value judgements are denied, so that using this language presupposes these judgements, which are, therefore, beyond

continued.....

issue. According to Peters, for example;

> If it could be shown that certain principles are necessary for a form of discourse to have meaning, to be applied or to have point, then this would be a very strong argument for the justification of the principles in question. They would show what anyone must be committed to who uses it seriously. Of course, it would be open for anyone to say that he is not so committed because he does not use this form of discourse or because he will give it up now that he realizes its presuppositions. ..... But (this) .... would be a very difficult position to adopt in relation to moral discourse. [30]

Rather than weary the reader, I will leave it to him to apply exactly the same considerations made against Warnock to Peters' position.

Yet a third variety of the same type is provided by those philosophers who use the supposed absurdity of the question 'why by moral?' to justify some privileged value judgement which does not then stand in need of support from any further value judgement. A rather cunning example of this sort is provided by Nielsen. He wishes to base all moral evaluation upon the principle of least suffering (J), so that no moral proof of J is possible, since any moral proof at all presupposes J. Nielsen then considers someone who accepts this, but asks why he should accept the moral proofs based upon J. This is an impossible question

continued.....

(really a pseudo-question), because the reasons stemming from J 'are just the reasons that are to count as 'good reasons' in this context'. [31]

Unfortunately all the by now familiar arguments apply against Nielsen. If J and what follows from it are to be justified in the way he indicates, then Nielsen's assertions about the special nature of J must themselves be justified, and so we find the regress continuing. Also, we loose the connection between morality and action. If 'morally good reasons' is a term which applies to anything which follows from J, why should we act upon these reasons?

(b) Fundamental Valuations are Arbitrary

Medlin observes that since an evaluative conclusion demands an evaluative premise;

> ...we can go back, indefinitely but not forever. Sooner or later, we must come to at least one ethical premise which is not deduced but baldly asserted. Here we must be arational; neither rational nor irrational, for here there is no room for reason even to go wrong. [32]

Werkmeister tells us that;

> In any actual situation of (value) conflict we have but to ask: What is our highest commitment? And what obligations does it entail? Given a clear conception

continued.....

> of this commitment, we can determine the harmony and integrative consistency of all entailed commitments and, therefore, of our obligations and of the rights we bestow on others. [33]

For Hare, 'ought' sentences;

> .... can only be verified by reference to a .... set of principles which we have by our own decision accepted and made our own. [34]

The idea here is that commitment to some arbitrary value principle enables all value claims which follow from the principle to be regarded as justified. Once the commitment has been made, nothing is allowed to falsify it and so it cannot stand in peril of being overthrown, so that it can be regarded as fundamental; justified though not justified by any other value claim. This commitment is, of course, supposed to be, by its very nature, free. Nothing can prevent a person making what commitments he choses. Thus for Werkmeister, 'our ultimate commitment ..... must be freely made and must be personal'. [35] According to Parker,

> Each person's heirachy of values is a personal decision (at the heart of which) is a 'basic egocentricity' which entails an equally 'basic relativity of values'. [36]

continued.....

In a similar vein, Williams states that;

> The whole pursuit of ethics consists of using all available evidence ..... first, to examine into this fundamental commitment (i.e. to find out 'what I really want'), and thereafter to deduce what acts ..... are best conformable to it. [37]

The sceptical arguments of the first chapter apply to this view in exactly the same way as before. According to the present view, a value judgement J is justified when it has been shown to follow from an arbitrarily chosen ultimate value judgement U. Any such justification will, therefore, depend upon claims of the form (5) and (6) below.

> If value judgement J is entailed by a arbitrarily chosen value judgement U, then J is justified. ...(5)

> J is entailed by an arbitrarily chosen value judgement U. ...(6)

According to the first sceptical argument, any attempt to justify (5) will lead to a non-terminating regress of reasons so that (5) will be incapable of justification. It comes as no surprise, therefore, when we fail to find any kind of justification for a principle of form (5) in the authors quoted above. Mostly, there is bald assertion, more

rarely a totally inadequate argument. But even if some principle of form (5) is justified, it can be used to justify particular value judgements only when coupled with a sentence like (6) which is also justified. The second sceptical argument tells us that any attempt to justify (6) will lead to an infinite regress of reasons, so that (6) cannot be justified.

There, are, moreover, independent arguments against the view that justification ends with arbitrary commitment. Firstly, the whole business of 'commitment' becomes very murky once we inquire into it. How, for example, is a commitment to an arbitrarily chosen principle to be made; how do we know when we have made such a commitment - does it require writing down or will a gentlemen's agreement suffice; can a commitment be retracted and, if so, how; can a commitment or a retraction be made inadvertantly, and so on?

More importantly, it is hard to see how a commitment to an arbitrarily chosen principle can have the authority which it is supposed to have. An individual is supposed to accept those value judgements which are entailed by the principle he is committed to, and reject value judgements which contradict it. The ultimate principle is supposed to provide a reason for the adoption of value judgements which it entails and the rejection of ones which it contradicts. It is hard to see, however, how ultimate principles can have this authority if commitment to them is arbitrary. Imagine a man who feels inclines to make a value judgement J which contradicts an ultimate evaluative principle he has committed himself to, U. Why should he reject J? Why not, since

continued.....

commitment is arbitrary, change to a new ultimate principle consistent with J? Someone consistently acting in this way will do exactly as he pleases, but will always be glad to justify what he does by appeal to some 'ultimate principle', even though it be freshly invented and destined for retraction as soon as it becomes inconvenient. There is no reason for thinking such a person any less reasonable than one who dogmatically retains some ultimate principle no matter what consequences it is found to have, since the choice of a principle is arbitrary. It would seem that the belief that ultimate principles are arbitrary contradicts the belief that ultimate principles have authority.

(c) Ultimate Evaluative Principles May be Argued For

Many moral philosophers, especially those favouring objectivist views of value, have offered arguments for the ultimate evaluative principles which they have supported. Mill, for example, (at least on one interpretation) argues from the fact that people desire happiness to his ultimate utilitarian principles;

> ....if human nature is so constituted as to desire nothing which is not either a part of happiness or a means of happiness - we can have no other proof, and we require no other, that these are the only things desirable. If so, happiness ..... must be the criterion of morality.[38]

Mill then goes on to argue that happiness is, indeed, the only end of human action.

continued....

There is something odd in arguing, as Mill does, for some ultimate principle, for the principle is supposed to be the terminus of all argument. On closer inspection, these suspicions are confirmed. The premise of any argument in favour of an ultimate principle must itself be justified, and so we are back in a regress. The whole point of ultimate principles is to terminate such regresses, but if ultimate principles need to be supported by argument, these regresses are without end. As might have been expected, if there are ultimate evaluative principles, they cannot be supported by argument.

Consider, for example, Mill's argument. It depends upon two claims:

> If human nature desires only happiness, then only happiness is desirable ....(7)
>
> Human nature desires only happiness ....(8)

Trying to justify these claims generates a regress of reasons in each case which does not terminate. Thus, we can ask what reasons there are for accepting (7) and (8), and what reasons there are for believing these reasons and so on indefinitely. This means that neither (7) nor (8) can be justified, so that Mill's claim that only happiness is desirable cannot be justified and cannot be a fundamental value judgement.

To conclude this section; it seems that the view that value judgements may be justified by some ultimate evaluative principles, which are themselves justified independently of other value judgements, must fail. These supposed ultimate principles must be self-evident, arbitrary or

continued.....

else argued for, but I have tried to show that none of these is adequate.

4. Non-Evaluative Foundations for Value Judgements

The above section has considered those value theories which attempt to show how value judgements may be justified in terms of some fundamental reason which is also a value judgement. A variant of this approach still sees the justification of a value judgement as requiring some fundamental reason, a reason, that is, which does not require support from a further reason, but the fundamental reason chosen is not itself a value judgement. The most developed theory of this kind has been proposed by C.I. Lewis.

For Lewis, basic to all evaluation is the direct experience of goodness or badness - 'felt goodness' and'felt badness' as he calls it.

> Without such direct value - apprehensions, there could be no determination of values, or of what is valuable .... Without the experience of felt value and disvalue, evaluations in general would have no meaning.[39]

Sentences which describe a felt-value experience are fundamental reasons. They do not need evidence from further reasons. Such a sentence is

continued

> .....self-verifying (for him who makes it) in the only sense in which it could be called verifiable, and subject to no possible error, unless merely linguistic error in the words chosen to express it. [40]

This is not to say that Lewis is committed to a subjectivist view of value. He is adamant that objects and states of affairs may possess value, and that this value is an objective property of the thing in question. Felt-value, of course, 'has the status of the apparent' and is subjective, but:

> There is no apprehension of the empirical whatever except by apprehension of appearances. ..... there is no basic difference between value-apprehensions .... and apprehensions of any other empirical character. [41]

The parallel between immediately apprehended value as compared with value as a property of an object, and seen redness or straightness as compared with the objective property of redness or straightness in a thing, is so obvious as hardly to call for extended comment. [42]

How, then, does Lewis suppose judgements about the objective value of a thing may be tested by the experiencing of felt-value and felt-disvalue? Lewis distinguishes between two kinds of value judgements - terminating judgements and ascriptive judgements. A terminating value

judgement holds that a certain value-quality will be experienced if a particular state of affairs pertains. For example, 'if I touch this red-glowing metal, I shall feel pain' is a terminating value claim. Such judgements are, according to Lewis, decisively and completely verifiable or falsifiable. If I touch the metal of our example above, and feel pain, then the sentence of the example is verified. If no pain is felt, it is falsified. Ascriptive evaluations, on the other hand, are those which ascribe an objective value property to a thing. According to Lewis, these judgements

> .... are not, at any given time, decisively and completely verified, but always retain a significance for further possible experience and are capable of further confirmation ... Any particular confirmation of such a judgement comes by way of finding true some terminating judgement which is a consequence of it.[43]

Thus:

> ... the end by relation to which alone anything is ultimately to be judged genuinely valuable, is some possible realization of goodness in direct experience.[44]

A somewhat similar view to that of Lewis is held by proponents of the so-called 'good reasons' school in ethics. For these writers,

continued.....

justification of a value judgement ultimately rests upon facts because facts provide good reasons for such judgements. As an example of this approach, we may briefly consider the work of Baier. [45]

According to Baier, the means-end model of action and evaluation has been an unfortunate one for ethics. Instead, he proposes that the best course of action is the one which is supported by the best reasons. This immediately raises the question of when something is a reason for action. To arrive at an answer to this question in a particular instance, deliberation is required. This is a double operation, consisting of a survey of all the related facts and a subsequent weighting of those facts found to be directly relevant to the decision. Thus facts provide reasons for action, and the regress of reasons generated by trying to justify a moral judgement can terminate at factual sentences. Facts alone are, of course, inadequate. They must be supplemented by what Baier calls 'consideration making beliefs'. The discovery of a reason supporting a course of action A consists of conjoining a fact F with a consideration making belief of the kind 'F is a reason for doing A'. Reasons of this sort do not finally justify the performance of A, but in the absence of countervailing reasons, we may assume that A is the correct course of action. Where reasons exist for and against doing A, rules of superiority must be applied to see which reason counts most.

A rule of superiority ranks reasons of various kinds, e.g. one might rank moral reasons above aesthetic reasons and another political above prudential reasons, and so on. Of particular interest for Baier are moral reasons, or reasons required by or acceptable to the 'moral point of view'. According to this point of view; the rightness of an act is dependent upon its performance as a duty, principles are meant to

continued

apply to everybody and to be for the good of all. Whether or not a principle meets this last condition is to be determined by someone who adopts the moral point of view- an independent, unbiased, impartial, objective, dispassionate, disinterested observer no less. Moral principles are superior to all others because:

> ... being moral is following rules designed to overrule self-interest whenever it is in the interest of everyone alike that everyone should set aside his interest. ..... The best possible life for <u>everyone</u> is possible only by everyone's following the rules of morality.[46]

As before, I shall now consider the adequacy of Lewis' view of the way in which value judgements may be justified, first looking at the force of the sceptical arguments of the previous chapter, and then at shortcomings which are independent of these very general arguments. As to the first sceptical argument; Lewis claims that all sentences reporting a felt-value experience are self verifying, but how is this claim justified? A sentence of this sort is supposed to terminate the regress of reasons generated when we attempt to justify a value claim, but asking this question restarts the regress all over again. Looking for justification that such sentences really are self-verifying, means that they cannot be employed to terminate any regress of reasons. If it is admitted, contrary to all these difficulties, that sentences reporting felt-values are justified, then if we can identify a sentence as a report of a value-experience, then we know that it is justified. Here, of course, we run foul of the second sceptical argument. Justifying the claim that a particular sentence is a report of a value-

continued.....

experience inevitably leads to a regress of reasons, so that the sentence cannot be used to terminate a regress of reasons.

We may now briefly mention the chief independent failing of Lewis' view. It is, as before, that it cannot explain the connection between evaluation and action. Lewis recognises the connection - 'the immediately good is what you like and what you want in the way of experience; the immediately bad is what you dislike, and do not want' -[47] but ultimately, he cannot explain the connection. If, as he claims, an object is good to the extent to which it produces the sensation of felt-goodness, why should we favour this particular sensation over that of felt-badness? Maybe we do have this preference, and maybe men have always preferred felt-goodness to felt-badness, but even so, the connection, on Lewis' theory, can only be an accidental, contingent one. This, however, is not enough.[48] I leave the reader to extend these criticisms of Lewis to the 'good reasons' school of philosophers.

5. Value Judgements May be Justified by a 'Way of Life'

A popular view in contemporary value theory is that value judgements can only ultimately be justified by reference to 'a way of life'. Feigl, for example, suggests that any evaluative argument takes place in the context of certain accepted norms and standards, and that these can 'validate' a particular judgement. In searching for a justification of such standards and norms we are trying to 'vindicate' them, and vindication needs to appeal to a way of life.

continued.....

> The purposes which may be adduced in vindicating arguments for a whole system of moral norms are embodied in the individual interests and social ideals which we have come to form in response to life experience. The principle of justice ..., may for example, be vindicated by reference to the ideal of a peaceful, harmonious and co-operative society. [49]

Hare also recognises the need for a regress of reasons generated by attempting to justify a value judgement to terminate. Ultimately, we must come to a way of life of which that particular value judgement is part. In practice, Hare tells us, it is impossible to specify such a way of life fully, but if we could, then we can give no further justification. If a questioner insists upon justification:

> We can only ask him to make up his own mind which way he ought to live; for in the end everything rests upon such a decision of principle. He has to decide whether to accept that way of life or not; if he accepts it, then we can proceed to justify the decisions that are based upon it- if he does not accept it, then let him accept some other, and try to live by it. [50]

A somewhat baroque theory of the same kind is proposed by Taylor. [51] An individual value judgement he supposes to be <u>verified</u> by being derived from a more powerful evaluative principle. Such a principle may than be <u>validated</u> by being exhibited as a consequence of a higher-

continued....

order principle. This may proceed until the highest order evaluative principle is reached. Here validation is impossible, but the principle may, nevertheless, be justified by appeal to a particular 'point of view'. A point of view, such as the political, aesthetic, moral and pragmatic, is a statement of what kinds of consideration are relevant to the adoption of a highest-order evaluative principle. Within each point of view, however, there are a number of competing 'value systems' or heirarchically arranged sets of principles. A value system may be vindicated by appeal to a 'way of life', consisting of a set of value systems of different points of view arranged in order of relative precedence. If a value system has worth relative to a chosen way of life, then it is vindicated by this way of life. In turn, a way of life may be justified by being shown to be rational. This leads us to Taylor's proffered explication of the concept of rational choice.

Taylor's list of necessary and sufficient conditions for rational choice are rather complex. They fall into three groups. Putting them briefly, they are; [52]

1. Conditions of freedom
   (a) The choice is not decisively determined by unconscious motives.
   (b) The choice is not at all determined by internal constraint.
   (c) The choice is not at all determined by external constraint.
   (d) The choice is decisively determined by the person's own preference.

continued.....

2. Conditions of enlightenment.

   (a) The nature of each way of life is fully known.

   (b) The probable effects of living each way of life are fully known.

   (c) The means necessary to bring about each way of life are fully known.

3. Conditions of Impartiality.

   (a) The choice is disinterested.

   (b) The choice is detached or objective.

   (c) The choice is unbiased.

We may now turn to the adequacy of the view of justification presented by proponents of the 'way of life' approach to evaluation. Any such theory must hold some principle of the form:

> If S is a way of life with property P and if value judgement J is a consequence of S, then J is justified. 9

For Taylor, the property P is being rational, for Hare P is being freely chosen by an agent, and other versions are conceivable, but in every case appeal to a principle of this sort is supposed to terminate the regress of reasons generated by the attempt to justify J. It is easy to see, however, that this is a forlorn hope. J may only be justified by such a principle if the principle itself is justified, but trying to justify the principle, by, for example, developing

theories of value like Taylor's or Hare's, necessarily starts a regress of reasons. The regress of reasons produced by trying to justify J, therefore, cannot terminate at a principle of the above type.

The second sceptical argument also operates. If we assume that some principle of the above form is justified, before we can employ it to justify a particular value judgement J, we need to know that S really is a way of life with property P. But justifying this claim again continues the regress which we are trying to stop by appealling to the principle. The regress of reasons from J cannot terminate at a principle of the above form, even if such a principle were itself justified, as finding evidence that S is a way of life with P will necessarily continue the regress. This is particularly obvious in the case of Taylor's account, where conditions for S to be a rational way of life are so complex that they could never possibly be shown to hold.

A criticism of the 'way of life' approach which is independent of the sceptical arguments considered earlier is that it fails to account for the connection between action and evaluation. On Taylor's account, for example, we want to ask why should we prefer a way of life which is rational over one which is not rational? Taylor claims that:

> The decision to commit oneself to a way of life which is rationally chosen over other ways of life .... is the most reasonable, least arbitrary, and best founded of all. It is the decision to live the way of life one is most justified in living, all things considered.[53]

continued....

This seems reasonable enough, until we remember Taylor's curious definition of rational choice. Given this definition, he has actually said nothing as to why a way of life rationally chosen should be preferred above all others. Why should we, that is, prefer a way of life selected under the conditions he lists to one chosen under a quite different set of conditions?

Hare comes across exactly the same problem. For him, justification terminates at a commitment to a way of life. Once such a commitment has been made, reasons for evaluations can be given, but not before. Reasons cannot, therefore, be given for the commitment to a way of life. Such decisions are not, however, arbitrary.

> Far from being arbitrary, such a decision would be the most well founded of decisions, because it would be based upon a consideration of everything upon which it could possibly be founded. [54]

This, however, is so much whistling in the dark. That a particular way of life has certain consequences can, as Hare admits, be no justification for commitment to that way of life, nor can it provide so much as a reason for such commitment, since reasons and justification wait upon commitment. What right has Hare, therefore, to talk about basing a choice between ways of life upon consideration of the consequences of each way of life. At best, considering such consequences may lead to favouring one way of life over its rivals, but only as a matter of psychology. It is not permissible to think that the favou[illegible] way of life is somehow more justified that its rivals.

To be consistent, Hare must admit that the commitment to a way of life is arbitrary, in which case he runs into the difficulties discussed above in section 3 (a); that such a view does not allow any connection between evaluation and action since it cannot explain the authority of the initial commitment.

The view that evaluations may be ultimately based upon a way of life, therefore, runs into a dilemma. Either it is held that one way of life should be chosen over others because it possesses some particular feature or else the choice between competing ways of life is seen as arbitrary. In the first case, it must be shown why a way of life with the selected feature should guide action rather than any other way of life, and this cannot be done. In the second case, the gap between the chosen way of life and action is equally wide.

6. Value Judgements May be Justified by Special Arguments

A view stemming from the Kantian tradition in value theory is that at least some value judgements may be justified by the employment of a special kind of argument. The chief proponent of this approach is Singer, who has given an extended account of what he calls the 'generalization argument'.[55] This has the form 'if everyone were to do X, the consequences would be undesirable; therefore no one ought to do that'. The inference of the argument is mediated by the 'generalization principle'; 'what is right (or wrong) for one person must be right (or wrong) for any similar person in similar circumstances'. The similarity between two individuals and between two sets of circumstances is determined once a reason is given why an act is right

continued....

(or wrong) for a person in the stated circumstances. For example, if it is held that Smith ought to diet because he is fat and wheezy, then the generalization principle tells us that all people who are fat and wheezy ought to diet. Singer places various restrictions on what kind of reasons are possible, to avoid the sort of special pleading which threatens to make the generalization principle vacuous.

The obvious question, of course, is how the generalization argument can be shown to be valid, or, what amounts to the same thing, how the generalization principle can be known to be true. As we have seen, Singer admits that no evaluative principle can be finally justified, since all justification presupposes some assumed values, so that the generalization principle cannot be justifed. We may agree with him here. The regress of reasons generated by attempting to justify the generalization principle will be an infinite one, so that the attempt must fail. Thus the generalization principle cannot be justified, nor can it confer justification upon its consequences. Moreover, the same will be true of any kind of argument which is held to occupy a similar place in ethics. Whatever argument is proposed, it may be enquired whether it is valid; in other words, whether the principle it depends upon is true. Asking such a question will generate a regress of reasons which will not terminate. It will, therefore, be impossible to justify the use of the argument for the assessment of value judgements.

A second difficulty with Singer's account is that the connection between values and action is broken. If, for example, the generalization argument is used to convince an agent that what he

continued.....

proposes to do is wrong, why should the agent perceive this as a reason for not doing what he proposes? Why, in other words, should an agent be persuaded not to do something just because if everybody did it the consequences would be bad? According to Singer, the agent's act would itself be bad, but if this is part of what is meant by 'bad', why should an agent consider the badness of an action a reason for not doing it?

CHAPTER TWO - FOOTNOTES

1. A. Ayer, Language Truth and Logic, 2nd edn., Gollancz, 1964 pp.. 107-8.

2. C. Stevenson, Ethics and Language, Yale Univ. Press, 1944.

3. A term first used by C. Ogden and I. Richards, The Meaning of Meaning, Harcourt, 1923.

4. For brief histories see, M. Warnock, Ethics Since 1900, O.U.P., 1966; G. Warnock, Contemporary Moral Philosophy, Macmillan, 1967. See also D. Collingridge, 'The Failure of Language in Ethics', Journal of Value Inquiry, 9. 1975, pp. 81-94.

5. M. Singer, Generalization in Ethics, Eyre and Spottiswoode, 1963, p.335. For the same strategy applied to a different problem see A. Ayer, The Problem of Knowledge, Penguin, 1956 and P. Strawson, Introduction to Logical Theory, Methuen, 1952, Chapter 9. See also S. Toulmin, The Place of Reason in Ethics, C.U.P. 1961, pp. 191-195 and K. Hospers, Human Conduct, Hart Davis 1961, pp. 499-507.

6. L. Becker, On Justifying Moral Judgements, Routledge and Kegan Paul, 1973. Compare J. Rawls, 'Outline of a Decision Procedure for Ethics', Philosophical Review, 56, 1957, pp. 177-197, and N. Rescher, 'Reasoned Justification of Moral Judgements', Journal of Philosophy, 60, 1958, pp. 248-255.

7. op cit, p.64.

8. op cit, p.65

9. op. cit. P.67

10. op cit, pp. 67 and 73.

11. M. Singer, Generalization in Ethics, Eyre and Spottiswoode, 1963, p.336.

12. J. Rawls, A Theory of Justice, O.U.P., 1972, P.21

13. op cit, p.579. Rawls cites Quine as the originator of this view of justification - see W. Quine, Word and Object, M.I.T. Press, 1960.

14. op cit, p.49.

15. W. Ross, The Foundations of Ethics, O.U.P. 1939, p.320.

16. op cit, p.186.

17. D. Prall, 'Metaphysics and Value', University California Publications in Philosophy, 5, 1924, P.126.

continued.....

18. C. Garnett, Reality and Value, New Haven, 1937.

19. op cit, p.273

20. G. Moore, Principia Ethica, C.U.P., 1903, p.143.

21. H. Prichard, Moral Obligation, O.U.P., 1949.

22. C. A. Campbell, Moral Intuition and the Principle of Self-Realization, Annual Philosophical Lecture, Henriette Hertz Trust, 1948.

23. W. Ross, The Right and the Good, O.U.P., 1930, p. 29.

24. op cit, Chapter 1, section D.

25. J. Kemp, Reason, Action and Morality, Routledge and Kegan Paul, 1964, p.176.

26. A. Ayer, Language Truth and Logic, 2nd edition, Gollancz, 1964, p.106. See also S. Toulmin, 'Knowledge of Right and Wrong', Proceedings of the Aristotelian Society, 50, 1949/50, pp. 146-149 and A. Ewing, 'Subjectivism and Naturalism in Ethics', Journal of Philosophy, 1937, pp. 588-597.

27. op. cit.

28. G. Warnock, The Object of Morality, Methuen, 1971, p.125. See also K. Nielsen, 'The Good Reasons and Ontological Approaches to Justification of Morality', Philosophical Quarterly, 9, 1959, pp.116-130.

29. See J.W.N. Watkins 'Farewell to the Paradigm Case Argument', Analysis, and following articles, 18, 1958, 25-33.

30. R. Peters, Ethics and Education, George Allen and Unwin, 1966, p.115. See also R. Downie and E. Telfer, Respect for Persons, George Allen and Unwin, 1969, p.64 and appendix.

31. op. cit, p.125.

32. B. Medlin, 'Ultimate Principles and Ethical Egoism', Australasian Journal of Philosophy, 35, 1957, pp. 111-118.

33. Werkmeister, Man and his Values, Johnsen, 1972, p.162.

34. R. Hare, Freedom and Reason, O.U.P., 1964, pp.77-78.

35. op. cit, p.163.

36. D. Parker, The Philosophy of Value, Ann Arbor, 1957, p.209.

37. D. Williams, 'Ethics as Pure Postulate' in W. Sellars and J. Hospers, Readings in Ethical Theory, Appleton-Century-Crofts, 1952, p.667.

38. J. S. Mill, Utilitarianism, Chapter 4. See also J. Kemp, 'Foundations of Morality', Philosophical Quarterly, 7, 1957, pp.305-318.

continued.....

39. C.I. Lewis, The Analysis of Knowledge and Evaluation, Open Court, 1946, p.375. For similar views see H. Feigl, 'Validation and Vindication' in W. Sellars and J. Hospers, Readings in Ethical Theory, Appleton-Century-Crofts, 1952, pp. 667-680.

40. op cit. p.375

41. op cit. p.409

42. op cit. p.411

43. op cit. p.375-376

43. op.cit. p.375-376

44. op.cit. p.387

45. K. Baier, The Moral Point of View, Cornell Univ. Press, 1958. See also P.Foot, 'Moral Arguments', Mind, 67, 1958, pp.502-513, and 'Moral Beliefs', Proceedings of the Aristotelian Society, 59, 1958-59, pp.83-104; S. Toulmin, The Place of Reason in Ethics, C.U.P., 1961; W. Frankena, 'Recent Conceptions of Morality' in H. Castaneda and G. Nakhnikian, Morality and the Language of Conduct, Wayne State Univ. Press, 1965, and Ethics, Prentice Hall, 1963.

46. op cit. p. 164.

47. op cit. p.404

48. See chapter 7.

49. op cit. p.678

50. R. Hare, The Language of Morals, O.U.P., 1952, p.69. See also K. Baier, The Moral Point of View, Cornell Univ. Press, 1958.

51. P. Taylor, Normative Discourse, Prentice Hall, 1961.

52. op.cit. p.166.

53. op.cit. P.188.

54. op.cit. p.69.

55. M. Singer, Generalization in Ethics, Eyre and Spottiswoode, 1963.

CHAPTER 3 - FACTUAL ERROR

In this chapter I shall assume a result which will be fully considered in chapter 5, that value judgements can have factual consequences. Whilst there is some novelty in this thesis, my proof of it is completely straightforward and does not depend upon any particular analysis of value judgements, nor upon any doubtful semantics or technical logic. The validity of argument 1 below will hardly be questioned.

Socrates is a man

All men are good

Socrates is good 1

1 has the form F and V_1, hence V_2, where F is a factual sentence and V_1 and V_2 value sentences. Any argument of this form may be re-written to give V_1 and not-V_2, hence not-F. 1 then becomes

All men are good

not-Socrates is good

not-Socrates is a man 2

whose validity is transparent. 2 is an example where two value judgements have a factual sentence as a logical consequence, and any number of examples may be constructed having the same pattern. This

is, of course, in clear breach of Hume's Rule, but the force of the examples seems undeniable. In chapter 5 another way of constructing similar examples will be considered.

1 and 2 may be re-written again to give an argument of the form F, hence not (V_1 and not V_2). 1 and 2 give;

Socrates is a man

not-(All men are good and not-Socrates is good) 3

3 is an example of a factual sentence which has evaluative consequences. If the factual premise is true, then it is not possible to hold both the value judgements that all men are good and that it is not the case that Socrates is good.

The proof above exploits no semantic principles about terms like 'ought' and 'right' and 'good' nor does it use any but the simplest logic, but it is also neutral towards different views of value judgements. Whatever views are taken about the nature of such judgements, the examples above will still hold. Thus, even if value judgements are held to have no truth value and to be mere imperatives, it must be conceded that the same relationships as above apply to imperatives. Thus, for example; 4 and 5 are valid.

Salute all men

not - salute Socrates

not - Socrates is a man 4

Socrates is a man

not - (Salute all men and not - Salute Socrates) ...5...

An unavoidable conclusion from the above is that some value judgements may be factually false. Consider, for example, a man X who thinks himself good and who thinks that all Jews are bad. From these it follows that X is not a Jew, thus;

All Jews are bad

X is good

not - X is a Jew ...6...

What happens, then, when X learns that he is, after all, a Jew? If X is a Jew, then he cannot hold the two value judgements in 6. One or both must be rejected. The two value judgements are not, of course, inconsistent and had not X been Jewish there would have been no problem about accepting them both. What can we say but that X's values have been falsified by the fact that he is a Jew? It is undeniable that some value judgements have factual consequences. If the factual consequences entailed by a value judgement are false, then, by modus tollens, the value judgement itself is false.

In chapter 6, this result will be seen to be of crucial importance for the theory of value, but for the present we shall concentrate on how the possibility of factual falsification affects

the justification of value judgements. If a Popperian point of view is adopted, then no factual claims may be justified (see Chapter 4). It follows immediately that no set of value judgements with factual consequences can be justified. On Popper's view, such a set of value judgements is always open to the discovery that one of its factual consequences is false, a discovery which must falsify the set of value judgements.[1] It might be thought that this possible embarrassment might be avoided by ensuring that the set of value judgements adopted has no factual consequences. This is, however, possible only if a very weak set is adopted. For the sake of simplicity, consider only value judgements which ascribe some value to an object, and which we may write $Va = q$ (the value of object a is q as measured on some appropriate scale). Now;

$$\begin{array}{l} Va = q \\ \underline{Vb = q'} \\ \text{Not} - (a = b) \end{array} \qquad \ldots.7$$

For example;

$$\begin{array}{l} \text{The value of the fastest steam engine in the World} = 3 \\ \underline{\text{The value of The Mallard} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = 2} \\ \text{not} - (\text{The Mallard is the fastest steam engine in the World}) \end{array} \qquad \ldots.8$$

If the attribution of some numerical measure, as on a utility scale, is seen as suspicious, consider

$$\frac{Va > Vb}{\text{not } (a = b)} \qquad \ldots.9$$

For example;

The fastest steam engine in the World is more valuable than The Mallard

not - (The Mallard is the fastest steam engine in the World)

....10

Confining ourselves to singular value judgements of the form Va = q, it can be seen that the only way to avoid factual consequences is to adopt a set of the form, Va = q, Vb = q, Vc = q, Vd = q ..... . As soon as two objects are ascribed different values, a factual consequence emerges as I have shown.

If we now consider universal value sentences of the form (x) (Rx ⊃ Vx = q), a similar result appears.

$$\frac{(x)\,(Rx \supset Vx = q)}{(x)\,(R^1x \supset Vx = q^1)}$$

$$(x)\,(Rx \supset \text{not} - R^1x) \qquad \ldots.11$$

For example,

$$\frac{(x)\,(x \text{ is a Rolls Royce car} \supset Vx = 5)\qquad (x)\,(x \text{ costs more than £10,000} \supset Vx = 3)}{(x)\,(x \text{ is a Rolls Royce car} \supset \text{not - } x \text{ costs more than £10,000})}$$

....12

Again, if numerical values are found objectionable they may be easily eliminated. As before, factual consequences can be avoided only by adopting a set of value judgements of the form $(x)\,(Rx \supset Vx = q)$, $(x)\,(R^1x \supset Vx = q)$, $(x)\,(R^2 x \supset Vx = q)$ .... . As soon as two values appear in the set, factual consequences arise. Thus, if Popper's sceptical view about the possibility of justifying factual sentences is accepted, all sets of value judgements, except of the very weak kind above, will be beyond justification. All but the weak sets above have factual consequences, and if these consequences stand forever in peril of falsification, so too do the sets of value judgements which entail them.

Even if a Popperian position about factual sentences is rejected, and it is held that some such sentences can be justified, the justification of value judgements is still fraught with difficulties. The value judgements which an agent holds are never clear cut and closed, but always roughly formed, ill-articulated and open-ended.[2]

It is, therefore, very difficult to identify just what the factual consequences of the agent's values are. This means, in practice, and not in some philosophical text, that it is always possible for unexpected factual consequences to emerge from the agent's set of value judgements and, of course, some such consequences may well be discovered to be false. If this happens, then the agent's set of value judgements is also falsified. Thus in practice, whatever view is taken of the justification of factual sentences, an individual's set of value judgements is always open to unexpected falsification. It follows that his set of value judgements cannot be regarded as justified. This will perhaps become clearer in the discussion of various case studies in chapter 7.

We are now at the end of the first part of this work. What I have tried to show is that the traditional view that value judgements can be justified is mistaken. In chapter 1 I developed a highly general argument to this effect, and the argument was applied to various justificationist views of value in chapter 2. There, several attempts at showing value judgements to be justifiable were criticised both by the general argument of chapter 1 and by more particular arguments. The present chapter also provides an independent argument for the impossibility of justifying value judgements. These sceptical arguments call for an entirely new approach to ethics. The most depressing conclusion which we could draw from these arguments is total scepticism or nihilism, which holds that any value judgement is as good or as bad as any other and that reason cannot assess such

judgements. If this is to be avoided, then a fallibilist approach to value judgements is called for which admits the impossibility of justification; but maintains that some value judgements are rationally assessible by criticism. Whilst nihilism would accept both of the sceptical statements below, a fallibilist theory of value would accept the first but seek to deny the second.

1. No value judgement can be justified.
2. No reasons can be given for favouring one value judgement over another.

The development of such a theory of value is the task of part 2.

CHAPTER 3 - FOOTNOTES

1. Here I am using the usual convention that a set of statements is true if and only if all members of the set are true. It follows that a set of statements is false if and only if at least one member of the set is false.

2. This does not matter very much provided these judgements can be made precise when this is called for, and this is called for when the agent's value judgements are under critical attack (see chapter 6).

PART II

CRITICISM AND VALUE

Chapter 4 - Popper's Theory of Science

So far we have arrived at the sceptical conclusion that no value judgement can be justified. This is a disturbing result for it seems to indicate that any value judgement is as good as any other, and that reason can have no part to play in evaluation. This is, however, a mistaken conclusion as I hope to show in this part of the work. What I hope to show is that there is a way between the impossible optimism of justificationist accounts of value and the darkness of total scepticism, which relies upon the possibility of subjecting the value judgements we make to criticism. To repeat what was said at the close of the previous chapter, I hope to develop a theory of value which admits the first sceptical claim below, but denies the second.

1. No value judgement can be justified.
2. No reason can be given for favouring one value judgement over another.

In developing such an account of value, reference to the theory of scientific inquiry proposed by Karl Popper will be very rewarding, for the problem which Popper addresses bears a close analogy to our present problem. Popper is convinced that traditional views about the certainty of observation statements are wrong and that induction

cannot be shown to be a legitimate form of argument. It follows from this that no scientific claim can be justified. Popper does not, however, surrender to complete scepticism about the external world. Instead, he develops a theory of science which admits the first, but denies the second sceptical claim below:

1. No scientific claim can be justified.
2. No reasons can be given for preferring one scientific claim to another.

Popper's view is that scientific claims may be tested and compared by being submitted to criticism, or, in other words, by being exposed to falsification and rejection. The theory of value to be proposed in chapter 6 will follow him in this; maintaining that value judgements, although impossible to justify, may be assessed by exposure to criticism. The present chapter lays the foundation for chapter 6 by considering Popper's theory of science. What is offered is not, however, a bland description of Popper's views, but an account which highlights those features of his views which are particularly relevant to the development of an analogous theory of value. For this reason, deep and detailed discussion has generally been avoided, as have comments on Popper's numerous critics.

1. The Rationality of Popper's Methodology

Popper proposes a methodology for the testing and assessment of scientific claims which provides substance to the basic idea that such claims may be tested by being subjected to falsification and rejection. Before discussing the details of his methodology, however, the first question is why this methodology is appropriate to the assessment of scientific claims. What reasons are there, in other words, for adopting the methodological rules proposed by Popper rather than some rival set of rules?

Popper's first reply to this question is that the methodological rules he proposes be regarded as comprising a definition of empirical science, so that 'they might be described as the rules of the game of empirical science'.[1] The definition is not arbitrary, however:

> It is only from the consequences of my definition of empirical science, and from the methodological decisions which depend upon this definition, that the scientist will be able to see how far it conforms to his intuitive idea of the goal of his endeavours.
>
> The philosopher too will accept my definition as useful only if he can accept its consequences. We must satisfy him that these consequences enable us to detect inconsistencies and inadequacies in older theories of knowledge, and to trace these back

> to the fundamental assumptions and conventions from which they spring. But we must also satisfy him that our own proposals are not threatened by the same kind of difficulties. This method of detecting and resolving contradictions is applied also within science itself, but it is of particular importance in the theory of knowledge. It is by this method, if by any, that methodological conventions might be justified, and might prove their value.

The problem with Popper's original suggestion is that no link is forged between the proposed methodology and truth. The search for a scientific method is the search for ways of assessing our ideas about the world if we are interested in the truth, and any adequate methodology must connect with this interest in truth. Popper sought to frame his methodological proposals without reference to truth because at the time he regarded it and related concepts as thoroughly contaminated with bogus metaphysics. Tarski was, however, able to dispel Popper's suspicions about the propriety of these concepts with his semantic theory of truth.[2] Having appreciated the need to link his methodology with truth, Popper attempted to do this through his notion of verisimilitude. The verisimilitude of a scientific theory is a measure of the difference between its truth content (the set of true sentences it entails) and falsity content (the set of false sentences it entails). Popper sees

the aim of science as being the development of theories which get closer to the truth in the sense of having ever larger verisimilitude.

If T_2 has a greater logical and empirical content than T_1 (e.g. if T_2 entails T_1), then the truth content of T_2 is at least as great as the truth content of T_1. Hence, the verisimilitude of T_2 will be at least as great as that of T_1 unless the falsity content of T_2 is greater than the falsity content of T_1. The search for high verisimilitude, therefore, reduces to the investigation of T_2's falsity content, i.e. the search becomes an attempt to falsify T_2. This, Popper tells us, 'forms the logical basis of the method of science'[3].

There are two problems with this attempt to connect methodology and truth. The first is that the concept of verisimilitude is far more difficult to formalise than Popper originally suggested. Indeed, a whole cottage industry devoted to the concept's elucidation seems to have grown up and to be prospering.[4] Given the present state of confusion, it would seem unwise to base a whole methodology upon considerations of verisimilitude.

The second objection is even more fundamental. Even if the concept of verisimilitude is eventually clarified, the search for high verisimilitude cannot possibly constitute the 'logical basis of the method of science'. If science aims at high verisimilitude, then it may be appropriate to construct bold theories of high logical content and then attempt to

falsify them in order to assess their falsity content, but what is to count as falsification? A theory may always be saved from falsification by employing one of a whole number of conventionalist strategems (see below), or else the falsification may be taken as proper criticism of the theory. According to Popper, the latter is always needed. To assess a theory's falsity content, for Popper, involves searching for experimental evidence against the theory, and taking this, if found, as criticism of the theory, conventionalist strategems being prohibited. Indeed, these prohibitions form the backbone of Popper's methodology. But why, is assessing a theory's falsity content should such strategems be outlawed[5]? An answer is desperately called for, but simply appealing once again to the search for theories of high verisimilitude cannot settle the issue. Assessing verisimilitude involves probing a theory's falsity content, and, for Popper, this requires the prohibition of conventionalist strategems. Why conventionalist strategems are prohibited cannot, therefore, be explained by the need to assess the verisimilitude of theories. Verisimilitude cannot, therefore, provide the connection between truth and methodology which Popper so urgently requires.

A third suggestion about the reasons for adopting Popper's rules of method is due to Lakatos. Like Popper, he is aware of the need to connect methodology and truth and so sees Popper's methodology as requiring the inductive principle that its application produces theories of ever larger verisimilitude.

Popper claims to explain why one theory should be preferred to another, but[6];

> Preference is only a pragmatic concept within the context of this game [of science]. This preference can only assume epistemological significance with the help of an additional...., inductive ....principle which would somehow assert the superiority of science over pseudoscience. Such an inductive principle must be based on some sort of correlation between "degree of corroboration" and "degree of verisimilitude".

Popper has, however, exposed the fatuousness of this suggestion and no words of mine are required[7].

Before considering a positive suggestion about the reasons for adopting Popper's methodology of science, a final suggestion must be considered. According to Popper, the central problem for rationalism, the thesis that problems can be solved by critical discussion, is that no arguments can be given for it. Before an argument can be taken seriously, a rationalist attitude must already have been adopted. The choice between rationalism and its antithesis, irrationalism, is, therefore, seen by Popper as a moral one[8]. Although admiting that arguments cannot decide moral issues, Popper nevertheless thinks that the choice between rationalism and irrationalism

can be helped by argument, and he therefore presents arguments pointing to the beneficial consequences of the former. Since Popper's methodology of science is part of an overall rationalist scheme, the decision to adopt it also appears as a moral issue. How scientific claims are to be assessed, in other words, will depend on what moral commitments are made.

Popper's case is not, however, acceptable. To give arguments in favour of rationalism, as he does, be they determining or merely persuasive, presupposes that a commitment to rationalism has already been made[9]. Rationalism, therefore, can only be based upon an ethical commitment in the way suggested by Popper at the cost of denying the possibility of any kind of reason for this commitment, and so for rationalism itself[10]. If Popper's methodology of science is seen as part of a rationalist scheme, then since there can be no reason for adopting such a scheme, there can, in the final analysis, be no reasons for adopting Popper's methodology.

We may now turn to what I think is a happier way of connecting Popper's methodology with truth. In making this suggestion I am fully aware that Popper has not explicitly propounded it, nor has anyone else, and I have no wish to attribute implicit views to anyone. What I claim is that the following view connects Popper's methodology with truth in the way required, and that it fits naturally with the rest of Popper's thinking, whether or not he or anyone else would agree with the view. I wish to suggest that Popper's methodology

may be derived from two claims, both of which are made by Popper, although he does not seem to have explicitly noted the connection. These are that scientific statements have a truth value in a perfectly straightforward way, and that no scientific statement can be justified. The first claim may seem trivial at first until the long history of views of science which deny it, such as instrumentalism, phenomenalism and conventionalism with their many variants, is remembered. Against such views, Popper is a realist[11]:

> ... I am a realist in holding that the question whether our man-made theories are true or not depends upon the real facts; real facts which are, with very few exceptions, emphatically not man-made. Our man-made theories may clash with these real facts, and so, in our search for truth, we may have to adjust our theories or to give them up.

In support of his second claim, Popper has several arguments. All scientific claims must be theory laden because they must contain some universal term which cannot be correlated with any finite body of sensory data. The claim 'here is a glass of water', for example, contains the universals 'glass' and 'water', both of which denote bodies which behave in certain law-like ways[12]. In addition, any attempt to justify a particular scientific claim must lead to an endless proliferation of tests[13]. Finally, any

attempt to justify a scientific theory from observational reports founders on Hume's problem of induction[14].

If we are interested in the truth then we must attempt to resolve any contradiction between statements which we wish to hold, for the contradiction means that not all of the statements are true. Scientific statements have a truth value and so, if we are interested in the truth, contradictions between scientific statements need to be resolved. If a set of scientific statements is contradictory then at least one of them must eventually be abandoned. Since none of the statements is justified there can be no reason for regarding any one of them as sacrosanct. Any one of the statements may be abandoned; all are open to rejection. This is Popper's 'supreme rule of methodology' which states that all other rules of methodology are to be designed to ensure that no scientific claim is immune from criticism and rejection. This simple consideration also generates his criterion of demarcation which states that a statement is scientific only if it is possible to falsify it empirically[15]. To put it another way: any scientific statement, having a truth value and being beyond justification, may be false. It follows that any such statement must be open to criticism and rejection, an openness to be ensured by Popper's rules of methodology. If we seek truth, then our scientific conjectures must be submitted to assessment by a methodology which ensures that all can receive criticism and that any may finally be rejected. In Popper's own words[16]:

> I saw that what has to be given up is the quest for justification . ....All theories are hypotheses; all may be overthrown.
>
> On the other hand, I was very far from suggesting that we give up the search for truth: our critical discussions of theories are dominated by the idea of finding a true (and powerful) explanatory theory; and we do justify our preferences by an appeal to the idea of truth: truth plays the role of a regulative idea.

The connection between the methodology and truth is here extraordinarily simple, especially when compared with the complex nature of verisimilitude. If we are interested in the truth, contradictions between statements must be resolved, and since none of the statements is justified, it is possible to abandon any one of them. This, I suggest, for all its simplicity provides an adequate basis for Popper's methodology. We may speak of the two claims which generate this methodology, that scientific statements have a truth value and they cannot be justified, as constituting a metatheory. This simple metatheory is all that is required for Popper's methodology.

2. Test Sentences

According to Popper all claims within science are to be exposed to criticism and rejection. Principal among these

claims are scientific theories, expressed in universal sentences. It is important, therefore, to find a type of sentence which can be used to falsify universal sentences. Popper calls such sentences 'basic statements', but I prefer the term '(scientific) test sentences'. These test sentences must be of the form 'there is such and such a thing at place p and time t'. There is a further condition which test sentences must fulfill. Since their function is to test scientific theories there must be some way in which the acceptance of test sentences can be agreed upon. Without such agreement, which may, however, only be provisional, science would resemble the tower of Babel. Observation provides a way of achieving agreement about test sentences, as almost universal agreement can often be reached by different observers, for example, about sentences like 'this pen now on my desk is yellow', 'the duck now on the bridge is quacking' and so on. We can say, therefore, that scientific test sentences must be of the above form and must refer to some observable state of affairs.[17] Science has, however, no interest in stray, unreproducible effects, so that a scientific test sentence should be accepted only if it describes a reproducible effect.[18]

For Popper, test sentences are just as uncertain as the theories which they test. There must, therefore, be some way of testing test sentences by exposing them to criticism. This can be done in two ways. In the first, test sentences are tested in exactly the same way as theories are tested. If a test sentence falls into question, then

deductive consequences may be drawn from it and these consequences exposed to falsification by accepted test sentences. Suppose, for example, a piece of litmus paper is seen in bad light, so that there is some doubt about the test sentence 'there is blue litmus paper at p,t'. The test sentence entails 'if the litmus paper is observed in daylight then it will appear blue'. This sentence, and hence the original test sentence, will be falsified if observation leads to the acceptance of the test sentence 'there is daylight at p,t and the litmus paper appears pink at p,t'[19].

The second way in which a test sentence is criticizable is through falsification of the theory it depends upon. Once it is agreed that all sentences are theory-laden to some degree, we can accept a sentence such as 'there is a star at p,t' as a test sentence, because we have instruments for observing the heavens which lead to agreement about this sentence. Astronomers observing with their telescopes will generally reach agreement about a sentence such as 'there is a star at p,t', just as laymen, observing with their eyes, will reach agreement about 'there is a cat at p,t'. If, now, the theory of the telescope is falsified, then all grounds for accepting 'there is a star at p,t' may be destroyed.[20]

Popper's views of the nature of observation must be distinguished from traditional, justificationist views. These hold that some sentences referring to observation are justified by the observation,

without need for evidential support from further sentences, so that they provide a foundation for all knowledge. We have seen that Popper rejects this view and maintains that knowledge has no foundations. Experience can motivate an observer to accept a sentence, but in no way can it justify a sentence. Test sentences, for Popper, are special only in that agreement about them can be reached by observation. This agreement, however, is only provisional and tentative, for an accepted test sentence may at any time be faced with falsification. Thus[21]

> The empirical base of objective science has thus nothing 'absolute' about it. Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry the structure, at least for the time being.

It is usual to view the attempted falsification of a scientific theory T in the following way. T is conjoined with some set of initial conditions I and an auxiliary hypothesis A (though not every test will involve such a hypothesis), to yield the negation of some test sentence O. If observation leads to the acceptance of O, then T is falsified, whilst if not-O is accepted, T is corroborated. For example, let;

T = For all metallic conductors, the current flowing is directly proportional to the potential difference across the conductor.

A = In the circuit C, meter X reads the potential difference across the metallic conductor M and Y reads the current through M.

I = When the reading on X is 2 and 4 volts, the reading on Y is 1 and 2 amps respectively.

O = The reading on X is 6 volts and the reading on Y is not 3 amps.

T.A.I entails not-O. Hence, if observation of the two meters leads to the acceptance of O, T is regarded as falsified, whilst T is corroborated to some extent if observation leads to the acceptance of not-O.

The logic of such a test can be put in other ways, which we will find useful later on in chapter 6. For example;

T.A $\longrightarrow$ not (I.O)

Here, the theory plus auxiliary hypothesis forbids the conjunction of I and O, which can sometimes itself be regarded as a test sentence. In our example, this is the case, as observation can lead to a verdict about both I and O.

A third way of viewing any test is given by observing that since

$$T.A \longrightarrow \text{not } (I.O)$$

then, $$I.O \longrightarrow \text{not } (T.A)$$

T.A and I.O can, therefore, be seen as mutual counterarguments. This will be particularly useful later on.

3. Empirical Content

The requirement that scientific claims be submitted to criticism naturally leads to a criterion for good scientific theories. A good theory is one which is easily tested, i.e. one which may be falsified by many test sentences whose negation it entails. The class of test sentences contradicted by a theory, Popper calls the empirical content of the theory. A good theory, therefore, has a high empirical content. But such a theory also says more about the world than one of low empirical content. Good theories are, therefore, powerful and bold.[22]

The simplest way of comparing the empirical content of two theories is by the entailment relations between them. Let us denote the empirical content of a theory T by C(T). If theory T_2 entails theory T_1 then:

$$C(T_2) \geqslant C(T_1)$$

If T_1 also entails T_2;

$$C(T_1) = C(T_2)$$

If T_2 neither entails nor is entailed by T_1, then the empirical content of the two theories is not comparable by entailment relations. Another measure, using the so-called method of dimensions, may be possible for two such theories, but this has little value for our purposes.

It is easy to show that, once conventionalist strategems have been banned, (see section 5) empirical content follows logical content $C_L(\ )$, the set of sentences entailed by a theory.

$$C(T_2) \geqslant C(T_1) \text{ if } C_L(T_2) \geqslant C_L(T_1)$$

This suggests that the upper limit of empirical content, which we can assign the number 1, be ascribed to a contradiction, and that the lower limit, to which we may allot 0, be attributed to a tautology Logical content is the complement of probability. Whilst a contradiction and a tautology provide the upper and lower limit of logical content respectively, they respectively provide the lower and upper limit for probability. Thus, if p(T) is the (absolute logical) probability of T, we may measure empirical content by

$$C(T) = 1 - p(T)$$

There are two ways in which one theory T_2 may entail another T_1. It may be because T_2 has a greater universality than T_1, as in A and B below, or because T_2 is more precise than T_1, as in B and C. Hence, Popper's demand for theories of high empirical content leads to the demand for theories of high universality and precision, two features which scientists intuitively favour. Popper also claims

that theories of high empirical content are simpler than those of low content, so that his account can also explain the need which scientists often express for simple theories.[23]

A. All heavenly bodies move in circular orbits.

B. All planets move in circular orbits.

C. All planets move in elliptical orbits.

4. Corroboration

In evaluating a theory we need to know how it has stood up to the tests which have been made of it. It is not enough to know what and how many tests it has passed or failed, for we also need to know how severe these tests have been. A test is severe if we can expect the theory to fail it. A theory is, therefore, only severely testable if it has some novel, unexpected consequence, normally a prediction but occasionally a retrodiction, and the more unlikely the consequence, the more testable the theory. A theory is successful if it passes some test, and the more severe the test, the greater the success which accrues to the theory. When a theory passes a test Popper talks of it as being corroborated by the test, and a theory's degree of corroboration is a measure of how successful it has been in withstanding falsification. A theory which makes novel predictions can acquire a higher degree of corroboration than one which does not. This is perfectly reasonable, for the first theory has risked its neck whilst the second has risked virtually nothing. Compare, for example,

the predictions made by Einstein's theory of relativity with the predictions made by today's horoscope, which are so vague and imprecise that they can hardly help but come true. Obviously relativity should acquire a greater success in having its predictions confirmed than that acquired by astrology through the confirmation of its predictions.[24]

In talking in this way of success and corroboration, it must not be thought that some inductive principle has been smuggled in by Popper. Nothing can be further from the truth. For Popper degree of corroboration is;[25]

> ... nothing but a measure of the degree to which a hypothesis h has been tested, and of the degree to which it has stood up to tests. It must not be interpreted, therefore, as a degree of the rationality of our belief in the truth of h ... Rather, it is a measure of the rationality of accepting, tentatively, a problematic guess, knowing that it is a guess - but one that has undergone searching examinations.

Any critical discussion must assume a background of sentences which are accepted, though perhaps provisionally and undogmatically, as being true. Such sentences may be said to form 'background knowledge' and exist far from the centre of critical debate. They provide something against which more controversial claims may be tested. Examples might be, 'grass is green', the theories underlying

the use of ordinary measuring devices such as the galvonometer and manometer and 'wheat contains protein'. An item of background knowledge may at any time fall under suspicion during a debate, when it will have to leave the periphery for the debate's centre.

For Popper a test of a theory is severe if the theory can be expected to fail the test on the basis of our background knowledge. A theory can be tested severely only if it yields predictions which, on the basis of present background knowledge, must be regarded as highly unlikely. If a theory T is to be tested through its prediction of E, then the severity of the test for the theory given background B, S(E,B) is measured by;

$$S(E,B) = 1 - p(E,B)$$

The more unlikely E, given B, the more severe the test. This may be normalized to

$$S(E,B) = \frac{1-p(E,B)}{1+p(E,B)}$$

The requirement that T entails E may be weakened, noting that the severity of the test must fall as p(E,T.B). The more general formula, in its normalised form is;

$$S(E,T,B) = \frac{p(E,T.B) - p(E,B)}{p(E,T.B) + p(E,B)}$$

The explanatory power of T for E given B may be defined thus;

$$E(T,E,B) = S(E,T,B)$$

The degree of corroboration of a theory T by a test E is a measure of the success which T has in surviving the criticism offered by E. It will, of course, be relative to background knowledge. Popper lists a number of requirements for a satisfactory measure of degree of corroboration, and suggests a number of possible measures. What is more important than the selection of a particular measure, however, is the fact that all of Popper's requirements can be satisfied together. Perhaps the simplest measure is given by;[26]

$$Co(T,E,B) = \frac{p(E,T.B) - p(E,B)}{p(E,T.B) - p(T,E.B) + p(E,B)}$$

5. Protection from Falsification

Popper is aware that it is all too easy to protect some cherished theory from falsification and that 'any nice adaptation of conditions will make any hypothesis agree with the phenomena'.[27] A strategy for avoiding falsification, Popper calls a conventionalist (or immunizing) strategem. If our aim is to expose all scientific claims to falsification, according to the supreme rule of methodology, then these conventionalist strategems must be banned from science. These prohibitions form lower order rules of methodology. When a theory is protected from falsification, it no longer belongs to science, but to pseudo-science (or 'metaphysics' as Popper calls it). This is,

of course, Popper's famous criterion of demarcation.[28] This states that the defining characteristic of a scientific claim is the possibility of its clash with experience, or, more formally, with test sentences accepted through observation. If this clash is rendered impossible by conventionalist strategems, then science is deserted for pseudo-science.

To give just one example, in 1936 Eddington proposed a theory which entailed that the ratio of the masses of the proton and electron is 1847.6, which was, unfortunately, outside the experimentally determined range of 1836.56 ± 0.56. Ten years later Eddington amended his theory, incorporating a mysterious factor of $\left(\frac{137}{136}\right)^{5/6}$ in order to obtain a value for the mass ratio which was within the experimental range. His amended theory fits the experimental facts, but this cannot count as success for the theory. Success can only be achieved at the risk of failure, and in this case Eddington's theory has not been exposed to possible failure. The fact that the theory gives the correct answer for the mass ratio of the proton and electron in no way corroborates the theory. We can see this from the measure for degree of corroboration given in the previous section. Since the value of the mass ratio is already experimentally determined, it forms part of the background knowledge, B, so that the probability of the ratio having this value, given B, is 1. In this event, the degree of corroboration afforded to Eddington's theory is 0. Such ad hoc devices for getting theories to agree with known facts must,

therefore, be outlawed from science by suitable methodological rules.

6. When is One Theory Better than Another

We have already seen that preference should be given to bold theories because they are easily tested, but it would be more useful to have a criterion for one theory being better than another. Popper provides just such a criterion. It is not enough to say that theory T_2 is better than theory T_1 if T_2 has a higher empirical content than T_1, for we could then produce an endless sequence of better and better theories simply by conjoining them with some sentence so as to increase their empirical content. Nor is it enough to insist that T_2 has a greater empirical content than T_1 and that T_2 has some novel consequence not shared by T_1, through which it may be highly corroborated. This would allow us to make any theory better than another simply by conjoining it with some novel sentences such as 'all kettles freeze when heated on Tuesdays'. These difficulties can all be overcome if we insist that T_2 is better than T_1 only if T_2 has some novel consequence not shared with T_1 which is corroborated. The final criterion for T_2 being a better theory than T_1 is, therefore, threefold. Firstly, T_2 must ensure the redundancy of T_1 by explaining all that T_1 can explain. Secondly, T_2 must make some novel prediction(s) not made by T_1, and, thirdly, some of these novel predictions must be corroborated by experiment.[29] Hence, special relativity is a better theory than Newtonian mechanics. It explains all that the earlier theory can explain and has corroborated novel predictions not shared by Newton's theory, such as the gravitational bending of light. Newton's theory

was, in its turn, a better one than Kepler's theory of the planets, because it could explain all that Kepler's theory could explain and because it predicted many new phenomena whose existence was confirmed by experiment.[30]

The passage from Kepler's theory to Newton's and from there to Einstein's illustrates what Popper regards as a central feature of science, its growth. Growth is achieved, not by the assembly of ever more facts and observations, but by the replacement of old theories by new ones which are even more testable, and so even more bold. The falsification of a theory by an experiment is not a sufficient reason for the rejection of the theory. The theory should be rejected only when a better theory is ready to replace it.[31]

Chapter 4 - Footnotes

1. K. Popper, The Logic of Scientific Discovery, Hutchison, 1959, section 11.

2. K. Popper, Conjectures and Refutations, Routledge and Kegan Paul, 1969, 231-233.

3. P.Schillp(ed.), The Philosophy of Karl Popper, Open Court, 1974.

4. For the most recent discussions see the various contributions to Synthese , 38, 1978.

6. I.Lakatos, Popper on Demarcation and Induction, in P.Schillp(ed) The Philosophy of Karl Popper, 257.

5. See K.Popper, Conjectures and Refutations, 246.

7. K.Popper, Reply to My Critics, in P.Schillp(ed.), The Philosophy of Karl Popper, 999-1011.

8. K.Popper, The Open Society and Its Enemies, Routledge and Kegan Paul, 1945, chapter 24.

9. For this reason W.Bartley has proposed his theory of comprehensively critical rationalism in The Retreat to Commitment, Open Court, 1962.

10. See the discussion of ethical commitment in chapter 2.

11. K. Popper, Objective Knowledge, Oxford University Press, 1972, pp. 328-9. See also pp. 317-329 and 38-45 and Conjectures and Refutations, Routledge and Kegan Paul, 1969, pp. 114-119.

12. K. Popper, The Logic of Scientific Discovery, Hutchinson, 1959, section 25 and appendix* V. See also Conjectures and Refutation pp. 118-9 and 387-388.

13. Conjectures and Refutations, pp. 21-24.

14. Logic of Scientific Discovery, section 1 and appendix* i; Objective Knowledge, pp. 1-13 and 85-101.

15. Logic of Scientific Discovery, p. 54.and sect.6.

16. Objective Knowledge, pp. 29-30.

17. Logic of Scientific Discovery, sections 27-30; Conjectures and Refutations, pp. 365-8.

18. Logic of Scientific Discovery, pp. 86-88.

19. The conjunction of two test sentences is a test sentence.

20. I. Lakatos, 'Falsification and the Methodology of Scientific Research Programmes' in I. Lakatos and A. Musgrave, Criticism and the Growth of Knowledge, Cambridge University Press, 1970, pp. 127-131.

21. Logic of Scientific Discovery p. 11. See also section 8.

22. Logic of Scientific Discovery, Chapter 6.

23. Logic of Scientific Discovery, Chapter 7.

24. Logic of Scientific Discovery, sections 82 and 83; Conjectures and Refutations, pp. 33-39 and 220-221.

25. Logic of Scientific Discovery, p. 145.

26. Logic of Scientific Discovery, appendix *ix.

27. Joseph Black, quoted in Logic of Scientific Discovery, p. 82.

28. Logic of Scientific Discovery, section 6.

29. Conjectures and Refutations, pp. 240-248.

30. See Objective Knowledge, Chapter 5.

31. Logic of Scientific Discovery, section 85; Conjectures and Refutations, 240-248; Objective Knowledge p. 30.

CHAPTER 5 - FACT AND VALUE

It is our aim to construct a fallibilist theory for the assessment of value judgement following Popper's fallibilist theory of science. A problem central to this enterprise is to discover what kind of sentences may be used as test sentences for value judgements. Test sentences, it will be remembered must satisfy three conditions; they must be able to contradict the statements they test, there must be some way of achieving at least provisional agreement about their adoption or rejection, and they must be testable. What I hope to show in this chapter is that factual sentences can play the role of test sentences for value judgements. For factual sentences the last two conditions are so obviously satisfied that they require little comment. Whatever the fine points of the final account given of factual knowledge, it is clear that we are able to reach agreement at least on some factual sentences, and that we have ways of testing such sentences. This chapter will, therefore, concentrate on the first condition. I shall try, therefore, to show how factual sentences can entail the negations of value judgements. This could be done in a matter of a couple of lines, but the relationship between facts and values has been much discussed and what I have to say, even though I use no extravagant argument, will be of some novelty. I shall, therefore, try to place my account of the matter in perspective and I shall state my position in what is, perhaps, an overdefensive way.

In a famous passage in the Treatise, Hume tells us that no 'ought' sentence can be derived from premises which are purely factual.[1] It is usual to extend this rule to forbid the derivation of any evaluative sentence, 'X is right', 'X is obligatory' etc., as well as 'X ought to be done', from factual premises. I shall refer to this strong rule as Hume's Rule. This rule is of the greatest importance in ethics because it seems to guarantee the autonomy of values. Proponents of autonomy hold that an agent is free to adopt values in a way in which he is not free to adopt factual claims. A man who holds a particular scientific theory can be presented with objective evidence which he must recognise as counting for or against this theory; refusal of this recognition amounting to unreasonableness or irrationality. Factual claims must always stand in jeopardy from the discovery of counter-evidence. Once Hume's Rule is accepted, evaluations are insulated from fact. Whatever facts are discovered and whatever factual sentences seem acceptable, these have no consequences of an evaluative nature. Facts can never force us to a particular evaluation, nor prevent us from holding the values which we do hold. If Hume's Rule is correct, then in Hare's words:[2]

> .....it follows that we are free to form our own moral opinions in a much stronger sense than we are free to form our own opinions as to what the facts are.

Popper adopts the same positions when he argues that the duality of facts and standards means that:[3]

> Neither nature nor history can tell us what we ought to do. Facts, whether those of nature or those of history, cannot make the decision for us, they cannot determine the ends we are going to choose.

If Hume's Rule is correct, then it seems clear that facts and values are autonomous and that we enjoy a special freedom in our evaluations. The reverse is, however, not so clear. As we shall see, it may be possible for Hume's Rule to be incorrect and yet for the doctrine of autonomy to be true. The correctness of

Hume's Rule is a sufficient, but perhaps not a necessary condition for the autonomy of values.

The thesis of this chapter is:

(1) There are many arguments which clearly breach Hume's Rule

(2) Of these arguments, some are also clear counter-examples to the doctrine of the autonomy of values.

(3) Consideration of these arguments seems to leave us no option but to admit that evaluations may be factually mistaken.

In section 1, well known counter-instances to Hume's Rule will be reviewed, whilst section 2 considers whether these counter-instances also falsify the doctrine of autonomy. Section 3 suggests some new objections to Hume's Rule and to autonomy. The final section then considers what remains of the doctrine of autonomy.

(1) Against Hume's Rule

Here I shall briefly review proposed counter-instances to Hume's Rule which have appeared in the literature.

(a) Institutional Obligations

In 1964, John Searle reviewed the argument, proposed much earlier by Reid and Carritt[4], that from the factual report of a person's promise making, his obligation to keep the promise follows by logic alone, in breach of Hume's Rule.[5] Without going into the refinements of the argument, we can put it briefly thus:

(1) Jones uttered the words, 'I hereby promise to pay you, Smith, five dollars'.

(2) Jones promised to pay five dollars to Smith.

(3) Jones placed himself under an obligation to pay Smith five dollars.

(4) Jones is under an obligation to pay Smith five dollars.

(5) Jones ought to pay Smith five dollars.

Various additions are needed to ensure that the relationship between successive sentences is one of entailment. Conditions for effective promise making must be added, for instance, but the above gives the core of the argument. Searle sees the argument as a reflection of the 'institutional' nature of promise keeping. This

enables him to identify an indefinitely large class of counter-examples to Hume's Rule. Thus:

> ...'one ought not to steal' can be taken as saying that to recognise something as someone else's property necessarily involves recognising his right to dispose of it. This is a constitutive rule of the institution of private property.

The problem with Searle's derivation, and a similar argument proposed by Max Black,[6] is that it depends upon the meanings of the terms 'promise', 'obligation', 'ought' and so on. This is a severe weakness, since, for all the pints of ink dedicated to the subject, there is still no recognised and unquestionably valid way of determining such meanings. It follows that whether the relationship between successive sentences in Searle's argument is one of entailment or not is not a matter which can receive a definitive answer, at least in the present state of the art. His argument does not, therefore, present an impregnable counter-example to Hume's Rule.

(b) Necessary Evaluations

Another approach to criticising Hume's Rule is to show that certain evaluative doctrines which are held to be logically necessary enable the Rule to be breached. It is often maintained, for instance, that 'ought' imples 'can', i.e. that it can be said of a person that he ought to do something only if he is able to do it. If, however, the evaluative '_X_ ought to do _Y_' entails the factual '_X_ can do _Y_' then the factual '_X_ cannot do _Y_' entails the evaluative 'it is not the case that _X_ ought to do _Y_', in breach of Hume's Rule.[7] A similar argument is used by Sen against Hare's theory of ethics. Hare maintains Hume's Rule and also insists that two items with identical descriptive properties must receive the same valuation. Sen shows that these two claims are, however, contradictory, since '_X_ and _Y_ have exactly the same descriptive properties' is factual and yet entails, according to Hare, the evaluative '_X_ is exactly as good (or as bad) as _Y_'.[8]

This approach, like the earlier one, is clouded by the problem of meaning. The logical transformations involved are impeccable, but the argument depends upon the claim that some evaluative sentence is logically necessary. Once it is admitted that

we have no clear cut and unobjectionable way of deciding whether sentences such as 'X ought to do Y and X cannot do Y' are logically false, the above objections to Hume's Rule are open to considerable doubt.

(c) Prior's Arguments

Prior employs elementary logic against Hume's Rule. This has the advantage of being clear cut and precise, unlike the attempts we have so far considered. Prior produces a number of counter-examples to Hume's Rule which are logically impeccable.[9]

1. Tea drinking is common in England, therefore either tea drinking is common in England or all New Zealanders ought to be shot.

2. There is no man over 20 feet high, therefore there is no man over 20 feet high who is allowed to sit in an ordinary chair.

3. Undertakers are church officers, therefore undertakers ought to do whatever all-church-officers-ought-to-do.

4. One should always wear a coat on a rainy day but there's no need to wear a coat today, therefore it is not raining.

It should be observed that none of these examples depend in any way upon a special view of value or ethics; in particular, on any view which gives to evaluative claims a truth value. Geach has shown, for example, that even imperatives can be related to facts in the same way as Prior's 4. Thus: 'If the 12.55 weather forecast says it will be showery, cancel this afternoon's match' and 'don't cancel this afternoon's match' clearly entail the factual 'the 12.55 weather forecast did not say it will be showery.'[10]

Resting as they do on straightforward logical manipulation, Prior's examples give us the clearest and best counter to Hume's Rule. Given his examples, it can never again be maintained that no evaluative sentence is entailed by sentences which are purely factual.

(2) Against Autonomy

So far, we have identified a set of clear cut breaches of Hume's Rule proposed by Prior. Other counter-examples have been seen to be murkier. In this section, I shall consider whether the examples discussed above are also counter to the doctrine of the autonomy of evaluation.

Searle's argument arises from the institutional nature of promise keeping, but

an agent may be perfectly free to adopt the institution in question or to refrain from adopting it. If he does adopt it, then uttering 'I hereby promise...' will give rise to obligations, but only because of the agent's free adoption of the institution of promise keeping.[11] Even if Searle's argument is accepted, it is, therefore, unclear to what extent it casts doubt upon the autonomy of values.

If the ought-implies-can thesis is logically necessary, then facts may place restrictions upon the range of evaluations which we may make. The fact that I cannot fly, for example, prevents the adoption of the evaluation 'I ought to fly'. Nevertheless, we have seen the difficulties in deciding the status of the thesis, so that it does not provide an unobjectionable counter-example to autonomy.

This leaves us with Prior's counter-examples. The deductions these involve are all formal, so questions of meaning do not obscure the issue. Prior's examples, therefore, provide the most clearcut and unquestionable counter to Hume's Rule. Things are not so crystalline, however, when we consider the force the examples have against the doctrine of the autonomy of evaluation. In his first example, for instance, it is clear that any evaluative sentence would do in place of the one used by Prior. If F is a factual sentence and N is normative, then F entails F or N but also F or not-N. Similarly with his second example. If there is no man over 20 feet high, there is no man over twenty feet high who is allowed to sit in an ordinary chair, just as there is no man over twenty feet high who is not allowed to sit in an ordinary chair. Prior recognises this, calling the occurrences of the evaluative expressions in (1) and (2) 'contingently vacuous'. Prior's third example contains no contingently vacuous expressions, but, as he admits, it fails to inform anyone that they have some particular duty which they should perform.[12] As Shorter comments, 'the inference in question is quite useless for anyone who wants to decide what concrete action he ought to perform'. Thus the first three of Prior's examples cannot really be taken as showing that our freedom to evaluate in whatever way we choose is circumscribed by facts. These examples contradict Hume's Rule, but not the doctrine of autonomy. Just this is shown, however, by his final example, which deserves special attention.

Consider the argument 5:

You ought to wear a coat on a rainy day.
It is not the case that you ought to wear a coat today.
Today is not a rainy day. ...5

Here the evaluative premises entail a factual conclusion, as in Prior's example 4. Re-arranging 5 yields 6, where a factual sentence entails an evaluative conclusion.

Today is a rainy day
not (You ought to wear a coat on a rainy day and it is not the case that you ought to wear a coat today.) ...6

I hope now to show that examples like 6 provide genuine counter-examples, not only to Hume's Rule, but also to the doctrine of autonomy. Notice first that no expression in 6 is contingently vacuous, and that it contains no 'ought' within an 'ought'-clause as in Prior's 3. I interpret the force of 6 in the following way. The fact that it is raining today shows that the two evaluative sentences, 'you ought to wear a coat on a rainy day' and 'it is not the case that you ought to wear a coat today' cannot be held together.[13] Clearly, the two evaluative sentences are not contraries, since there are circumstances, for example sunny weather, when they could be both accepted. What are we to say of someone who holds the two evaluative sentences, despite the fact that it is raining today? Odd as it sounds, I see little option but to say that this person's evaluations are factually incorrect. He is not being logically inconsistent, but merely unaware that his evaluations have a consequence which is, as a matter of fact, false.

There is, of course, nothing startling in the logic of argument 6. In applying any general evaluative principle, one has to conjoin it with a factual sentence as in 7.

All men are wicked
Fred is a man
Fred is wicked ...7

The argument has the structure N_1 and F, therefore N_2 where F is a factual sentence and N_1 and N_2 value judgements. Any such argument can be re-arranged to give an argument of the same kind as 6 by writing; F, therefore not (N_1 and not N_2). Consider a more serious example. X accepts both the evaluations 'all Jews are bad

men' and 'X is a good man'. What happens when X learns that his true parentage has been kept from him, and that he is himself a Jew? Obviously he cannot retain his values. The newly revealed fact about his origins conflicts with his original evaluations. Since we cannot change facts to suit our values, a truism we all learn sooner or later, X has no choice but to revise his evaluation of Jews or of himself. His original position was certainly not inconsistent, but has merely been shown to be factually incorrect.

It must be observed that the above in no way rests upon any particular view of values, for instance, the view that evaluative sentences possess a truth value. Geach's example, mentioned earlier, shows that the same relation exists between value judgements and factual sentences even when the former are treated as pure imperatives, having, of course, no truth value. His example shows that facts can make imperatives 'inoperative' because the orders they contain run counter to fact.

(3) Autonomy and Consistency

Another way to establish the results of the previous section is to consider how the consistency conditions which are imposed upon any set of evaluations immediately restrict autonomy. Upholders of autonomy insist that a person is free to adopt whatever values he or she may chose, subject only to the condition that these values are logically consistent. What I wish to show in this section is the surprising result that whatever consistency conditions are imposed upon values, this alone is enough to breach Hume's Rule and restrict autonomy. I shall consider five eminently reasonable consistency conditions which have been proposed, illustrating in each case how Hume's Rule may be broken and how autonomy is restricted by their imposition.

(a) An inescapable consistency condition for any set of values is that it does not contain both the normative sentences N and not N. If a set of evaluative sentences does contain both sentences, then it is inconsistent and must be modified. Consider an agent Z who so hero-worships another agent, X, that he freely choses to adopt the evaluation 'all of X's evaluations are correct'. Z then slavishly follows whatever

value judgements X makes. Formally, we may say that Z adopts the value judgement

(x) (X places the value q on x ⊃ Value of x is q) 8

Now suppose that X is inconsistent in his evaluation of some object, a, because he ascibes to it two quite different values. The fact that X 's evaluation is inconsistent means that Z cannot adopt 8 without inconsistency. Formally,

X places the value γ on a → not-(x) (X places the value q on x ⊃ Value of
and X places the value γ' on a x is q) 9

The left of 9 is a factual statement about the value judgements made by X, whilst the right is the denial of Z's original value judgement. This is, therefore, a clear breach of Hume's Rule, but it is also counter to the doctrine of autonomy. Z's evaluation is factually wrong. Z may only adopt 8 if his hero is consistent in his own value judgements - and whether or not he is consistent is, of course, a factual matter. Facts about the hero's evaluations, therefore, place a restriction on what values others may adopt.

(b) If objects are to be ascribed values, an essential condition is that one object may possess only one level of value. We may think of one and the same object as possessing different kinds of value, aesthetic, moral, prudential, etc., but to ascribe two different levels of, say, aesthetic value to the object is to be inconsistent. This enables us to argue in the following way:

The Mallard is the fastest steam locomotive in the World. → not-(The Mallard is more valuable than the fastest steam locomotive in the World.) 10

This, again, breaches Hume Rule, but it also falsifies the doctrine of autonomy. Imagine someone who thinks that the Mallard is more valuable than the fastest steam locomotive in the World. If we can demonstrate to him that the Mallard is the fastest steam locomotive, we thereby show his value judgement to be mistaken. It is not inconsistent, but we can point to facts which reveal it to be erroneous. It seems fair to say that his value judgement is factually

mistaken. Facts, once again, place restrictions on the value judgements which we may make.

(c) The relationship 'better than' or 'preferable', written '...pref ---', is normally held to be transitive. If this consistency condition is imposed upon values, we may argue:

Strawberries are the only fruit with seeds on the outside. $\Rightarrow$ not-(Eating strawberries pref eating grapes pref eating fruit with seeds on the outside) 11

Once again, we have a clear breach both of autonomy and Hume's Rule.

(d) According to Hare, to attribute value to an object is to commend it in virtue of some descriptive property which it possesses. It follows that if two objects, A and B, have all their descriptive properties in common, written A I B, then they must have the same value. Attributing different values involves inconsistency. This seems perfectly reasonable, but Sen has pointed out that this involved a breach of Hume's rule.[14] A I B is a descriptive sentence, and yet it entails the evaluative sentence 'A and B have the same value'. This is also a breach of autonomy, for the fact that A and B have the same descriptive properties forbids us placing different values upon them.

(e) A key concept in the theory of utility is the lottery, a chance of winning prizes $A_1 \ldots\ldots\ldots A_n$ with probability $p_1 \ldots\ldots\ldots p_n$. The standard axiomatisation of utility entails that the value (or utility) of a lottery is equal to the sum of the values of the prizes weighted by their probability.[15] Writing $V(A_i)$ for the value of the ith prize:

$$\text{Value of lottery} = \sum_{1}^{n} V(A_i)p_i \qquad \ldots\ldots.12$$

If the axioms are regarded as forming a definition of the concept of utility, in the usual way, then attributing any other value to the lottery leads to contradiction. This enables Hume's rule to be broken. For simplicity, consider L, the next lottery to be played, which happens to have two prizes, A_1 and A_2, with probabilities p_1 and p_2. This contingent description of L is

enough to determine its value and to rule out all others. In other words:

$\underline{L}$ has prizes $\underline{A}_1$ and $\underline{A}_2$ with probability p_1 and p_2 $\longrightarrow$ not-(The value of $\underline{A}_1$ is $V(\underline{A}_1)$ and the value of $\underline{A}_2$ is $V(\underline{A}_2)$ and not - $V(\underline{L})$ = $V(\underline{A}_1)p_1 + V(\underline{A}_2)p_2$) 13

The left of 13 is a contingent, factual sentence, whilst the right is evaluative, since it specifies what value $\underline{L}$ must have in relation to the value of $\underline{A}_1$ and $\underline{A}_2$. Contingent facts about the next lottery to be played determine relationships between values. This places a restriction on our freedom to evaluate. If values are attributed to the prizes, then facts about the lottery determine the lottery's value, and we are not free to attribute any other value to it.

These examples have an important feature in common. The imposition of a constraint on a set of evaluations in the form of a consistency condition enables the derivation of an entailment of the form $\underline{F} \longrightarrow$ not-$\underline{N}$, thus breaking Hume's Rule. In addition, each example can be interpreted as a restriction upon the autonomy of evaluation. In each case, the adoption of a set of values is constrained by facts. For proof of my thesis that <u>any</u> consistency condition leads to a breach of autonomy, I would only observe that an example of the kind (a) can be constructed for whatever consistency condition is imposed.

(4) Autonomy - Some Rump Hypotheses

I have presented a number of instances where some factual sentence, F, entails the negative of some conjunction of value sentences, not-(N_1 and N_2), all of which are in clear breach of Hume's Rule and all of which run counter to the doctrine of the autonomy of values. It seems undeniable, in the face of these examples, that facts may restrict our freedom to evaluate. Nevertheless, weaker versions of the autonomy hypothesis may still be true. In particular, I wish to suggest two such weak claims.

First of all, I have given no example where a factual sentence entails the negation of a single value, and I doubt if any cases will be found. Secondly, all my examples show how the negation of an evaluative sentence may be derived from a factual sentence. In all cases, facts serve to limit what evaluations may be made, but nowhere is it shown that a factual sentence can entail a non-negative evaluative sentence. It seems a reasonable conjecture that this is impossible. I cannot argue for this conjecture, except to say that I know of no counter-example to it. If there really is no counter-example, then the evaluations which we make are bounded, but not determined by, facts.

1) Treatise III, 1.1. The interpretation of Hume's views on the relationship between fact and value has generated a considerable literature. For some of the main arguments here, see W Hudson, The Is-Ought Question, Macmillan, 1969.

2) R Hare, Freedom and Reason, Oxford, 1963, p.2.

3) K Popper, The Open Society and Its Enemies, vol. 2, Routledge, 1966, p.278.

See also Vol. 2, addendum 1, Sect. 13, and Vol. 1, p. 73.

4) For a history, see A Prior, Logic and Basis of Ethics, Oxford, 1949.

Prior would add Hobbes to the argument's early proponents.

5) J Searle, 'How to Derive "Ought" from "Is" ', Philosophical Review, 73, 1964, pp 43-58 and Speech Acts, Cambridge, 1969, pp. 175-198. For objections see Hudson op. cit.

6) M Black, 'The Gap Between "Is" and Should" ', Philosophical Review, 73, 1964, pp 165-181. For objections see Hudson op. cit.

7) J Brown, "Moral Theory and the Ought-Can Principle', Mind, 84, 1977, pp. 206-223; G Mavrodes 'Is and Ought', Analysis, 25, 1964, pp. 42-44 and K Tranoy, ' "Ought Implies "Can" ', Ratio, 14, 1972, pp. 115-130 and Ratio, 17,1975, pp 147-175. For a possible escape route see Atkinson, 'The Autonomy of Morals', Analysis, 18, 1958, pp. 57-62, N. Cooper, 'Presuppositions of Moral Judgements', Mind, 75, 1966, pp.46-54, and P Shaw, 'Ought and Can", Analysis, 25, 1964, pp196-197, who use Strawson's notion of presupposition. But see D Collingridge, ' "Ought" Implies "Can" and Hume's Rule', Philosophy, 52, 1977, pp 348-351.

8) A Sen 'Hume's Rule and Hare's Rule', Philosophy, 41, 1966, pp.75-79.
See also Collingridge op. cit.

9) A Prior, 'The Autonomy of Ethics', Australasian Journal of Philosophy, 38, 1960, pp. 199-206.

10) P Geach, 'Imperatives and Deontic Logic', Analysis, 18, 1958, pp.49-56.
See also H Castaneda, 'Imperatives and Deontic Logic', Analysis, 20, 1960, pp.42-48. Again, this contradicts Hare. See his op. cit. p.28.

11) R Hare in Hudson op. cit.

12) J Shorter, 'Prof Prior on the Autonomy of Ethics', Australasian Journal of Philosophy, 39, 1961, pp286-287.

13) If anyone wants to object that 'it is not the case that X ought to do Y' is not evaluative, see Collingridge op. cit.

14) See Footnote 8.

15) J Von Neumann and O Morgenstern, Theory of Games and Economic Behaviour, Wiley, 1943, Chapter 1.3.

CHAPTER 6 - A FALLIBILIST THEORY OF VALUE

The ground being prepared, it is now possible to construct a fallibilist theory of value. In chapter 1 it was observed that most theories of value have attempted to find a way in which evaluations of some kind may be justified, and the whole of part 1 was aimed at exposing the futility of such attempts. If part 1 is correct, then no evaluations can be justified and traditional justificationist theories of value are impossible. In their place, I wish to propose a fallibilist theory of value which admits that justification is impossible, but maintains that some value judgements are rationally assessible by criticism. Such a theory, it will be remembered from the close of chapter 3, admits the first of the two sceptical claims below, but denies the second.

1. No value judgement can be justified.
2. No reasons can be given for favouring one value judgement over another.

The theory I shall develop here will be closely modelled on Popper's account of science which was discussed in chapter 4. In their attempt to show how scientific claims can be justified, justificationist philosophers of science have endowed science with special features such as inductive logic and protocol sentences or sense data reports. The great power of Popper's view is that science does not need such special apparatus; instead it employs the critical method - the method of conjecture and criticism - and

this method underlies all our attempts to apply reason to the problems which we have. On Popper's view scientific method can be seen as a particularly precise application of critical method, and one which can, therefore, serve as a model for the development of methodologies in other areas, such as ethics.[1]

In developing a theory of value, an immediate problem is the selection of a suitable starting point. Should we begin with what people _like_, what they _ought_ to do, what their _duties_ and _obligations_ are, or what? As my starting place I shall take _preference_, for two reasons. An agent's views of what ought to be done or what duties and obligations exist depend upon the values he places on the actions which are open to him, i.e. upon his preference for one action over others. It might prove fruitful to begin with the value ascribed to individual items rather than the comparison of values implicit in judgements of preference, but the logical apparatus is simpler for preference. The second reason for developing the theory in terms of preference is that it does not confine us to considerations of particular kinds of value, such as aesthetic or moral value, as developing the theory in terms of, say, duty might. The theory of value I hope to arrive at is, therefore, a perfectly general one. If there are varieties of value, then the ascription of each of them is to be assessed in the manner stipulated by the general theory of value to be developed here. The choice of a starting place is not,

however, crucial as I hope to show later that a fallibilist theory can be constructed for 'ought', 'obligation', 'duty', etc. in more or less the same way as for 'preference'.

1. Metatheory

The first step in the construction of a fallibilist theory of value is the statement of a metatheory upon which a methodology for the assessment of preference judgements may be based. Popper's metatheory, discussed in chapter 4, consists of two claims; that scientific statements have a truth value in a perfectly straightforward way and that no scientific statement can be justified. These two claims entail Popper's supreme rule of methodology, that all scientific claims are to be kept open to criticism. The metatheory for our theory of value consists of essentially the same two claims:

1. A statement of preference (X pref Y) is a sentence having a truth value.
2. No statement of preference can be justified.

The first claim may appear trivial until it is put into context. Many philosophers have wished to deny 1. Logical positivists, for example, thought that 'X pref Y' could not be a sentence since it has no empirical consequences (see chapter 2). At best, the utterance of 'X pref Y' might serve to evince certain feelings. Claim 1 is a denial of all such views and cannot, therefore, be

accused of triviality. It will be remembered that the analogous claim of Popper's seems trifling until its many historical denials are brought to mind.

Claim 2 has been defended in part 1. The view of preference proposed here is undoubtedly counter to the 'common-sense' view. According to 'common sense', an agent has privileged access to his own preferences - if he says that he prefers a to b, then he does prefer a to b (at least if we cannot show him to be lying, to be linguistically confused etc.), just as when he says that his big toe hurts, then his big toe hurts. This is a plausible view, although a mistaken one, when a and b are taken to be such things as colours, tastes, smells, sounds etc. It is not outrageous to say that if a man prefers celery to cabbage, then he prefers celery to cabbage and this is the end of the matter. To argue with him is to forget that he knows about his preferences in much the same way as he knows about his pains. Error about his own preferences is impossible. But even common sense admits that mistakes are possible in more complex situations. In chosing between motor cars, for example, a person may very well be mistaken about his own preferences. If he lists say, ten cars in order of preference and then makes a pair-wise comparison between the cars, we can perform an interesting experiment. An order of preference for the ten cars may be calculated from his list of pair-wise preferences[2], and it generally happens that this conflicts with his original ordering. Since the person's

original set of preferences is thus shown to be inconsistent, he must have been in error about his own preferences.

It is interesting to see why this is so common. In chosing a motor car there are a great many factors to be considered. The first source of error is that one or more factors may be overlooked. For example, a person may select one car as the best, but revise his decision when it is pointed out that his employer will pay him a lower mileage allowance because of the car's small engine, a factor quite overlooked in the original decision. A second cause of error is that no one car will generally be the best on all counts. It will, therefore, be necessary to 'trade-off' preferences. The person's chosen car may, for example, have the best upholstery but not the lowest fuel consumption, the extra fuel consumption being more than compensated for by the quality of the upholstery. In such complex decisions errors about the trading between different factors is almost unavoidable.

Common sense is wrong to think that the same possibilities of error do not infect mundane decisions, such as prefering celery to cabbage. It is thought that in such a judgement there is only one factor to be considered - in this case taste - so that factors cannot be overlooked, and trading between different factors is not called for. The simplifications involved here are, however, quite unrealistic. Many, many factors are involved even in

this simple choice; the nutrition to be obtained from celery and from cabbage, possible health risks from eating them, the need to teach children how delightful cabbage is, the effects of demand for the vegetables on celery and on cabbage growers, the digestibility of the vegetables, and so on. Mistakes about preferences are, therefore, possible even here. The eater may have an incomplete knowledge of the factors involved; he may, for instance, be quite ignorant about nutrition, and he may miscalculate the trading he does between the various factors, just as in chosing a motor car. There is nothing revolutionary about this view, for it is a fact of everyday experience that a person's preference for even such things as food can be altered by the acquisition of knowledge, about nutrition and the ability of his own digestive system, for example, or the influence his choice has over his children. Similarly, the same experiment mentioned above, concerning the choice of a motor car, can show a person that he is wrong in how he trades between factors when chosing food.

Here it might be objected that even if 'I prefer celery to cabbage' is corrigible, then something as weak as 'I prefer the taste of celery to the taste of cabbage' is not. The judgement may be weak, but it nevertheless involves many factors, and is therefore open to the kind of errors we have been discussing. For example, suppose that the taste of celery leads to an uncontrollable urge to eat more and more celery with disastrous consequences to the intestines, or that the taste itself drives people insane. It is a contingent fact that these are not consequences of tasting

celery, but there may be as yet unknown facts of the same sort which would make us reverse our preferences once disclosed to us. There can, of course, be no certainty that such facts will not be discovered. Exactly how facts may influence preference will be explored below.

My claim that statements of preference are corrigible may run counter to ordinary language. It does sound odd to say that a person prefers one thing to another, but is mistaken about it, but this locution is unavoidable on my view of value. To some extent it would appear that the doctrine that an agent has privileged access to his own preferences is built into our everyday language, but I can see no harm in challenging the doctrine and exposing a common error. In the same way as an extension of our knowledge of the physical world has led, quite naturally and correctly, to changes in our employment of terms like 'magic', 'witch', 'supernatural' and so on, advances in the philosophy of value should lead to revision in how we speak of preferences. This may seem shocking to those who see ethics as the unearthing of the rules of everyday discourse about values, but I reject their approach to the subject. Such exploration seems to me vaccuous and trifling, but I cannot go into this here for reasons of length. All I shall say is that I see the prime function of the philosophy of value or ethics as producing a methodology by which evaluative claims may be assessed, and such a methodology cannot be achieved by any kind of linguistic analysis[3].

The two claims of the metatheory together yield a very different picture of the making of preference judgements than that given by traditional theories. For one reason or another an agent may be lead to accept the judgement 'I prefer a to b', and this will be, by 1, a genuine sentence. By 2, however, the agent can never be sure of the truth of the sentence, i.e. he can never be sure that he really does prefer a to b. If he genuinely wishes to search for the truth of the matter he must, therefore, hold his preference claim open to criticism. He must search for factors which are relevant to the judgement but which he has overlooked, and he must look critically at the trading between the various factors which he has recognised. Since none of his preference claims can be justified, all of them should be open to criticism. This is the supreme rule of methodology for the assessment of preference judgements and it is entailed by the metatheory consisting of 1 and 2, just as Popper's metatheory entails his supreme rule.

2. Test Sentences

In chapter five it was pointed out that factual sentences have three properties which enable them to be considered as test sentences for value judgements in general, including, of course, judgements of preference:

1. Factual sentences may contradict preference sentences.
2. There are ways of coming to agreement about the acceptance or rejection of factual sentences.
3. Factual sentences are testable.[4]

Much attention was given to the first of these properties because of its originality. Many examples were given of the form $F \longrightarrow \text{not}(N_1 \text{ and } N_2)$ where F is a factual sentence and N_1 and N_2 value judgements. Since no example of form $F \longrightarrow$ not-N is known, as far as I am aware, no single preference sentence will be falsifiable by a factual sentence, although some conjunctions of preference sentences will be so falsifiable. Exactly the same is true in science. No single scientific test sentence can falsify a scientific theory, but only a conjunction of a theory and some initial conditions.

We saw in Chapter 4 that an attempt to falsify a scientific theory can be seen in two ways; as involving the deduction of the negation of a test sentence from the theory plus auxiliary hypothesis and initial conditions /or the development of a counter argument to the theory from an auxiliary hypothesis and initial conditions. It may help to refresh memories by considering an example. Let the theory be T = All samples of urine are acid; the auxiliary hypothesis be A = All acid samples turn litmus paper red; the initial conditions I = there is a litmus paper in a urine sample at p, t; the observation O = the litmus paper at p, t is not red.

T.I $\longrightarrow$ There is an acid at p, t

A.O $\longrightarrow$ not-(There is an acid at p, t)

Here we can see A.O as providing a counterargument to T.I, and thence to T itself. We can also view the situation as:

T.A $\longrightarrow$ not-(I.O)

Here the theory plus auxiliary hypothesis entails the negation of the test sentence I.O (remember that a conjunction of test sentences is a test sentence). If I is accepted by observation and if O is similarly accepted, then T.A must be regarded as falsified, which may be taken as a criticism of T.

Exactly the same is true of the falsification of a universal preference sentence by a factual sentence. Reasons for paying particular regard to universal sentences will be given in the next section. Imagine an individual who is greatly impressed by the need for man to live in harmony and balance with his environment and thinks, therefore, that any action which retains the balance of nature is preferable to an action which alters the balance of nature. In other words, he adopts the preference judgement U below. We may also suppose the individual to believe B below, which together with U entails that tolerating smallpox is preferable to eliminating the disease. The preference judgement V we may also attribute to the agent. Someone wishing to persuade the agent to alter his preferences may point out that tolerating smallpox will lead to more deaths than eliminating smallpox (C), so that eliminating the disease is, after all, preferable to tolerating it.

U = (x)(y) (x does not alter the balance of nature and y alters the balance of nature $\supset$ x pref y).

B = Eliminating smallpox alters the balance of nature and tolerating smallpox does not alter the balance of nature.

V = (x)(y) (x leads to fewer deaths than y $\supset$ x pref y)

C = Eliminating smallpox leads to fewer deaths than tolerating smallpox.

Formally, V.C provides a counterargument to B.U since;

B.U $\longrightarrow$ Tolerating smallpox pref eliminating smallpox.

C.V $\longrightarrow$ Eliminating smallpox pref tolerating smallpox.[5]

Faced with such a counterexample, the agent has no choice but to revise his values, giving up either V or U. If he gives up U and retains V, his opponent has convinced him that his original views on the eradication of smallpox are wrong.

The logic of the situation may be viewed in another way by noting that

U.V $\longrightarrow$ not (B.C)

Here the conjunction of two preference sentences entails the negation of a factual sentence (B.C) which we may, therefore, take as a test sentence. If B.C is accepted, by whatever rules are appropriate for the acceptance of factual sentences, then U.V must be regarded as false. Once again, our original agent is forced to reject either U or V. If U is rejected and V retained, then, as before, the agent has been convinced that his original preference for tolerating smallpox over eliminating it is wrong. Any testing of preference sentences by factual ones can be seen in both these ways.

To reinforce this conclusion consider another, more mundane, example. A person is offered a bowl of fruit containing apples and oranges after a good meal and immediately choses an apple. He then remembers that eating apples after large meals always makes him sick, whereas he can happily digest oranges. The realization of this fact makes him reverse his preferences and select an orange. Let:

W = Eating apple A pref eating orange O.

X = (x)(y) (x does not make me sick and y makes me sick ⊃ x pref y).

D = Eating apple A will make me sick and eating orange O will not make me sick.

One way to view the logic here is that X.D is a counterargument to W since

X.D ⟶ Eating orange O pref eating apple A

Alternatively, we may observe that

W.X ⟶ not -D

Here the conjunction of two preference sentences entails the negation of a factual test sentence which, if accepted, must lead to a rejection of either W or X. If W is rejected and X retained, then the agent has reversed his original preference for the apple over the orange.

What I wish to stress by considering such a lowly example is the obvious truth that a thousand times a day our preferences are altered by our awareness of facts. How this can be explained if it is denied that facts can falsify preference sentences in the way described, I fail to see. If facts lead us to alter our preferences, then facts must reveal our original preferences to be wrong.

3. Factual Content

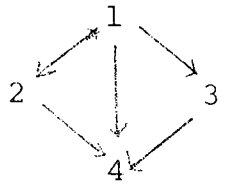
So far we have seen that an individual trying to decide what his preferences are cannot justify any preference claim which he finally adopts; all he can do if he is serious about trying to discover his own preferences is hold whatever preference judgements he ventures open to criticism. This is our supreme methodological rule, from which we may draw lower order rules of method. Criticism of preference claims is, as we have seen, to be by factual test sentences. In adopting a preference judgement, an agent is not committing himself to the judgement come what may, as some justificationists have thought (see chapter 2); rather he is accepting the judgement provisionally to see how it stands up to being tested. What kind of preference claim, then, should be adopted for testing? If the agent is seriously interested in investigating his preferences, then obviously he should adopt preference sentences which are easy to test in favour of ones which are difficult to test. He should, that is, favour highly testable (or highly falsifiable) preference sentences.

A preference sentence is to be tested by drawing from it the negation of some factual test sentence, so we may measure a preference sentence's testability by what I shall call its factual content - the set of factual test sentences which it forbids. Absolute values for factual content are impossible, since the number of factual test sentences contradicted by a given preference sentence is either zero or infinity, but relative values may still be given. If, for example, one preference sentence P_2 entails another, P_1, then any factual test sentence contradicted by P_1 will also be contradicted by P_2, so that the factual content, or testability, of P_2 is at least as great as that of P_1 (once immunizing strategems are prohibited - see section 5 below). The logical content of a sentence is the set of sentences which the sentence entails, so that the logical content of P_2 is at least as great as that of P_1. This suggests that logical content may also be used as a measure of factual content and of testability (again, once immunizing stratagems are prohibited).

Consider the preference sentences below:

1. Any £5 win pref any £4 win
2. A £5 win today pref any £4 win
3. Any £5 win pref a £4 win tomorrow
4. A £5 win today pref a £4 win tomorrow

The entailment relations between the four preference sentences may be represented:



It follows that the logical content of 1 is greater than that of 2 and 3, which are both greater than that of 4. 2 and 3 cannot be compared for logical content. From this we may say that 1 is at least as testable (has at least as high a factual content) as 2 and 3 which are, in turn, at least as testable as 4. The testability of 2 and 3 cannot be compared, at least by logical content. It follows that an agent seriously investigating his own preferences should, all things being equal, prefer to adopt 1 to 2 or 3 and 2 or 3 to 4, since this is the order of the ease with which the sentences can be tested.

The favour which should be shown for the adoption of preference sentences which are highly testable gives special place to universal preference sentences of the form 'all ---- pref all ....', or, better, (x) (y) (----x and ....y$\supset$ x pref y). 1 above is of this form, since it may be phrased (x) (y) (x is a £5 win and y is a £4 win$\supset$ x pref y). Such universal sentences are highly testable, as the discussion above illustrates.

4. Corroboration

Our picture is of an agent who has provisionally adopted some preference sentence for testing, favouring high testability in his choice, and who is trying to test his adopted sentence as strenuously as he can against factual test sentences. If his sentence is falsified in such a test, he must look for a better one, but if it passes a test he should regard it as successful. The problem for this section is to see how successful the sentence is in passing tests. When a preference sentence passes a test, we may say that it is corroborated by the factual test sentences involved, and we may use the term 'degree of corroboration' to measure the cumulative success which accrues to the sentence. Success for a preference sentence is the passing of a test, and the more severe the test, the greater the success of the sentence in passing it. A test is, of course, an attempt to falsify the preference sentence and a severe test is one which we expect the preference sentence to fail. A severe test is, therefore, one which we expect to falsify the preference sentence under test. If a preference sentence manages to pass such a test; then it is highly corroborated. If, on the other hand, the preference sentence passes only tests which we already expect it to pass, its success must be counted as small, and it is only weakly corroborated, though the tests be as numerous as we please. Success goes only to guesses which risk failure, and the greater the risk of failure, the greater their success in surviving. It must always be remembered, however, that a preference sentence's degree of corroboration is in no way a

measure of its probability or the likelihood of its being true, and in no sense is it predictive of future performance. Degree of corroboration is a measure of a preference sentence's success in passing tests. As such it says nothing about how future tests will turn out. A highly corroborated preference sentence may very well fail the very next test that is applied to it, and its degree of corroboration tells nothing about the likelihood of such failure.

The idea of corroboration may be illustrated by recalling the example concerning smallpox discussed in the second section. There it was shown how an agent might have his preference sentence U falsified, where

U = (x)(y)(x does not alter the balance of nature and y does alter the balance of nature ⊃ x pref y)

U, with the factual B, 'eliminating smallpox alters the balance of nature and tolerating smallpox does not alter the balance of nature', leads the agent to prefer tolerating smallpox to eliminating it. The counterargument employs the two claims:

V = (x)(y)(x leads to fewer deaths than y ⊃ x pref y)

and

C = Eliminating smallpox leads to fewer deaths than tolerating it.

V and C, of course, entail that eliminating smallpox is preferable to tolerating smallpox, counter to the agent's original preference. C is expected to be true, but suppose the agent can

provide us with arguments that show C to be false, against our expectations and that its contrary D is true.

D = Tolerating smallpox leads to fewer deaths than eliminating smallpox.

We may then observe that:

V.D ——→ Tolerating smallpox pref eliminating smallpox

which, of course, is in agreement with the agent's original conclusion. What has happened here is that the agent has drawn a conclusion from his preference sentence U and then provided us with an <u>independent</u> argument for the same conclusion. This argument uses a preference sentence which is acceptable to nearly everybody and which is not at the centre of any critical debate - we may call it a 'background' preference sentence - and a <u>highly novel</u> factual sentence. The novelty of D means that the argument gives a high degree of corroboration to U. U is highly corroborated because it has successfully challenged our expectation that C is true.

As with falsification, the logic may be viewed in a different way.

U.V ——→ not (B.C)

Here, U.V entail the negation of the factual test sentence B.C. If B is accepted, then since

U.V.B ——→ not C

the only way to retain U.V is to show that C is false. This is a great challenge, since C is expected to be true. Hence, if the agent manages to convince us that C is really false, this must be

taken as excellent corroboration of U.

A measure for the degree of corroboration of preference sentences may be based on Popper's measure for the degree of corroboration of scientific sentences. It will be remembered that Popper's measure for the corroboration of a theory x by a scientific test sentence y, C(x,y) is;

$$C(x,y) = \frac{p(y,x) - p(y)}{p(y,x) + p(y) - p(x.y)}$$

It is undoubtedly odd to talk of the probability of a preference sentence, though this is perfectly permissible on the views being developed here. To talk in this way may produce unnecessary obstacles to the understanding though, and so I will avoid doing so. If u is a preference sentence and V the negation of a factual sentence and if u entails V, then the corroboration of u by V is measured by:

$$C(u,v) = \frac{1 - p(v)}{p(v) + C_L(u)}$$

where $C_L(u)$ is the logical content of u (= 1 - p(u) if this locution is allowed).[6]

Absolute measures of corroboration are generally impossible, but comparison may be made of the degree of corroboration of two preference sentences or of the same sentence by two factual test

sentences. In particular, it can be seen that C(u,v) increases as p(v) decreases, so that for a high corroboration we require the negation of the test sentence entailed by the preference sentence to have a low probability. In the smallpox example, for instance, not-C is very unlikely given our present state of knowledge, so that finding not-C to be true confers a high corroboration on u.

5. Protection from Falsification

Popper has pointed out how it is possible to protect a cherished theory from falsification by making a whole series of what he calls ad hoc manoeuvres or immunizing (conventionalist) stratagems. A theory protected in this way never finds itself in disagreement with the facts, since all disagreement is discretely avoided, so that it may appear extremely successful. This success is, however, wholly illusory for it is achieved without risk of failure. According to Popper's supreme rule of methodology, all scientific theories are to be held open to falsification. It is necessary, therefore, to banish such protected theories from science, and they may be labelled 'pseudo-scientific'. The stratagems which produce these theories must also be prohibited, and these prohibitions form some of the lower order rules of Popper's methodology of science.

Exactly the same is true of preference sentences. The supreme rule of the methodology being developed here states that all preference sentences should be open to falsification, and we know that such

falsification can come from factual test sentences. It is possible, however, to protect any chosen preference sentence from falsification by factual sentences, and so we need rules to ensure that this does not happen. It is best to show this by means of an example.

In his defence of the utilitarian principle, Smart states boldly that whenever the principle comes into conflict with any value judgement, he is going to retain the principle and reject the recalcitrant judgement.[7] Let us consider how this can work. Smart accepts the utilitarian principle which we may briefly paraphrase as:

S = (x)(y)(x produces more happiness than y ⊃ x pref y).

Imagine now a possible criticism of S; that there are some acts which produce the greatest amount of happiness but which are not preferable to other actions because they cause avoidable deaths. Let:

E = a produces more happiness than b

F = b leads to fewer deaths than a

V = (x)(y)(x leads to fewer deaths than y ⊃ x pref y)

Now:

S.V ⟶ not (E.F)

E.F may be regarded as a factual test sentence for S.V. If E and F are both accepted as true, then at least one of S and V must be regarded as falsified. Smart would, of course, always insist that V be rejected, since he is committed to the defence of S. Obviously

if this policy were to be consistently followed, there would be no chance of falsifying S. S is always protected from criticism. The conformity of S to the facts cannot, however, be regarded as corroborating S since success can only be achieved at the cost of risking failure, and S has risked nothing. Returning to the measure of corroboration discussed in the previous section, it will be seen that this is indeed the case. If values like V are always adjusted to make S fit factual sentences known or expected to be true, then the probability of these factual sentences is extremely high, so that they confer very little corroboration on S. Employing Popper's notion of corroboration, therefore, effectively prevents Smart from claiming any success for his protected utilitarian principle.

This discussion obviously raises the question of when it is legitimate to save a principle like S from falsification by rejecting some other preference sentence, such as V, which figures in the test. Briefly, we may say that this is only allowable when the falsified preference sentence can be replaced by one which is <u>better</u>. What, then, are the conditions which make one preference sentence better than another?

6. <u>When Is One Preference Sentence Better than Another?</u>

Consider an agent who is seriously investigating his own preferences by submitting them to possible falsification from factual test sentences. Suppose that one particular preference sentence is falsified by such a test, when should the falsified sentence be <u>rejected</u> by the agent? In science, Popper points to the

need for a certain degree of toleration towards falsified theories, recognizing that nearly all theories are falsified, at least at their inception, so that rejecting a theory as soon as it fails a test would mean that science would be empty. The problem posed by the falsification may be solved without changing the theory; the falsification may, for example, be attributed to an erroneous auxiliary hypothesis, or false initial conditions. It will, of course, take time and effort to explore these possibilities, so that immediately rejecting the theory on falsification would be premature. What Popper suggests, as we have seen, is that the theory should be rejected only if it can be replaced by a better theory, and he gives us criteria for one theory being better than another.

The same problem arises in the testing of preference sentences. Any test of a preference sentence will involve other preference sentences and factual assertions, so that any falsification may be directed at these and not the preference sentence under test. As before, adopting the rule that a preference sentence is to be rejected as soon as falsified means that it is impossible to explore these possibilities. As in science, toleration towards falsified preference sentences is essential. My suggestion is that a preference sentence should only be rejected if it can be replaced by a better one, and that preference sentence P_2 is better than preference sentence P_1 if:[8]

1. All factual test sentences forbidden by P_1 are also forbidden by P_2
2. P_2 forbids some factual test sentences which are expected to be true and which are not forbidden by P_1
3. Some of the factual test sentences of 2 are shown to be false (thus corroborating P_2).

The reasoning behind this criterion is exactly the same as Popper's for the corresponding criterion for scientific theories. 1 ensures that P_1 and P_2 are rivals and that nothing is lost if P_2 replaces P_1, while 2 and 3 are needed to ensure that the testability of P_2 exceeds that of P_1 and that this is not the consequence of ad hoc stratagems.

7. Perspective

Having outlined a fallibilist theory of value, it may now be useful to place it in perspective by considering how the theory is related to some of the topics traditionally dealt with in ethics, and to traditional ethical theories.

(a) Objectivist and Subjectivist Theories of Value

Most traditional theories of value are either objectivist, holding that values are objective and independent of what people think them to be, or subjectivist, maintaining that values are reflections of inner feelings or dispositions so that no wedge may be driven between what an agent thinks his values are and what his values really are. A central theme in the history of ethics is the tension between theories of these two kinds. This tension exists

because subjectivist theories have been able to explain how value judgements can guide action, but have not been able to find any room for reason in ethics; whilst objectivist theories have found it easy to explain the role of reason in evaluation, but impossible to explain why the outcome of this reasoning should exercise any control over our actions. What I hope to show now is that the theory developed here has characteristics which are typical of traditional theories of both sorts, and that it can, therefore, account for the connection between values and action and also explain the need for evaluative reasoning. It will be best to start by a brief discussion of the failures of subjectivist and objectivist theories.

Rather than enter upon a tedious, and probably fruitless, search for a final definition of 'subjective', we can best begin by considering the crudest subjectivist theory, according to which value judgements are merely reflections of inner feelings of approbation or disapprobation. Such a theory holds that:

> The value of an object for a person is solely determined by the feelings of approbation or disapprobation which the object causes in the person. A

A explains the connection between value judgements and action very neatly, for there is no more reason to ask why a man seeks what he values and avoids what he thinks worthless, than there is to ask why he seeks pleasure and shuns pain. Where the theory fails is in finding a role for reasoning about values. At the root of this failure is a view such as B.

A person has immediate and complete knowledge of his own feelings of approbation and disapprobation. B

Since a person's feelings of approbation and disapprobation determine what value he places on things, B tells us that a person can have immediate knowledge of his own values. He knows immediately whether his feelings are of approbation or of disapprobation, and in either case he knows immediately the degree of his feelings. It follows from A that he also knows immediately the value of whatever it is that produces these feelings. But if this is so, then there is no need to reason or to argue about values, any more than there is need to offer reasons for or against 'I am in pain'. We can see, therefore, that the very way in which the theory connects judgements with action precludes it from finding any room for reasoning about values.

All traditional subjectivist theories of value have found themselves in a similar situation. In order to forge a connection between value and action, a person's values have been held to be a function of his likes, dislikes, desires, aversions, longings, loathings and so on; a collection we may subsume under the generic term 'interests'. At the same time it has been thought that immediate knowledge of such interests is possible, and, as before, the combination of these two claims has precluded the possibility of reasoning about value judgements.[9]

Turning now to objectivist theories, we may best begin with a typical theory of this sort, one which holds the utilitarian principal that:

> The value of an object is determined solely by the sum of the happiness or unhappiness which it produces for people.

This theory has an obvious place for reason, since reason is needed to determine how much happiness and unhappiness an object produces. The theory cannot, however, explain why values exert control over our actions. In judging that an object is valuable, a person is judging that it will produce a great surplus of happiness over unhappiness, but this, by itself, gives him no cause to cherish the object, or to seek its possession, or to advocate its adoption. He will do so only if he happens to possess a subjective desire to promote the general happiness of the human race.

We saw above how the assumption that immediate knowledge of interests is possible precludes any subjectivist theory from finding a place for reason. We can now see how this same assumption prevents any kind of objectivist theory from linking value and action. Any theory which explains this link must do so by showing how value is a function of an individual's interests - his desires, feelings, longings etc. If immediate knowledge of such interests is possible, then an individual has immediate knowledge of his own values. It follows that a person's opinions about his own values are final and beyond question, a state of affairs quite incompatible with any

objectivist view of value.[10] In other words, given the assumption that immediate knowledge of interests is possible, no objectivist theory of value can explain the authority which value judgements have over action.

Two responses to this problem have been traditionally explored. The first maintains that the value of an object is determined by some objective property, and then adds that, as a matter of fact, men favour and desire objects with this property. Bacon makes such a suggestion when he argues that self-interest is the sole determinant of value, and that people happen to seek what is in their own interest. Although this kind of move succeeds in linking value and action, it must fail, since the link it forges is only a contingent one. It hardly does to say that people just happen to seek what is valuable and avoid what is valueless. The link between value and action must be a necessary one.

The second kind of response to the problem of getting an objectivist view to explain the authority of value judgements is typified by Frankena's proposal.[11] He argues that we must distinguish justifying reasons from motivating reasons. Stating an objective evaluative fact may, he argues, provide a reason which justifies a value judgement, although it may fail to provide a motivating reason because it is irrelevant to the interests of the person to whom it is addressed. This suggestion is not, however, very helpful. Suppose that an objective evaluative fact justifies us in accepting that of two actions, P and Q, P is the better. Suppose also that the same

evaluative fact fails to provide a motivating reason to anyone, which is quite possible on Frankena's account. When we are thinking about whether to do P or Q, this means that the evaluation 'P is better than Q' is completely irrelevant to our decision. This is, at least *prima facie*, paradoxical. Unless someone can persuade us that such a striking situation is really possible, we must reckon it as contradictory. Unfortunately, Frankena says nothing to so persuade us. It would seem, therefore, that traditional objectivist theories of value cannot, after all, explain the connection between value and action.

We find, therefore, that the traditional inability of subjectivist views of value to find a place for reason, and the traditional failure of objectivist views to explain the authority of value judgements over actions both stem from the same assumption. This is that immediate knowledge of our own interests - our desires, aversions, longings and loathings etc. - is possible.

The theory developed here, of course, explicitly denies the possibility of such immediate knowledge, and so is able to explain both the role which reason plays in evaluation and the link between evaluation and action. Preference falls under the generic term 'interests' as defined above. That is, preferences guide actions. There is no more need to explain why a man strives for what he prefers than there is to explain why he seeks pleasure or avoids pain. My theory resembles subjectivist theories, therefore, in its ability to explain the authority which values have over actions. An agent

strives for what he prefers, but he can never be sure what his real preferences are. Reason is, therefore, required if the agent is to test his guesses about what his preferences are, and I have tried to show how these tests should proceed. In its ability to find room for reasoning about value judgements, my theory resembles objectivist theories of value.

(b) Dogmatism

It has often been commented that dogmatism is a bad thing in evaluative reasoning and the pejorative force of the term 'casuistry' is well known and perfectly fitting.[12] It is, however, difficult for traditional theories of value to explain why dogmatism is wrong. Many such theories have held that there exist some fundamental value judgements which are either arbitrary or self-evident (see chapter 2), but in either case the fundamental judgements must be held to be immune from any kind of criticism. If a fundamental value judgement clashes with any other value judgement, it must always be the latter which yields. My theory, however, finds a natural explanation for the repulsion felt against dogmatism in evaluation. Since we can only guess at what our real values are, any of our guesses may be wrong and so all must be open to criticism.

(c) Fact and Value

If I am correct, then many traditional ethical theories have grossly distorted the relationship between fact and value in

upholding the autonomy of value. It is an everyday event, however, for us to alter our preferences in the light of newly revealed facts. There is nothing startling or alarming about this, and yet if the doctrine of autonomy is held it is extremely difficult to see how this is possible. My theory, which explicitly denies autonomy can, of course, account for this phenomenon perfectly easily.

A second point here is that values often fall into a coherent pattern, so much so that we often talk of a system of values. There is nothing surprising about this, but often the system extends to facts as well as values. This will become clearer from the case studies to be considered in Part 3, Chapter 7 but for now simply consider two politicians having different value systems. Knowing their value systems it is often extremely easy to predict what their opinions on <u>factual</u> matters will be - what, for example, they think will follow from the abolition of corporal punishment, wage restraint, large scale unemployment, increasing student grants, abolishing the Monarchy and so on. This is difficult to explain on many traditional views of value, but perfectly easy on the view presented here.

(d) <u>Value-Neutrality of Philosophical Ethics</u>

It is often held that philosophical theories of value, or metaethical theories as they are sometimes called, should be value-neutral in the sense that they should entail no value judgements. The theory developed here is value-neutral in this sense. What the

theory lays down are rules of method which should be followed by an agent if he wishes to seriously investigate his own preferences. The theory does not entail that any agent ought to follow such-and-such rules, for whether or not the agent wishes to seriously investigate his own preferences is a decision which is his.

(e) Logic

Many traditional theories of value see their task as to reflect the sort of reasoning which actually takes place about value judgements, and many find this best done by postulating that some kind of 'weak' logic is involved in evaluative reasoning.[13] According to Hare, for example, moral words like 'ought' are not ambiguous, but they 'have two or more aspects to their meaning, one of which may on occasion be emphasized to the neglect of the others'. Presumably, in changing the emphasis between different aspects of its meaning the word's logical consequences also change. This, of course, provides an enormous opportunity to escape from criticism. If, for example, the statement 'X ought to do Y' is criticised on the grounds that it entails the unacceptable 'Z ought to do Y' (by something like Hare's universalizability), it can always be argued that this time 'ought' did not mean quite what the critic thought, and that its meaning here does not justify the inference to the unacceptable claim. Weak logic, of the sort proposed by Hare, makes it just too easy to dodge criticism. As Popper has stressed, if we wish to criticise statements of a certain sort, then we should employ the strongest logic available to us.[14] If we wish to criticise preference

judgements, therefore, we should apply to them the strictest logic we can. This is what I have done in developing my theory, where I have employed standard first order predicate calculus.

(f) Value Aggregation

In almost any decision, some people will receive what they regard as benefits from the decision and others will receive what they regard as costs. It is nearly always necessary, therefore, to balance what one group of people gain with what others lose. The problem of finding this balance is much discussed. In chapters 8 and 9 I hope to show that orthodox answers to this problem are inadequate, and in chapters 10 and 11 I will argue that the theory developed here can give a better solution.

(g) Preference and Principle

The following sort of objection is one which I expect many to make of the theory developed here. Consider a soldier who has a choice between staying safely in his trench where he can inflict no harm on the enemy and risking his life to wound or kill some of the opposition facing him. If he takes the second course of action, it is not because he prefers risking his life to staying safely in his trench; it is because he is convinced that it is his duty to inflict as much damage on the enemy as he can, even if this involves risking his life. This, I can see it being argued, is a general

point. Some actions reflect the agent's preferences, but sometimes the agent acts counter to his own preferences because he knows that some action is his duty or is morally right. In dealing only with preferences, I will be accused of forgetting that people can sometime 2a act from principle and that is what makes them moral agents and lifts them above the animals. What needs explaining, and what I have remained silent about, is that people can act against their preferenc when they recognize their duty or what is morally demanded. 1

This criticism involves a serious distortion of reality. It assumes that all preferences are a reflection of selfish interests If an agent acts according to the moral law (whatever that is) rather than his selfish interests, why should we not speak of him as preferring to act according to the moral law rather than according to his selfish interests? In all cases like the soldier's, an agent is faced with a conflict between preferences. The soldier, for example, prefers safety to risking his life, but also prefers inflicting harm on his enemies to not doing so. The facts of his situation, however, means that he cannot act according to both preferences. I think that philosophers have invoked ideas of duty or moral principles because they cannot understand how the soldier can <u>prefer</u> risking his life and engaging the enemy to staying safely in his trench. The fact that he prefers safety to risking his life hardly calls for explanation, but the fact that he might reverse this preference seems to call for the invoking of special duties or principles. This is where they are mistaken. The soldier has a set

of very fundamental preferences which hardly call for explanation and may never receive critical attention. He prefers living to dying, remaining whole to being wounded, safety to the danger of death, warmth to cold, the respect of his companions to their contempt, congratulations to vilification, carrying out his orders to prison, the safety of his family to their danger and so on. Some of these preferences can be used to provide reasons for the soldier to leave his safe trench and attack the enemy. He may be reminded, for example, that his family will suffer if the war is lost, that his comrades will despise him if he remains in the trench and so on. This, of course, leads to a conflict of preferences which the soldier may resolve by deciding that it is preferable to attack than to remain safe. In this way, the existence of 'unselfish' preferences poses no mystery and does not call for the postulation of principles, laws or duties unrelated to the agent's preferences.

(h) Generality

It is often said that a high degree of generality is desirable for evaluations and, indeed, evaluations of a low generality suggest ad hoc maneouvring and special pleading. For instance, compare 'all men have lives valued at £Z' with the collection of less general evaluations; 'all North Sea divers have a life valued at £W ', 'all farm hands have a life valued at £Y ', 'all housewives have a life valued at £X ' and so on. On traditional theories of value, however, it is hard to explain why this regard for generality should exist. This is true even of

theories, such as Hare's and Singer's, which employ some principle of universalizability.[15] This is easily explained on the present theory, however, since preference judgements of high generality are more easily tested than those of low generality.

(i) Hypothetical Examples

Often in evaluative reasoning, points are made with the help of hypothetical examples rather than real ones, and the use of these examples has received considerable attention from contemporary philosophers. For Hare, the use of hypothetical examples is important in ethics, and the whole of Singer's approach, which we briefly considered in chapter 2, depends upon the use of hypothetical examples beginning 'if everyone acted in this way then ----'[16]. We cannot, however, allow hypothetical examples to falsify preference sentences. If F is a factual sentence and N_1 and N_2 preference sentences, and if $N_1.N_2$ entails not-F, discovering the truth of F falsifies $N_1.N_2$ and discovering that F is false corroborates $N_1.N_2$. But if hypothetical cases are allowed to count for or against $N_1.N_2$, all is confusion. It is possible to imagine that F is true, from which we would conclude $N_1.N_2$ to be falsified, but it is also possible to imagine that F is false, in which case we would conclude that $N_1.N_2$ is corroborated. One and the same consequence of $N_1.N_2$ counts both for and against the conjunction.

There is, however, a distinct appeal to hypothetical examples in ethics. Suppose, for example, that someone suggests that it is <u>always</u> preferable to tell the truth rather than lies. We then present him or her with the hypothetical example of a maniac chasing a man who hides in a wardrobe. What should we say to the maniac when he asks 'where is the man I've been chasing?'? Is it not obvious that here we should tell a lie in preference to the truth, so that it is not <u>always</u> preferable to tell the truth rather than lies. This example is plausible, but I hope to show how its plausibility is based on a logical misunderstanding. How are we to formulate the assertion that telling the truth is <u>always</u> preferable to lying? The following seems reasonable;

$$T = (x)(y)(x \text{ is telling the truth and } y \text{ is lying} \supset x \text{ pref } y)$$

This is an undated universal sentence, and what it means is that there is no place and no time where x is telling the truth, y is lying and x is not preferable to y. Indeed, we may re-write the sentence as;

$$T^{*} = \text{not-}[\exists x\ \exists y\ (x \text{ is telling the truth and } y \text{ is lying and not-}(x \text{ pref} y))]$$

What makes T and T^{*} false is the existence, at some place and time, of a case where x is telling the truth, y is lying and x is not preferable to y. It is not rendered false because such a case can be <u>imagined</u> to occur. A hypothetical test is, however,

possible for the claim that such a case cannot (and not merely does not) occur. What is being claimed here is stronger than T and T^*. Perhaps we can capture what is being said by T+, which employs the operator 'necessarily' (Nec).

T+ = Nec $(x)(y)$ (x is telling the truth and y is lying $\supset$ x pref y)

What this means is that in all possible, and not just all real, cases telling the truth is preferable to lying. If the hypothetical incident described above is really coherent, then it is a counterexample to T+, which must be taken as falsified by it. In developing the fallibilist theory of value here, I have used only sentences like T, finding no use for ones such as T+. Against sentences like T, hypothetical counterinstances are ineffective. Our task is to guide ourselves through this world by our preferences. This, it must be admitted, is hard enough without seeking for guidance through all possible worlds.

(j) Other Starting Places

I said at the beginning of the present chapter that the fallibilist theory of value here developed does not depend upon employing 'preference', and that the theory could be couched in any of the traditional terms such as 'duty', 'obligation' and 'ought'. The case against the possibility of justifying value judgements

in Part 1 was perfectly general. It follows that an agent has no way of justifying claims about what his duties or obligations are, or about what he ought to do. If such claims are to be tested at all, it follows that they must be held open to criticism. Once this is recognised, the next step is to identify test sentences for each sort of claim. When this has been done, the methodology may proceed along exactly the same lines as before. If I can show that factual sentences can act as test sentences for claims about duties, obligations and what ought to be done, then the methodology for testing these claims will be exactly the same as the one for testing claims about preferences. The following examples show how factual sentences can be test sentences for each type of value claim.

X has a duty to clean his teeth each morning

X has a duty to do as his mother tells him

not-(X's mother tells X not to brush his teeth each morning)

Y is obliged to pay Smith £5

Y is obliged to keep Smith sober

not-(If Y pays Smith £5, then Smith will be drunk)

Z ought to shop at the smallest store in town

Z ought not shop at Quibley's

not-(Quibley's is the smallest store in town)

CHAPTER 6 - FOOTNOTES

1. For some attempts to base a theory of ethics upon science see: R. Ackoff, 'On A Science of Ethics', Philosophy and Phenomenologica Research, 9, 1949, pp. 663-672; C. Baylis, 'The Confirmation of Value Judgements', Philosophical Review, 61, 1952, pp. 50-58; P. Caws, Science and the Theory of Value, Random House, 1967, pp. 50-58; A. Edel, Method in Ethical Theory, Routledge and Kegan Paul, 1963; Science and the Structure of Ethics, University Chicago Press, 1961; Ethical Judgement: The Use of Science in Ethics, Free Press, 1955; A. Gerwith, 'Positive Ethics and Normative Science', Philosophical Review, 69, 1960, pp. 311-330; D. Lackey, 'Empirical Disconfirmation and Ethical Counterexample', Journal of Value Inquiry, 10, 1976, pp. 30-35; H. Margenau and F. Oscanyon, 'A Scientific Approach to the Theory of Values', Journal of Value Inquiry, 3, 1969-70; pp. 163-172; E. Mesthene, 'On the Need for a Scientific Ethic', Philosophy of Science, 14, 1947, pp. 96-101; D. Monro, Empiricism and Ethics, Cambridge University Press, 1967; F. Schoeman, 'A Rational Approach to the Foundations of Ethics', Journal of Value Inquiry, 8, 1974, pp. 241-251; E. Walter, 'Reasoning in Science and Ethics', Journal of Value Inquiry, 8, 1974, pp. 252-265.

Amongst others who have proposed, or come near proposing a fallibilist view of ethics are: J. Brown, 'The Appraisal of Value Judgements', Ratio, 18, 1976, pp. 56-72; A. Edel, Method in Ethical

Theory, Routledge and Kegan Paul, 1963, p.318; R. Hare, Freedom and Reason, OUP, 1963, pp.87-8, 92 and 136; C. Humphrey, 'The Testability of Value Claims', Journal of Value Inquiry, 3, 1969-70, pp.221-7; E. Mesthene, 'On the Need for a Scientific Ethic', Philosophy of Science, 14, 1947, pp.96-101; and especially J.W.N. Watkins, 'Negative Utilitarianism', Proceedings of the Aristotelian Society, Suppl. Vol., 37, 1963, pp.95-114, and 'Decision and Belief' in Decision Making, BBC, 1967, pp.9-26.

2. Assuming that the relationship of preference is transitive. The best argument I know for this considers an agent who prefers a to b, b to c and also c to a. If we have a and the agent has b and c to begin with, since he prefers a to b he will swop his b for our a and also be prepared to pay us money to accomplish the exchange. We now have b and he has a and c. Since the agent prefers b to c he will pay us to exchange our b for his c. We now have c and he has a and b. Since he prefers c to a he will pay us to swop our c for his a, leaving him with b and c and us with a. This, of course, is the original position, so we can repeat the exercise until the agent has no more money or other valuables to give us. The only escape from this embarrassing situation is to insist that preferences are transitive.

3. D. Collingridge, 'The Failure of Language in Ethics', Journal of Value Inquiry, 9, 1975, pp.81-94.

4. Not all factual sentences possess all three properties, of course, but many do. All test sentences for preference sentences will be factual, but not all factual sentences will be test sentences.

5. Here, as elsewhere, a pref b means that a is more valuable than b; it does not mean that a is at least as valuable as b.

6. This is easily relativized to 'background knowledge' - see K. Popper Logic of Scientific Discovery, Hutchinson, 1959, appendix *ix.

7. J. Smart, Outline of a System of Utilitarian Ethics, Melbourne University Press, 1961.

8. The satisfaction of 1-3 is a sufficient, but not a necessary condition for P_2 being better than P_1. 1 may be generalized to cover instances where the two preference sentences do not forbid the same factual test sentences - see K. Popper, Conjectures and Refutations, pp. 215-250.

9. See for example, B. Blanchard, Reason and Goodness, George Allen and Unwin, 1961, pp.89 ff.; P. Nowell-Smith, Ethics, Penguin, 1954, pp. 210 and 320; S. Touluin, The Place of Reason in Ethics, Cambridge University Press 1958, chapter 3.

10. This is mentioned by a host of writers. W. Frankena mentions most of them in his 'Obligation and Motivation in Recent Moral Philosophy' in A. Melden (ed), Essays in Moral Philosophy, University of Washington Press, 1958, pp.40-81. To his list may be added: B. Blanchard op. cit., pp. 34-35; D. Monro, Empiricism and Ethics, Cambridge University Press, 1967, p. 105; W. Stace, The Concept of Morals, Macmillan, 1937, pp. 41-43. See also chapter 2 where many objectivist theories are criticised on this point.

11. W. Frankena, op. cit.

12. See, for example, W.W. Bartley, 'Theories of Demarcation Between Science and Metaphysics' in I. Lakatos and A. Musgrave (eds), Problems in the Philosophy of Science, North Holland, 1968.

13. R. Hare, Freedom and Reason, Oxford University Press, 1963, p. 73. See also: L. Becker, On Justifying Moral Judgements, Routledge and Kegan Paul, 1973; L. Cohen, '3-Valued Ethics', Philosophy, 26, 1951, pp. 208-227; J. Findlay, 'The Methodology of Normative Ethics', Journal of Philosophy, 58, 1961, pp. 757-764; G. Kerner, 'Approvals, Reasons and Moral Argument', Mind, 71, 1962, pp. 474-486; V. McGill, 'Scientific Ethics and Negotiation', Proceedings and Addresses of the American Philosophical Association, 42, 1968-9, pp. 5-20; D. O'Connor, 'Validity and Standards', Proceedings of the Aristotelian

Society, 57, 1956-7, pp.207-228; P. Shaw, 'Ought and Can', Analysis, 25, 1964, pp. 196-197; J. Stolnitz, 'Notes on Ethical Indeterminacy', Journal of Philosophy, 55, 1958, pp. 353-367.

14. K. Popper, Objective Knowledge, p.304 ff.

15. Don Locke, The Trivializability of Generalizability' , Philosophical Review, 45, 1969, 415 - 427.

16. R. Hare, op. cit.; M. Singer, Generalization in Ethics, Eyre and Spottiswoode, 1963.

CHAPTER 7

The Fallibilist Theory of Value - Case Studies

The aim of this chapter is to apply the fallibilist theory of value of the previous chapter to some evaluative debates. In this, the points of interest fall into three categories; the public nature of value debates, the relation between evaluative and factual elements in these debates, and the testing of value judgements by debate. Before considering the case studies in detail, it will be useful to say a few words under each of these heads.

(a) The Public Nature of Value Debates

A very common view of value judgements, some of whose versions were described in chapter 2, holds that an agent is free to adopt whatever values he choses, provided only that he is logically consistent. Factual considerations are not relevant to his choice of values; evaluation is autonomous, a view often ensured by upholding Hume's Rule. What values an agent holds may be seen as resulting from the agent making some sort of fundamental, free commitment. I have criticised such views in chapter 2, and also the general justificationist assumptions behind them in part 1. For our present purposes, however, it is necessary to point to a so far unnoted feature of these views of evaluation. This is that evaluation is a necessarily <u>private</u> affair. This is at its clearest when the adoption of a value judgement is seen as stemming from some fundamental commitment to the judgement itself, or to some other value judgement from which it follows. The commitment is subjective in the sense that only the agent can have first hand knowledge of it. Nobody else can argue with him that he has not committed himself when he denies this. Evaluation, on this view, is essentially private; a totally personal affair. While it makes sense to speak of a group of people, such as a company or a political party or a University department, as having certain values, this

must be seen as resulting from a series of happy accidents about what private commitments the individuals of the group happen to make.

On the theory of value proposed in the last chapter, however, a very different view has to be taken. Private commitment to values has no part to play in such a theory. Commitment is supposed to justify the agent's acceptance of a value judgement, but such justification is altogether impossible. Instead, the agent must realise that none of his value judgements is justifiable, so that all should be open to criticism and possible falsification and rejection. This criticism proceeds through trying to find factual sentences which falsify the value claim in question and for which there are reasons for accepting. This is a public matter. Nobody has privileged access to the facts which can test a particular evaluative claim. The agent himself may be aware of facts which falsify or corroborate one of his value judgements, or someone else may point to the existence of this set of facts. Whether the fact falsifies, or the extent to which it corroborates, the value judgement is an objective, public issue. Thus, the fallibilist theory of value emphasizes the public dimension of evaluation denied by justificationist views of value.

This may be illuminated by Popper's three worlds. Popper's world 1 is the world of physical objects; the second is the world of private mental states and world 3 is the world of intelligibles, or of ideas in the objective sense[1];

> ... it is the world of possible objects of thought: the world of theories in themselves, and their logical relations; of arguments in themselves; and of problem situations in themselves.

Traditional theories of knowledge place great emphasis on the second world, the world containing the scientists' perceptions and experiences, because it is upon these that the whole structure of science is supposed to

rest. Popper rejects this psychologistic approach; arguing that justification for scientific claims is not possible, but that they can nevertheless be subject to critical appraisal.[2] Criticism does not rest upon receiving special privileged information about one's own mental state. Instead it is a public enterprise conducted in the third world. An individual scientist must make up his own mind about, say, the rejection of some physical theory, and in doing so he will of necessity employ only what relevant knowledge he has himself managed to acquire during his professional career, but whilst this is a statement of great profundity on justificationist epistemologies,[3] for Popper this is a banality. If the individual scientist has been unaware of information relevant to the fate of the theory, then this shortcoming may be remedied by his colleagues. If his judgement about the relationship between fact and theory is shaky, his colleagues may criticise him on this score. The fate of the theory is something which is to be settled, not by private introspection, but by public debate - or debate in the 3rd world.

Exactly the same is true of the appraisal of a value judgement according to the fallibilist theory of value. In deciding, for example, that some value judgement of his is falsified, an agent must conduct his own logical deductions and can only use knowledge which he has managed to acquire for himself in the past. But, as before, there is nothing profound in any of this. If his logic is faulty, this can be pointed out; if he is unaware of relevant facts, these may be made known to him. In other words, the appraisal is best seen, not as an essentially private affair of mysterious commitments, but as a public, 3rd world matter involving, at least potentially, many people. Logic was depsychologized in the early part of this century, and Popper has continued the process to epistemology. It is now time to de-psychologize ethics and the theory of value. For this reason, all of the case studies given below concern highly public evaluative debates. Things are much clearer here, of course, than in purely 'private debates' between an agent and his conscience.

(b) Fact and Value in Debate

On the traditional view of value, any disagreement about a particular evaluation either rests upon a confusion over facts or is irreconcilable. If all disagreements about factual matters are resolved between the two parties, and if they still differ over some evaluation, there is nothing which can be done to settle this dispute, assuming that both parties use language properly and are logically consistent. Evaluation is essentially private and subjective, so if one agent values an object differently from another agent, there is really no more dispute between them than if one were to say he has a pain in his foot, and the other to say that his own foot is free from pain. On this view, we would expect debates about values to progress so far, when disagreements over factual matters are gradually cleared up, and then to halt stubbornly and finally, when the different values of the two parties meet head on.

Inspection of real debates over evaluative questions, like the ones which follow, show this to be a very mistaken view. Debates about values do not need to end with an irreconcilable clash of opinions; indeed the progress of such a debate can usefully be seen as the avoidance of such clashes. A second feature of debates about values which is very surprising on traditional views is that they can be very protracted and extended, and yet contain no explicit discussion of values whatever. This is, however, understandable on the fallibilist view of value.

Consider two parties to a debate A and B, A wishing to defend the value judgement V, B seeking to deny it or to defend $\bar{V}$. The interesting case is where V is a fairly highly universal and precise value judgement - i.e. one of high factual content. Assume that both parties have the values $V_1 \ldots V_N$ in common, so that these may be regarded as background values in the debate. How can A and B conduct a reasoned debate? A can assert V and

B assert $\bar{V}$, but this simply produces a head on clash between them and the end of all debate. The clash may, however, be avoided if the parties employ only values chosen from $V_1 \ldots . V_N$, because these they share. To argue his case, A could produce a factual claim F_1 which, when coupled with some set of background values from $V_1 \ldots . V_N$, entails V. This, however, is rendered impossible because, in general, the factual content of V will so greatly surpass that of any of the background values. There will, in general, be no pathway from background values to some strong claim such as V. There is, however, a pathway from background values to the negation of V, and this may be exploited by B. B may bring forward a factual claim F_2 which, when coupled with background values from $V_1 \ldots . V_N$, entails $\bar{V}$. If F_2 is shown to be acceptable, then V is falsified and A must revise his position. If, on the other hand, F_2 is expected to be true and yet is shown to be false, then V is corroborated, the degree of corroboration depending upon the novelty of $\bar{F}_2$.

Formally:

$$F_2 . V_1 \longrightarrow \bar{V}$$

So that

$$V_1 . V \longrightarrow \bar{F}_2$$

The first entailment represents B's counterargument to V. The second views the argument in a slightly different way. If V_1 is accepted, then F_2 falsifies V, whilst $\bar{F}_2$ corroborates V.

We can now see how the debate can progress and avoid the head on clash which threatens it. Instead of a stubborn clash of values, the debate can proceed to consider the crucial issue of whether F_2 is true. This factual issue is only made relevant to the debate by the set of background values, but generally these values receive no explicit mention in the debate. The

reason for this is that the values are shared and, generally, shared because they are rather trivial and mundane. Thus all the important issues over which an evaluative debate ranges are nearly always purely factual. This leads naturally to the third feature highlighted by the studies which follow; the testing of value judgements.

(b) The Testing of Value Judgements

The case studies will, I hope, illustrate the main features relevant to the testing of value judgements, as described in the previous chapter. Important here are the factual content of the value judgement being tested (and hence its universality and precision); the existence of immunizing strategems to protect some chosen value from criticism; the novelty of the factual claim involved in a test of a value judgement, and the degree of corroboration resulting from this novelty; the replacement of one value judgement by another, and better, one under force of criticism; and the testing of background values. A few remarks on this final point will link with the earlier discussion about the relationship between fact and value in an evaluative debate.

Suppose, in the example of the previous section, that B has argued for $\bar{V}$ by bringing forward the factual claim F_2, which entails $\bar{V}$ when conjoined with the background value V_1. Suppose too that A accepts F_2 because of the evidence adduced by B in its favour. A has to admit defeat in the debate unless he can argue that V_1 is false. It must be remembered, of course, that V_1 belonging to the background of the debate is no justification of V_1. A background value is as beyond justification as any more spectacular value judgement; all that distinguishes it is its provisional acceptance by both parties. If A can bring V_1 into question, then it can no longer be reckoned a background value, and it must leave the periphery of the debate

for its centre. A can question V_1 by appealing to some factual claim F_3, which together with the remaining background $V_2 \ldots . V_N$, entails $\overline{V}_1$. No doubt B will wish to deny F_3, and so the debate may continue, but will revolve around the acceptability of this factual claim. Once again, progress in the debate is only possible because of this shift to factual issues. If F_3 resists B's attempts at criticism, then B may attempt the falsification of the background values used in A's falsification of V_1, and so on.

(d) Value and Decision

Having noted some of the main points to be illustrated by the case studies which follow, it must now be noted that each case study concerns a decision. Making a decision is choosing the most preferred course of action from a number of options, and so is a special case of ordering preferences. It might be thought, however, that there are special features which debates about decisions have which are not generally found in evaluative debates where decisions are not involved. I hope that the analysis of the following debates according to the fallibilist theory of value will not be distorted in this way. It is true that additional questions arise when decisions have to be made, rather than some list of preferences drawn up in an armchair, but these questions are not our concern at the moment; they will be considered in the remaining part of this work. For our present purposes, there is no distortion introduced by considering only evaluative debates which must end in a decision.

Case Study 1

The Corporal Punishment Bill

Parliamentary debates are a rich source of material to which the theory of value developed earlier may be applied. Rather than consider the grand debates on urgent political issues of the day, I have chosen for our present purpose a Private Member's Bill receiving its second reading in the House of Commons on a sleepy Friday afternoon. Such debates often contain more real attempts at dialogue and less Party bickering than grander ones. The Bill is the Corporal Punishment Bill, moved by Mr. Graham Page on Friday, 29th April 1977.[4] The long title tells us that the Bill aims to "Permit a sentence of corporal punishment upon a person convicted of an offence involving bodily harm to another or malicious damage to property". The chief point I wish to bring out is that although the question whether or not to re-establish corporal punishment is clearly an evaluative one, the debate contains hardly any mention of values, all the important issues being factual. I hope to show how the theory developed here can accomodate what, on conventional views of value, must be a highly surprising feature of the debate.

The Bill would enable corporal punishment by birch, cane or strap, under circumstances and for crimes laid down in affirmative orders of the Secretary of State. To expedite punishment the Secretary of State may also lay before the House affirmative orders enabling a convicted person "to waive an appeal against conviction or sentence, sufficient to allow the sentence of corporal punishment to be inflicted before the expiration of the time allowed for appeal" (Clause 3(e)). The Bill's mover, Mr. Page, saw the operation of this clause in this way. If a Magistrate wishes to impose a sentence of corporal punishment on a convicted person, then he may ask whether the offender would appeal against such a sentence. If he does

not wish to appeal, then the sentence would be passed and would be carried out as quickly as possible. If the offender wishes to appeal, however, corporal punishment is likely to be ineffective due to the great delay between offence and punishment, and so some other sentence would probably be favoured by the Magistrate. For juvenile offenders, the choice would lie with the offender's parents. The Bill sets no upper or lower age limit for corporal punishment, nor would it exclude women from receiving such punishment, although Mr. Page saw both issues as open to debate.

The arguments in favour of the Bill brought forward by Mr. Page and his supporters are as follows:

F_1 Public opinion for corporal punishment is very strong.

F_2 The increasing incidence of vandalism and crimes against the person show that present punishments are not effective.

F_3 Corporal punishment would be an effective deterrent against vandalism and crimes against the person. (This would be helped by the short time between punishment and offence envisaged by the Bill).

F_4 A growing proportion of crime is committed by juveniles.

F_5 Corporal punishment would be especially effective as a deterrent to juveniles.

All of these arguments are entirely factual. As is perfectly normal, the Bill's supporters have not made the evaluative claims which lie behind their argument explicit. It is not hard to see why this should be so, for the value claims are all very mundane. We might reconstruct some of them as follows:

V_1.(x)(y)(If punishment x reduces the incidence of violent crimes and wilful damage, and punishment y does not, then x pref y).

V_2(z)(If public support for a measure z is very strong, then Parliament should effect z).

With F_2 and F_3, V_1 yields the conclusion that corporal punishment is preferable to existing punishments, a conclusion also drawn from F_2, F_4, F_5 and V_1. With F_1, V_2 shows that Parliament should introduce corporal punishment and so favour Mr. Page's Bill.

The thing about these evaluative claims is that they are more or less accepted by all parties to the debate. In the words of the theory proposed here, they are background values. Opponents of the Bill can argue against it in two ways. They can attack the evaluative parts of the case made out by the Bill's supporters, or else the factual parts. The first offers little hope since the Bill's supporters have used only background values shared by both sides - indeed, to have used more contentious evaluative claims would have been to play into the hands of their opponents. What the Bill's opposers should do, therefore, is question the factual claims used by Mr. Page and his friends and find further facts which, when coupled with background values, show the Bill to be bad. This is exactly what we find.

The factual claims criticised were F_3 and F_5 - that corporal punishment would be an effective deterrent against vandalism and crimes against the person, and especially for juveniles. Six reasons were given for denying this claim.

(a) The delay between offence and punishment would be too great for an effective deterrent even under Mr. Page's scheme. To this it was said that such delay could be reduced to a matter of a few days.

(b) Statistical analyses by two earlier committees, Cadogan (1938) and Barry (1960), showed that among convicted prisoners, those birched were more likely to return to crime than those punished in other ways. This is further supported by the experience of Magistrates who refrained from passing sentences of corporal punishment, even when it was in their power to do so, because they had concluded that it was ineffective. As the debate

developed four replies were made to this criticism by the Bill's supporters. It was stated that the findings of Cadogan and Barry were irrelevant because they were principally concerned with gang violence. It was also argued that figures concerning the Isle of Man, where corporal punishment was still administered, refuted the findings of the two reports and that anyway the great growth in crimes of violence made the reports outdated. Finally it was argued that even if corporal punishment was less effective as a deterrent to convicted offenders, it might still deter more people from ever beginning a criminal career than other punishments.

(c) It was argued that if corporal punishment is a deterrent then behaviour in British schools, the only ones in Europe retaining such punishment, should be better than in European schools; something for which there is no evidence. In reply to this criticism it was stated that behaviour in British schools would be even worse if corporal punishment were abolished there.

(d) Several critics of the Bill argued that corporal punishment would be chosen by an offender as the lesser of two evils, so that it could hardly be reckoned a deterrent. In reply, it was argued that it would still be a deterrent, even though a somewhat weakened one.

(e) It was also argued that corporal punishment might well increase the level of violence in society, by adding to the glamour and status of an offender so punished, especially a juvenile. One objection to this was that it ought to apply to all forms of punishment, so that imprisonment and fines would also lead to a higher crime rate. A second objection was that even if corporal punishment leads to more violence by some sections of the community, it might reduce the total level of violence by its deterrent effect.

(f) A slightly different point was made by one objector to the Bill who argued that corporal punishment, although quite possibly a deterrent, was not the best deterrent. A better one would be the full use of existing powers to impose stern penalties of imprisonment or fines. In reply to this, it was pointed out that these powers were not used for some reason, and that other punishments should, therefore, be attempted.

As well as these debates on purely factual questions, which occupied most of the discussion, evaluative questions were raised three times. Each of these illuminates the fallibilist theory of value in an interesting way. One opponent of the Bill objected to the value judgement V_2, implicit in the case made out by the Bill's supporters, by arguing that public opinion would strongly favour such barbaric penalties as castration for sexual offenders. It follows from V_2 that Parliament should introduce castration as a punishment for these offences, something which both sides in the debate would never accept. This is, of course, a perfectly typical criticism of a background evaluative claim by a factual one. Formally we can view it in the following way.

V_2 (z)(If public support for a measure z is very strong then Parliament should effect z).

V_3 Parliament should not effect castration for sexual offenders

---

$\bar{F}_6$ not (Public support for castration of sexual offenders is very strong).

It is pointed out that F_6 is true, so that at least one of V_2 and V_3 is false. V_3 is a background value even more basic than V_2, so that V_2 should be taken as falsified. Thus V_2 begins the day by being a background value, tentatively accepted by all parties to the debate, but as the debate develops it is questioned and must, therefore, abandon the periphery for the centre of debate and lose status of background value.

The second evaluative issue raised is slightly more complex. It was argued that corporal punishment is a degrading and inhuman punishment and so one that should not be used; indeed one that could not be used because of the European Convention on Human Rights (Article 3), signed by the United Kingdom. Whether corporal punishment is inhuman is clearly an evaluative question, but it is interesting to see how the debate's participants came to grips with it. This they did by considering various factual issues. The relevant factual questions were singled out by means of background value judgements. It was accepted by all that inhuman punishments should not be inflicted and that corporal punishment inflicted by a parent or teacher is not inhuman. The factual question then becomes whether the relationship between the state and the offender is similar to that between teacher or parent and offender. If it is, then corporal punishment by the state is no more inhuman than corporal punishment by a parent or teacher. If the relationship is significantly different, however, then corporal punishment by the state may well be considered inhuman. In this way what threatens to be a bald, intractable clash of values is converted into a factual question which can be handled in debate.

Formally, we may view the Bill's supporters as advancing the following argument.

V_4 If corporal punishment by parent or teacher is not inhuman and if the relationship between parent or teacher and offender is similar to that between the state and offender, then corporal punishment by the state is not inhuman.

V_5 Corporal punishment by parent or teacher is not inhuman.

F_7 The relationship between parent or teacher and offender is similar to that between the state and offender.

V_6 Corporal punishment by the state is not inhuman.

V_4 and V_5 may be viewed as background value judgements, whilst F_7 is factual. Opponents of the Bill argued that F_7 is false, because the parent or teacher has a far greater knowledge of the likely effects of the punishment on the offender than has the state, and because the relationship between the punishing parent or teacher and the offender can still be one of affection, an emotion unknown to the state. This does nothing to show that corporal punishment by the state is inhuman, but it effectively destroys the argument made out by the Bill's supporters.

The third evaluative issue to be raised was done so by an opponent of the Bill, Ms Maureen Colquhoun. She held that young people 'need to be left alone as much as possible by adults' and that they ought to be given rights as people. The Bill was wrong because it would lead to adults inflicting pain on children and would give adults too much power over young people. Ms Colquhoun's entry into the debate led to nothing. Nobody agreed with her and nobody thought it important to criticise her claims. This is hardly surprising because whatever Ms Colquhoun meant to say it obviously rests upon some value claim which is special to her and not common to the other participants in the debate. To engage in a debate about an evaluative question like the restoration of corporal punishment one must employ value judgements shared by one's opponents, or else the debate degenerates into a simple and totally intractable clash of rival value claims.

We cannot discover if anyone changed sides during the debate because of the arguments presented there. All we can record is that the Bill was lost by 17 votes to 6.

Case Study 2

Small is Beautiful

E.F. Schumacher's influential book Small is Beautiful[5] does not consist of a closely reasoned, continuous argument, but is a collection of ideas around a central theme. Schumacher is a vociferous critic of modern technology and its handmaiden, economics, both of which he wishes to reform drastically. His writing is, therefore, an intimate mixture of factual and evaluative claims which it may be rewarding to analyse from the point of view of the fallibilist theory of value. In particular, I shall look at the relationship between Schumacher's factual and evaluative claims.

Schumacher's criticism of conventional economics is that it regards consumption as the only good, and pays no regard to the burden which this consumption places on man's environment and on future generations who may have nothing left to consume because of our own greed. Despite its limited outlook, economics plays an overwhelmingly important part in our decision making.[6]

> Call a thing immoral or ugly, soul-destroying or a degradation of man, a peril to the peace of the world or to the well-being of future generations; as long as you have not shown it to be 'uneconomic' you have not really questioned its right to exist, grow and prosper.

> Economics, moreover, deals with goods in accordance with their market value and not in accordance of what they really are. The same rules and criteria are applied to primary goods, which man has to win from nature, and secondary goods, which .... are manufactured from them. All goods are treated the same, because the point of view is fundamentally that of private profit-making,

and this means that it is inherent in the methodology of economics to ignore man's dependence on the natural world.

(In) the market .... there is no probing into the depths of things, into the natural or social facts that lie behind them. In a sense, the market is the institutionalisation of individualism and non-responsibility. Neither buyer nor seller is responsible for anything but himself.

The dominance of economics amounts to;[7]

.... the religion of economics, the idol worship of material possessions, of consumption and the so-called standard of living, and the fateful propensity that rejoices in the fact that 'what were luxuries to our fathers have become necessities for us'.

This dependence on economics had led us to adopt technologies of ever increasing size. Our consumption of the Earth's non-renewable resources is now on such a gargantuan scale that many vital resources, in particular fossil fuels, are likely to be exhausted very soon. The pollution produced by our ever growing industries endangers the delicate mechanisms which make the Earth habitable. Work has become alienating and meaningless as more and more skills are sacrificed to machinery in the name of economic efficiency. In the developing world, production has been increased with scant regard to the distribution of the wealth so created, so that rich, sophisticated city dwellers form isolated islands in a sea of traditional poverty. For conventional economics;[8]

> The farmer is considered simply as a producer who must cut his costs and raise his efficiency by every possible device, even if he thereby destroys - for-man-as-consumer - the health of the soil and the beauty of the landscape, and even if the end effect is the depopulation of the land and the overcrowding of cities.

Schumacher sees as a solution to all these problems the adoption of what he calls Buddhist economics in place of conventional economics. The difference between these approaches is nowhere more marked than when they are applied to labour. In conventional economics, a man's labour is valuable if he can produce more saleable goods than he can by using his efforts elsewhere. The Buddhist point of view sees work as offering a man a chance to use his skills, and to develop them; as overcoming the worker's ego-centredness by his assisting other workers and as a means of producing the necessities of life.[9]

> To organize work in such a manner that it becomes meaningless, boring, stultifying, or nerve-racking for the worker would be little short of criminal; it would indicate a greater concern with goods than with people, an evil lack of compassion and a soul-destroying attachment to the most primitive side of this worldly existence.

Whereas conventional economics considers consumption as the sole good, Buddhist economics seeks an optimal pattern of consumption, one which produces a high degree of human satisfaction with a low rate of consumption. The modest use of local resources is favoured over a consumption so gross as to require imports from all parts of the world. In this, Buddhist economics is aware of the sharp difference between those resources which renew themselves and those which may be exhausted, a difference to which conventional

economics is blind. Similarly, Buddhist economics would never countenance technologies which do violence to the long term stability of the environment and the eco-system.

Schumacher is somewhat despairing about the possibility of replacing conventional economics in the West, but sees a chance for developing countries to employ a more human, Buddhist economics to guide their progress. Conventional economics sees advance as the production of more and more goods, irrespective of their distribution. Buddhist economics, on the other hand, sees the improvement of the lot of the poor as the chief goal. The first requirement is work for the poor, but the great obstacle here is finding capital to invest in the development of new workplaces. Schumacher suggests, therefore, that developing countries should invest in what he calls intermediate technology.[10]

> If we define the level of technology in terms of 'equipment cost per workplace', we can call the indigenous technology of a typical developing country - symbolically speaking - a £1-technology, while that of the developed countries could be called a £1000-technology. The gap between these two technologies is so enormous that a transition from the one to the other is simply impossible. In fact, the current attempt of the developing countries to infiltrate the £1000-technology into their economies inevitably kills off the £1-technology at an alarming rate, destroying traditional workplaces much faster than modern workplaces can be created, and thus leaves the poor in a more desperate and helpless position than ever before. If effective help is to be brought to those who need it most, a technology is required which would range in some intermediate position between the £1-technology and the £1000-technology. Let us call it - again symbolically speaking - a £100-technology.

> Such an intermediate technology would be immenseley more productive than the indigenous technology (which is often in a condition of decay), but it would also be immensely cheaper than the sophisticated, highly capital-intensive technology of modern industry. At such a level of capitalisation, very large numbers of workplaces could be created within a fairly short time; and the creation of such workplaces would be 'within reach' for the more enterprising minority within the district, not only in financial terms but also in terms of their education, aptitude, organising skill and so forth.

Such intermediate technology has the added advantage that it does not need to be sited in urban areas, so that its adoption may do something to stem the flow of people from the countryside to towns. Neither does it require foreign experts since its principles can easily be understood by the worker using the technology.

These are the main ideas within the covers of Schumacher's book. We cannot consider all of them, as this would soon become tediously repetitive, but it will be useful to look at a number of Schumacher's arguments from the point of view of the fallibilist theory of value proposed in the previous chapter. It is first necessary to state formally Schumacher's main thesis. This is that employing Buddhist economics leads to better decisions than using conventional economics.

Formally:

V_B (x)(y)(If x follows Buddhist economics and y follows conventional economics, then x pref y)

(a) The exhaustion of limited resources

Here Schumacher considers the consequences of following conventional and Buddhist economics with respect to the consumption of limited resources, especially the Earth's stock of fossil fuels. He argues that Buddhist economics is superior because it alone comes to terms with the finite nature of the Earth's material resources. It may help by first considering how V_B might be falsified. A defender of conventional economics might argue from the following premises.

F_1 $\exists x \exists y$ (x follows conventional economics and y follows Buddhist economics, and x consumes more resources than y)

V_1 (x)(y)(If x consumes more resources than y, then x pref y)

Clearly

$F_1 V_1 \longrightarrow \exists x \exists y$(x follows conventional economics and y follows Buddhist economics and x pref y)

i.e. $F_1 V_1 \longrightarrow \bar{V}_B$

Schumacher turns this potential falsification of his thesis V_B into a corroboration. He accepts F_1, so he must argue against V_1. This he does from the following premises

F_2 (x)(y)(If x follows conventional economics and y follows Buddhist economics and x consumes more resources than y, then x exhausts resources rapidly and y does not).

V_2 (x)(y)(If x exhausts resources rapidly and y does not, then y pref x).

Clearly,

$F_1 . F_2 . V_2 \longrightarrow \exists x \exists y$(x consumes more resources than y and y pref x)

i.e. $F_1 . F_2 . V_2 \longrightarrow \bar{V}_1$

If F_1 continues to be accepted and if V_2 belongs to the debates' background values, then if Schumacher can establish F_2, he will have succeeded in falsifying V_1, so rescuing his thesis from the falsification threatening it. Schumacher defends F_2, particularly concentrating on fossil fuels, in the opening chapter of his book and in chapter 2.3.

In this way, Schumacher corroborates V_B, but it must be observed that V_B is very imprecisely stated since nowhere does Schumacher give a clear characterization of Buddhist economics. This, of course, leads to a low degree of corroboration whenever V_B passes a test.

(b) Intermediate Technology

Again, Schumacher compares the consequences of adopting Buddhist and conventional economics in developing countries, arguing for his thesis V_B by trying to show that following the former has better consequences for a developing country than following the latter. As before, it is best to begin by considering how V_B might be falsified.

F_3 $\exists x \exists y$(x follows conventional economics for a developing country and y follows Buddhist economics, and x produces more wealth than y)

V_3 (x)(y)(If x produces more wealth in a developing country than y, then x pref y)

F_3 and V_3 form the premises of a counterargument to V_B, since

$F_3 . V_3 \longrightarrow \exists x \exists y$(x follows conventional economics for a developing country and y follows Buddhist economics, and x pref y)

i.e. $F_3 V_3 \longrightarrow \bar{V}_B$

Assuming that F_3 is accepted by both the followers of conventional and of Buddhist economics, Schumacher can rescue V_B only if he can falsify V_3. In doing so he uses the following premises;

F_4 (x)(y)(If x follows conventional economics for a developing country and y follows Buddhist economics and x produces more wealth than y, then x produces a greater gap between rich and poor than y).

V_4 (x)(y)(If x produces a greater gap between rich and poor in a developing country than y, then y pref x).

$F_3.F_4.V_4 \rightarrow \neg x \neg y$(x produces more wealth in a developing country than y and y pref x)

i.e. $F_3.F_4.V_4 \rightarrow \bar{V}_3$

If F_3 is accepted, and if V_4 belongs to the debate's background values, then if Schumacher can establish F_4, V_3 must be regarded as falsified. In this way he avoids falsification of V_B. Avoiding falsification in this way, of course, confers a certain degree of corroboration on V_B.

(c) Nuclear Power

The pattern of argument is exactly the same as before. V_B might be falsified using the following premises.

F_5 (x)(If x follows Buddhist economics then x does not involve the adoption of nuclear energy)

F_6 (y)(If y follows conventional economics then y does involve the adoption of nuclear energy)

F_7 Nuclear energy is the cheapest way of producing energy

V_5 (x)(If x is the cheapest way of producing a good, then x should be adopted)

$F_7V_5 \rightarrow$ Nuclear energy should be adopted

So

$F_5.F_6.F_7.V_5 \rightarrow \neg x \neg y$(x follows Buddhist economics and y follows conventional economics and y pref x)

i.e. $F_5.F_6.F_7.V_5 \rightarrow \bar{V}_B$

If F_5, F_6 and F_7 are agreed upon, then the threatened falsification of V_B can only be relieved by the falsification of V_5. This Schumacher attempts using the following premises

F_8 Nuclear energy produces more hazardous wastes than alternative sources of energy.

V_6 (x)(y)(If x produces more hazardous wastes than an alternative way of producing a good, then x should not be adopted).

$V_6.F_8 \longrightarrow$ Nuclear energy should not be adopted

i.e. $F_7.F_8.V_6 \longrightarrow \bar{V}_5$

If F_7 is accepted, and if V_6 belongs to the background values of the debate, then V_5 should be regarded as falsified if F_8 can be established. This is a point in which Schumacher invests considerable effort in chapter 2.4.

I have considered only three of Schumacher's many arguments for V_B, but the majority of those I have not considered are of the same form.

Case Study 3

The Removal of Lead from Petrol - U.S. Environmental Protection Agency vs the Ethyl Corporation

Motor vehicle exhausts have, for many years, been considered a serious nuisance and potential health hazard in the United States. Action was taken by Congress in 1970 to remedy these problems when a revision of the Clean Air Act was passed which placed severe limits on the gases causing the problems in vehicle exhaust. Under the amended Act, the Environmental Protection Agency (EPA) was given powers to control or prohibit the sale of any fuel which will endanger the public health or welfare or which would impair the performance of devices fitted to exhaust systems to control the level of noxious gases in accordance with the new limits. On February 23rd 1972, the EPA proposed a complex set of regulations which would require a phased reduction in the amount of lead added to petrol to 1.25 grms per gallon in 1977 and the general availability of a 91 octane (strictly a 91 Research Octane Number or RON) petrol free of lead by July 1st 1974 (37 Federal Register 3882-84 (1972)). The EPA's case was a double one. At the time no adequate pollution control device was available which could be attached to vehicles to reduce the level of noxious gases in their exhaust to below the new limits, but the most promising type of device employed a catalyst which would be unable to operate if lead was in the exhaust. For this reason EPA sought to ensure the availability of a lead free fuel by 1st July 1974, from when new cars would have to meet the strict pollution requirements. The EPA also used their second authority to propose a reduction in the amount of lead added to petrol on the grounds that this constituted a public health hazard. They supported this claim with the document Health Hazards of Lead.

A major manufacturer of lead additives, The Ethyl Corporation, strongly resisted the EPA's attempt to gain approval for its proposed

regulations. This case study traces some of the threads running through the long and complex debate between EPA and Ethyl. Although highly technical in many places, the debate offers a useful example to which to apply the theory of value developed in the previous chapter. It is very well documented and conducted by parties equipped with great resources, acumen and specialist knowledge. We might, therefore, expect to find in it examples of many of the points illuminated by the fallibilist theory of value. Chief of these is the point that the debate, being about what course of action should be taken is a debate about values, and yet values play no explicit part whatever in the debate. The debate between Ethyl and EPA concerns facts and facts alone.

The traditional view of value would see the present debate in something like the following way. Both EPA and Ethyl agree on the facts about the consequences of reducing lead in petrol and those of maintaining existing levels, or at worst there are a number of facts which they have not yet agreed upon but which can be decided, one way or the other, without too much difficulty. What makes the argument between Ethyl and EPA so persistent is that each is committed to quite a different set of values. The EPA is charged with maintaining the health of the general public and a decent environment, and so places great value on clean and healthy air. Being a government agency it places little value on the profits of big business. Ethyl, on the other hand, is a private business and so must make money to survive. It is not surprising to find that it places a lower value on clean air than it does on its own profits. The conflict arises from the clash of these two, fundamentally irreconcilable, systems of value.

When we look at the details of the debate, however, this view of it appears a travesty. The debate between Ethyl and EPA concerns factual questions and factual questions alone. No discussion of values ever takes place in the whole debate. Values are needed to make any decision but they do not arise in the debate because the values employed by both sides are

mundane 'background' values shared by all. What EPA tries to show is that there are facts which, when coupled with some set of background values, entail that it is better to remove lead from petrol than maintain existing levels. Ethyl, on the other hand, tries to show that there are background values which, coupled with facts lead to the reverse conclusion. The debate then centres, not on the values used by the parties, because these are common, but on the facts elicited by each side in support of their case.

In any such debate, it is of the greatest importance to separate what the parties said from why they said it. In the case of Ethyl it is quite obvious what motivated them to attack the regulations proposed by EPA, for they faced ruin should they be put into effect. But this provides motivation only. The anguish of Ethyl less they be ruined is not part of the critical debate, no more than EPA's angelical urge to trounce big business in the protection of the environment. Each side is motivated to argue its case as well as it can, but it is the argued case, and not the motivation which is important, at least for our present purposes.

Round 1

The debate between Ethyl Corporation and the EPA takes the form of a number of rounds, each side modifying its position in the light of the earlier round. The EPA open the first round with their publication of the proposed regulations to make lead free fuels available and to gradually reduce the amount of lead added to normal petrol. Their first argument is that lead free fuel must be available from 1st June 1974, when the new emission regulations come into force, so that catalytic emission control devices can be used.[11]

> The Administrator (of the EPA) has determined that emission products of lead additives will impair to a significant degree the performance of emission control systems that include catalytic

converters which motor vehicle manufacturers are developing to meet 1975-6 motor vehicle emission standards and are likely to be in general use if lead additives are controlled or prohibited for use in certain motor vehicle gasolines.

EPA's argument may be captured formally in something like the following way.

V_1 The Federal Regulations on vehicle emissions ought to be met

F_1 The Federal Regulations can only be met by catalytic systems by 1975

F_2 Catalytic systems can only work if lead free gasoline is available by 1975

V_2 Lead free gasoline ought to be available by 1975

Ethyl replied to this argument by attacking F_1, the claim that only control systems employing catalytic converters would be working by 1975. It argued that it was unlikely that any workable control system would exist by 1975, so that implementation of the new standards would have to be postponed, involving a postponment of the regulations governing the availability of lead free gasoline. Ethyl also argued that control systems which do not rely upon a catalyst might be found preferable to catalytic systems, and if these were used there would be no reason to insist that lead free fuel be available to the motorist.[12]

Ethyl also attacked V_1, but not by simply denying it and substituting some other value statement of their free choosing. V_1 was criticised by appealing to facts. Ethyl argued that meeting the new emission standards would be very expensive and that using thermal reactors and not catalytic converters in vehicle exhaust systems would lead to slightly higher emissions

than those allowed under the regulations, but would save a great deal of money. We might formalize their case as:

V_3 If X produces a marginal improvement in the environment at great cost then X ought not be done

F_3 Meeting Federal emission Regulations will produce a marginal improvement in the environment at great cost

---

V_4 The Federal emission Regulations ought not be met

V_3 is a background value shared by both parties to the dispute and yet when coupled with F_3 it yields V_4 which contradicts V_1, assumed by the EPA to be a background value. If the EPA accept F_3, then they must reject at least one of V_3 or V_1. The conjunction $V_3.V_1$ is shown to be factually incorrect. If V_1 is rejected, then EPA's original argument collapses and Ethyl have won the argument. If the EPA wishes to counter this challenge, it should be obvious that this is best done by arguing against F_3. Thus the debate, although involving values, always centres on factual issues.

The second argument of the EPA was that lead from vehicle exhausts is a health hazard, so that the amount allowed in gasoline should be decreased. Very briefly, we may see this as:

V_5 If X eliminates a health hazard then X ought to be done

F_4 Reducing the amount of lead in gasoline would eliminate a health hazard

---

V_6 The amount of lead in gasoline ought to be reduced

In support of F_4, EPA published the document Health Hazards of Lead (April 1972) largely based on a specially prepared report from the National Academy of Sciences, Airborne Lead in Perspective (1971). This concluded that:

F_5 Since lead has not been shown to have any biologically useful function in the body any increase in body burden of lead is accompanied by an increased risk of human health impairment.

F_6 In many cities air lead concentrations are slowly rising.

F_7 Human blood lead levels begin to rise appreciably with an exposure to airborne lead concentrations in excess of 2 micrograms per cubic meter.

F_8 Elevated lead intake for periods as short as 3 months produces an increase in blood lead levels.

F_9 Body burdens of lead increase with age, at least to 40 years and probably thereafter.

F_{10} Although the ingestion of leaded paint is the predominant cause of lead poisoning in children, some children may show high blood lead levels from the ingestion of dust contaminated by fallout from airborne lead.

F_{11} Average blood lead levels tend to be higher among urban residents than among rural residents and higher among groups occupationally exposed to vehicle exhaust (e.g. policemen and garage workmen).

The Administrator of the EPA, therefore, recommended that:

... airborne lead levels exceeding 2 micrograms per cubic meter, averaged over a period of 3 months or longer, are associated with a sufficient risk of adverse physiologic effects to constitute

endangerment of public health. Since airborne lead levels in many major urban areas currently range from 2 to somewhat over 5 micrograms per cubic meter, and since motor vehicles are the predominant source of airborne lead in such areas, attainment of a 2.0 microgram level will require a 60-65% reduction in lead emissions from motor vehicles.

Ethyl's reply to the case made out by the EPA on health grounds was very extensive,[13] but concentrated entirely on F_4, V_5 being accepted by both parties. Ethyl claimed that existing levels of lead in the air do not constitute a health risk, so that F_4 is false. In arguing this they proposed the following counters to the claims made by the EPA.

Against F_5: Years of experience with occupationally exposed groups show that blood lead levels well in excess of those found in the normally exposed population to be perfectly safe.

Against F_6: The evidence indicates that air lead concentrations in many cities are falling. US blood leads are of the same order as those for many non-industrialised populations, indicating that lead from industrial sources makes only a small contribution to blood lead levels.

Against F_7: The data used in the calculation of this 2.0 microgram per cubic meter limit is seriously suspect, as is the statistical device used in the calculation. More reliable data (the so-called 7 City Study) shows no correlation between air lead levels and blood lead levels. In addition, the EPA assumed that about 30% of lead inhaled is retained in the lung. The true figure is nearer 10%.

Against F_8: This may be the case, but there is no evidence to indicate that the 'excess' blood lead levels resulting from exposure to airborne lead are a health hazard.

Against F_9: The data on body burdens shows that lead body burdens do not increase with age. Even if they do, this would merely reflect the very long time (about 30 years) needed for the body to come into equilibrium with environmental lead.

Against F_{10}: There is no known correlation between lead levels in dust and earth and blood lead levels of children exposed to the dust and earth. There is no evidence whatever for EPA's hypothesis about dust being a significant contributor to the blood lead of children. The rate of lead fallout is so low that this can only be an insignificant source of lead.

Against F_{11}: The very large 7 City study reveals no correlation between air lead levels and blood lead levels.

In addition to offering these counters to the case made out by the EPA, Ethyl pointed to the economic and environmental cost of the EPA's proposal, arguing that these had been grossly underestimated by the EPA. The economic penalty would include, according to the Ethyl Corporation;

F_{12} 5% more crude oil consumption due to using less efficient engines. This could put $1.4 billion on the balance of payment deficit.

F_{13} The cost to motorists will be about ¢4.7 per gallon because of higher gasoline costs and lower engine efficiency.

F_{14} Extra refinery investment to meet the need for low lead gasoline will amount to more than $4 billion. Small refineries will be unable to make the large investment necessary and will close.

If lead is not added to gasoline, the only way in which the intended octane number can be achieved is by adjusting the chemical mix of the product. This is done by altering the refining process (hence the refinery costs above) to give a gasoline containing a higher concentration of aromatic hydrocarbons. This would, according to Ethyl, constitute an environmental hazard since:

F_{15} Emission of polynuclear aromatic (PNA) hydrocarbons would increase, and many of these are suspected of causing human cancer.

F_{16} Emission of those hydrocarbons responsible for the formation of photo-chemical smog and eye-irritant chemicals would increase.

F_{17} EPA dismiss these problems as they expect these hazardous chemicals to be removed by the catalytic emission control systems to be fitted to new cars, but no working system yet exists. In addition, it is more difficult to work such a system when using fuel containing a high percentage of aromatic hydrocarbons.

We may formalise this part of Ethyl's case in the following way;

V_7 If X constitutes, on balance and at best, only a marginal improvement in human health at a great cost then X ought not be done.

F_{18} Reduction of lead levels in gasoline constitutes, on balance at best, only a marginal improvement in human health.

F_{19} Reduction of lead levels in gasoline is very expensive

---

V_8 Reduction of lead levels in gasoline ought not be done.

The force of the argument is as before. V_7 is a background value shared by both parties to the dispute. What Ethyl have done, therefore, is find facts, F_{18} and F_{19}, which, when coupled with some background value yield the evaluative consequence they want. If the factual claims are accepted by the EPA then they have no option but to revise their regulations calling for a reduction in the amount of lead added to gasoline. Returning to EPA's original argument ($V_5.F_4$ hence V_6), if they accept Ethyl's factual claims, then they must reject V_6 and hence at least one of V_5 and F_4 - presumably F_4. Formally, the conjunction $V_6.V_7$. is falsified by $F_4.F_{18}.F_{19}$. Ethyl have argued that F_{18} and F_{19} are true, so that their opponent can either accept $V_6.V_7$ and reject F_4 or else accept F_4 and reject either of V_6 or V_7. Since both V_6 and V_7 are background values, it seems that the rejection of F_4 is the preferred move, in which case EPA's original argument is bankrupted.

Ethyl's alternative to the EPA's regulations consists of ensuring that lead free gasoline is available from when catalytic systems are known to work. New vehicles, from then on, would use lead free fuel and so there would be a gradual elimination of leaded fuel. If some other emission control system is found preferable to the catalytic systems, then if it is still thought desirable to remove lead from the air, this can be done very cheaply by using existing fuels but fitting new cars with lead traps at present under development. Thus ended round one of the contest.

Round Two

Following the criticisms of the Ethyl Corporation and others, the EPA revised their regulations. Previously, EPA's case for both parts of their proposed regulations, those governing the availability of lead free gasoline and the phased reduction of lead concentrations in normal gasoline, were based on considerations of human health. In their revision the EPA now based the lead free gasoline regulations solely on the need to have catalytic emission control devices and the regulations for the reduction of lead additives in gasoline on health considerations. The revised regulations were published on January 10th 1973 (Federal Register 1254-61), the health issues being published in a new document, EPA's Position on the Health Effects of Airborne Lead, (March 9th 1973). These two documents provided the main elements of round two of the debate.

The EPA dropped from their argument F_6 - F_9, the most important of which was F_7, the claim that blood lead levels rise on exposure to air containing more than 2 micrograms of lead per cubic meter. This was savagely attacked by Ethyl in the previous round and plays no more part in the debate. The EPA, however, reiterate F_5, that lead has no known biological function so that any increase in body burdens increases the risk of human health impairment, F_{10}, the claim that airborne lead fallout in dust may be a significant route of lead exposure in children; and F_{11}, the claim that blood lead levels tend to be higher in urban residents than country residents, a claim which is now supported with evidence from the very large 7 City Survey. The following claims enter the debate for the first time:

F_{20} Many city dwellers have abnormally high blood lead levels.

F_{21} The susceptibility of children may be greater than adults so that children may be suffering subtle but unrecognised neurological impairment due to lead.

F_{22} Newborn babies in cities have higher blood lead levels than newborn babies in rural areas.

F_{23} Chromosomal damage due to lead is possible.

V_9 Presently recognised blood lead limits are too high to protect the public.

Upper acceptable limits for blood lead for the following groups are:

Expectant mothers	- 30 micrograms per 100 ml blood
Newborn babies	- 30
Children	- 40 (and perhaps less)
Adults	- 40

In support of F_{21} the EPA cited the work of David who compared the blood lead levels of children who were hyperactive with no known cause, with blood lead levels of a control group. He found that the hyperactive children had significantly higher blood lead levels than the controls. In support of V_9, the EPA referred to a recent study of umbilical cord lead levels.

Before considering the response made by the Ethyl Corporation, something must be said about the status of V_9. I have taken the statement as evaluative because it states that certain blood lead levels are the greatest which we ought to accept. EPA's argument for this involves appeal to a background value V_{10} in the following way:

V_{10} The upper acceptable limit for a toxin in the body is the lowest level at which the health of someone in the population is impaired.

F_{24} The lowest blood lead level at which the health of some expectant mother (newborn child, child, adult) is impaired is 30(30,40,40) micrograms per 100 ml.

V_9 The upper acceptable limit for lead for expectant mothers (newborn children, children, adults) is 30(30,40,40) micrograms per 100 ml.

V_9 forms the most important part of the EPA's case and so it is not surprising to find that it comes under severe fire from the Ethyl Corporation. As before, the criticism centres not on the evaluative issues but on factual claims. Ethyl apparently accepts V_{10}, which may, therefore, be regarded as a background value, but it seeks to deny F_{24}, a move which would, of course, destroy EPA's argument for V_9. Ethyl denied that there was any medical evidence supporting F_{24}. The only evidence cited by the EPA concerned the study of umbilical cord blood levels, but this found no urban-rural gradient and concluded that there was no evidence to implicate airborne lead as a contributor to high cord blood lead levels. Ethyl accused the EPA of making an ad hoc move in so redefining upper acceptable blood lead levels without supporting evidence. They also pointed out that the Surgeon General regarded children with a blood lead level of less than 50 micrograms per 100 ml as safe provided there is no evidence of continuing high lead exposure.

The Ethyl Corporation also countered the other parts of the case made out by the EPA.[14]

Against F_{20}: The upper level for city dwellers' blood lead is around 40 micrograms per 100 ml which cannot be said to be 'abnormal'. Many higher values turn out to be the result of faulty analysis. For children, high blood leads are solely due to exposure to leaded paints.

Against F_{21}: The major evidence for this is the work of David referred to earlier. Other investigations have failed to discover the same effect. David's results were due to the higher incidence of pica (the habitual eating of curious substances, often including lead paint) in hyperactive children.

Against F_{22} This claim is in direct contradiction to the paper cited by the EPA.

Against F_{23} The evidence for such chromosal damage is extremely speculative.

Ethyl also expanded on some of the points in round one of the debate. An extensive survey of lead poisoning in children failed to find a single case implicating dust as the source of lead. 98% of cases were due to the eating of lead paint. Hence, EPA's claim that dust contaminated by fallout of airborne lead might be a significant source of lead exposure in children$^{(F_{10})}$ is thrown in serious doubt. Ethyl also pointed to animal experiments indicating that lead might, after all, be an essential trace element, against EPA's claim F_5.

Finally, Ethyl argued that, on EPA's own admission, the removal of lead paint from delapidated, aging housing would be considerably cheaper than the reduction of lead levels in gasoline. Such a programme of renovation would prevent the vast majority of existing cases of overt lead poisoning in children. We may formalize their case here in the following way.

V_{11} If X improves the health of the population less than Y and if X is more expensive than Y then Y pref X.

F_{25} Removing lead from gasoline will improve the health of the population less than removing lead paint from old buildings.

F_{26} Removing lead from gasoline will be more expensive than removing lead paint from old buildings.

V_{12} Removing lead from old buildings is preferable to removing lead from gasoline.

V_{11} is suggested as a background value, acceptable to Ethyl's opponent. If Ethyl can then show that F_{25} and F_{26} are true, then EPA must accept V_{12} and so give up V_6. To avoid this, EPA would, of course, attempt to counter one or both of these factual claims, so that, once again, the factual element of the debate is to the fore.

Round Three

EPA again revised their position on the health aspects of reducing the amount of lead added to gasoline in their final document, EPA's Position on the Health Implications of Airborne Lead, (28th November 1973) which was used to support their revised and final regulations of 6th December 1973. Unlike their earlier documents, this one was not open to public comment, and so Ethyl was not allowed a rejoinder. No doubt the EPA thought that the argument had gone on long enough. Despite this, however, the final document contained some major changes to the earlier one, several of them apparently the result of Ethyl's criticism.

1. It was admitted that there is no evidence suggesting that children are more susceptible to lead than adults. The special position of children is now based on their higher exposure from paint, dust and dirt.
2. The earlier recommendations for upper acceptable blood lead levels for expectant mothers and newborn children are dropped. Instead, limits for all are put at 40 micrograms per 100 ml of blood.
3. EPA argues against Ethyl's claim that there is no correlation between air lead concentrations and blood lead levels.

The EPA also expand considerably on two of their earlier claims, that low levels of lead may cause subtle 'subclinical' neurological changes in children, and that contaminated dust is a major source of child exposure to

the metal. They concluded that in each case the evidence was not conclusive, but taken together pointed to the correctness of their earlier claims. Quite new was a calculation of the likely increase in blood lead of a 'standard man' exposed to various air lead concentrations. These purported to show that to keep below the recommended blood lead of 40 micrograms per 100 ml, the ambient air lead concentration should be below 11.8 micrograms per cubic meter on optimistic assumptions and below 4.0 micrograms per cubic meter on pessimistic ones.

Round Four

Round four takes the form of a court case. On the day that EPA promulgated their regulations, 6th December 1973, Ethyl petitioned the Court of Appeals to have them put aside. The case is of little value for our present purposes and a discussion of it would take us rather deeply into the fine points of law and precedence. Ethyl's submission to the Court, however, contains many criticisms of EPA's final health document, in particular on the sub-clinical effects of lead, on the correlation between air lead and blood lead levels and on the claim that dust containing lead is a hazard to children. Whilst of considerable interest in themselves, the arguments will add little illumination to our central concern - the nature of debates about evaluation.

On January 28th 1975 the Court of Appeals gave a majority finding in favour of Ethyl. The EPA then petitioned the Court for a rehearing which was granted and opened in May 1975. The case was a very interesting one, but to consider it in any detail would take us far from the point. Suffice it to say that the EPA won their case and that their final draft regulations were eventually approved and are now in force.

Chapter 7 - Footnotes

1. K. Popper, On The Theory of the Objective Mind, in Objective Knowledge, Oxford University Press, 1972, 153-190.
2. K. Popper, Epistemology Without a Knowing Subject, in op. cit., 106-152.
3. See, for example, B. Russell, An Inquiry Into Meaning and Truth, Penguin, 1962, 127-129.
4. Hansard, 29th April 1977, 1736-1798.
5. E. Schumacher, Small is Beautiful, Blond and Briggs, 1973.
6. op. cit., pp. 38-40.
7. op. cit., p. 244.
8. op. cit., p. 97.
9. op. cit., p. 50.
10. op. cit., p. 167.
11. U.S. Federal Register, 37, No. 36, 23rd February 1972.
12. Catalytic exhaust controls are now fitted to all new U.S. cars.
13. Comments on EPA's Proposed Lead Regulations, Ethyl Corporation, 1972.
14. Critique of E.P.A.'s Position on the Health Effects of Lead, Ethyl Corporation, 1973.

PART III

CRITICISM AND DECISION

CHAPTER 8 - CONTEMPORARY VIEWS OF DECISION MAKING I

INDIVIDUAL AND SOCIAL VALUES

In this and the following chapter I shall consider the limitations of contemporary theories of how decisions ought to be taken. In this, I shall avoid all but incidental discussion of theories of how decisions are actually taken, though we will find the distinction between normative and descriptive theories of decision making less clear than might at first be imagined. When I speak of contemporary decision theories I mean to embrace all theories which come under the heading of welfare economics or of Bayesian decision theory. Welfare economics is clearly marked off from the rest of economics by its inclusion of normative principles. Its basic concern is the desirability of different economic states, where these may consist of different sets of consumables variously distributed to a group of consumers, or a fixed quantity of goods distributed in different ways. For the purposes of the discussion, I shall include cost benefit analysis as part of welfare economics. The central postulate of Bayesian decision theory is that in making decisions under risk, a rational agent should seek to maximize his expected utility.[1] It has undergone very great development in recent years, and has been articulated to provide a theory of decision making under uncertainty (for the distinction between risk and uncertainty, see below); a theory of games;[2] a theory of group decisions[3] and even the beginnings of a political theory.[4]

What unites theories of both camps, however, is their basic intention of showing how decisions may be justified. This is directly counter the sceptical conclusions of Part I. There is was argued that no value judgement can be justified. To reach a similarly sceptical conclusion about making decisions, all that is necessary is to observe that making a decision is a special kind of value judgement. Since no value judgement is justifiable, no decision can be justified.

If part I leads to scepticism about justifying decisions, part II suggests

an answer to the problem of how decisions may be made in a rational way even though they cannot be justified. In short, a fallibilist theory of decision making is suggested. Since decisions are value judgements, it should be possible to develop a fallibilist theory of decision making from the general theory of value described in the last two chapters. This approach to the making of decisions accepts the first sceptical thesis below, but seeks to deny the second:

1. No decision can be justified.

2. No reasons can be given for favouring one decision over another.

It is my intention to develop a theory of decision making along these lines in chapter 10. Before such a theory can be appreciated, however, it must be compared with the rivals which it seeks to replace, and the aim of this and the following chapter is to point to the central shortcomings of contemporary decision theories of the kinds mentioned. This will also help to bolster the sceptical argument about decision making stated above which, although having the merit of brevity, does nothing by itself to point the way to a better theory of how decisions should be taken. I have divided my criticism of contemporary decision theories into two parts, one dealing with values and the other with facts. In this chapter my concern is to argue that all contemporary theories incorporate a false view of individual values, and that none of them provides a satisfactory account of how individual values can be transformed into social values needed for making public decisions. In the next chapter, I shall argue that contemporary theories required such a large quantity of factual information before a decision can be assessed, that they can never hope to be of assistance in the making of any but the most trivial decisions. Chapter II will consider how a fallibilist theory of decision making can overcome these difficulties.

1 The Determination of Individual Values

All contemporary theories of decision making assume the traditional view of value, according to which a person has privileged access to his own values. On this view, if a person says that he attributes such and such a value to some item,

then (if he is not drunk, lying or misusing language etc.) such and such is the value which the item has for him. Many examples of philosophical theories of value which incorporate the doctrine were discussed in chapter 2, but the doctrine receives virtually no explicit mention in the not inconsiderable literature on decision making. It is, therefore, an assumption of the very worst sort; one which is not only uncriticised, but unrecognised. The function of privileged access, of course, is to show how value judgements of a particular kind may be justified. Any view which incorporates the doctrine of privileged access, therefore, falls to the critical conclusions of Part I, where general arguments against justificationist views of value were deployed. In Part 2, I have argued that any evaluation can only be a guess which should be open to criticism, and have tried to characterise the nature of this criticism. It follows that all contemporary theories of decision making embody a false view of individual values. I shall illustrate the point by first looking at welfare economics.

Basic to welfare economics is the claim that all consumers can list all possible bundles of commodities in order of preference, which requires a postulate of rationality, a typical version of which holds that for all consumers:

1) For all possible pairs of alternatives A and B, the consumer knows whether he prefers A to B, B to A or is indifferent to them.

2) One and only one of the three possibilities is true for any pair of alternatives.

3) If the consumer prefers A to B and B to C, then he prefers A to C.

Before we can assess the reasonableness of these three conditions, we need to be clear about their status. Are the conditions supposed to be descriptive of the behaviour of real consumers, or are they about how an intelligent consumer should behave? The conventional answer is that the conditions are descriptive, but the development of welfare economics can only encourage doubt on this point in the mind of the external observer. If the three conditions are really taken by welfare economists as descriptive of consumers' behaviour, it is suspicious that

what must surely be highly suspect empirical claims have been submitted to experimental test in such a desultory way.[5] This is even more remarkable considering the enormous weight of theory which is built upon the postulate of rationality.

It seems more appropriate to the actual spirit of welfare economics, whatever its letter, to interpret the postulate of rationality as a normative claim about how consumers should behave. As such, it has a lot more to commend it. Consider, for instance, a consumer Q whose preferences fail to satisfy the third, or transitivity, condition. Imagine that he prefers A to B, B to C and C to A and that he possesses A, B and C being the property of some trader. The trader may suggest a deal by which Q exchanges A for C, and since Q prefers C to A he will be prepared to pay the trader for the exchange. This leaves Q with C and the trader with B and A, but since Q prefers B to C, the trader may extract more money from Q by exchanging B for C. This leaves Q with B and the trader with A and C, but since Q prefers A to B he will pay the trader to exchange A for B. This restores the original position, Q having A and the trader having B and C, and so the game can continue until all of Q's resources are transferred to the trader. This cautionary tale is a good reason for consumers to make transitive preferences orderings, even if, in practice, some consumers, perhaps through ignorance, fail to manage this. Similar cautionary tales can be constructed to persuade consumers that they ought to follow the other two conditions of the rationality postulate as well.

A second reason for viewing the postulate as normative rather than descriptive is that an exactly analogous criterion of rationality is, as we shall see later, traditionally applied to social preferences. Here, however, there can be no pretence that the conditions of the criterion are descriptive of real social preferences and the criterion is widely admitted to be normative.

Our interest in how decisions <u>ought</u> to be made makes welfare economics of concern to us only if it is given the normative interpretation above. Since this seems an acceptable interpretation, we may proceed to ask if the three conditions

of the rationality postulate are reasonable requirements which ought to be placed on consumers' preferences. (2) and (3) seem eminently reasonable taken normatively, but (1) must be treated with scepticism. All that an individual can do is guess that he prefers one thing to another and submit his guesses to criticism if he thinks them important enough. Any claim to know that one alternative is better than another must be unfounded, especially a claim on the scale envisaged by (1). As was pointed out in chapter 6, choosing between two items of consumption involves comparing all relevant features of the two items and, in general, also involves trading one desirable feature with another. Error is, therefore, possible in two ways; some relevant features may be overlooked entirely, and recognised features may be traded in a way which reflection shows to be mistaken. The bland statement of (1) conceals both of these problems.

Having seen the assumptions underlying the postulate of rationality of welfare economics, it will come as no surprise to find that the techniques used for the determination of consumers' values in the offshoot of welfare economics known as cost benefit analysis are deeply suspect. One way in which this is done is by asking those who benefit from a project the maximum they would be willing to pay to receive their benefits, this being taken as a measure of the value of the benefits to the individuals. Those on whom the project imposes costs are asked what they would just accept to compensate them for their loss, this being taken as a measure of the negative value the individual places on this loss. In the famous cost benefit analysis undertaken on the siting of the third London airport[6] it was recognised that the noise from aircraft would lower the enjoyment people living near the airport obtained from their houses. The market price of the houses would fall because of the noise nuisance, but it was realised that the owner of a house might lose much more than this drop in market price. His loss was reckoned as the money he would just accept in compensation for the noise, which might be greater than the drop in market prices. People living near the airport were, therefore, asked the question 'suppose your house was wanted to form part of a large development scheme and the developer offered to buy it from you. What

price would be just high enough to compensate you for leaving this house and moving to another area?'

Such a question betrays the stupidity of the whole approach to determining individual values. In deciding how much to ask for his house an individual must consider a whole host of factors - the cost of removal, loss of neighbours, loss of garden and local amentites, the gain of new amenities and neighbours, travelling problems, school problems with the children, whether the cat will roam, and so on. How can he be sure that no factor is overlooked altogether, and how can he be confident about the trade off he makes between the various costs and benefits? And yet a once and for all answer is required of him. A sensible answer, of course, can only be given after a critical scrutiny of the many factors involved.

Some realise the shortcomings of using the willingness to pay or the compensation criterion, but express the pious hope that 'errors' in individual values will cancel out provided a sufficiently large number of people are questioned.[7] Actual studies cast doubt on this. The well known study of Ridker on the costs and benefits of reducing air pollution,[8] for example, came to the conclusion that people affected by foul air were willing to accept on average about $10 per year in compensation. Self's doubts seem well founded when he writes that the results of such questioning:[9]

.....are likely to be somewhat artificial. When somebody states a price he normally does so in the context of an exchange relationship which gives point and precision to his calculations; he has to estimate how much something is worth in terms of other claims upon his income. Abstract from these conditions and the replies will either be somewhat casual.....or else will be latently influenced by the possibility of a subsequent bargain.....

An alternative method of estimating individual values is by appealing to the costs which people incur to obtain a particular benefit. For example, a person may enjoy visiting a national park to view the scenery, and in doing so he incurs various costs such as buying petrol, depreciation on his car, using time which he could put to productive use and so on. Whilst these costs can be measured fairly easily, the benefit the visitor receives from his visit cannot be readily measured. It is, therefore, assumed that he is rational and that he views the benefit as at least equal to the costs incurred.[10] Again this is spurious since the person has not necessarily subjected the matter to any kind of critical scrutiny. If, for example, he is unaware of the depreciation of his vehicle, as many people are, a major factor in the decision has been altogether overlooked.

The same criticism applies to the use of von Neumann utility in decision making. Basic to this is a postulate of rationality which takes the form of a set of axioms which include the three claims of the postulate made by welfare economics.[11] If the axioms are followed, a decision maker's values can be determined by presenting him with a series of lotteries. The idea here is very simple. Suppose there are three choices to evaluate, A, B and C, of which A is best and C worst for a particular person. What we want to know is how close the value of B is to either extreme for this person. We present him with a choice between having B for certain and having a lottery ticket which gives him a chance P of winning A and a chance 1-P of getting C. P is then adjusted until he is indifferent between the two options, B for sure and the lottery ticket with $P=P^1$. If the person values B just below A, then P^1 will be large (if B is valued equally as highly as A, P^1 will be 1 - i.e. the person will be indifferent between having B for certain and A for certain). On the other hand, if B is valued

close to C, then P^1 will be small (if B is valued equally to C, then $P^1 = 0$ i.e. the person is indifferent between B with certainty and C with certainty). In this way the individual's evaluation of a set of possibilities can be made and his utility function constructed. Such a function is a powerful guide to action since it follows from the axioms of the rationality postulate that a decision maker should maximize his expected utility (the sum of the utility of the various outcomes multiplied by their probability).

As before, however, no attempt is made to get the decision maker to subject his judgements about his indifference between a certain outcome and a lottery between the best and worst outcomes to critical scrutiny. The utility approach embodies the same mistaken view about individual values as welfare economics.[12]

2 Social Values

So far we have considered an individual decision maker and the problems he has in deciding on his preferences. He cannot know what his preferences are, and so he must guess and submit his guesses to as severe a critical scrutiny as the situation demands. When he acts, he should base his actions on the best guess he has about his preferences, always bearing in mind that even this guess, be it ever so well corroborated, may later have to be revised. In the remainder of this chapter the problems connected with corporate decision making are to be considered. Our basic question here is 'when a group of individuals must ascribe values to various items, how can this ascription properly reflect the values of the individuals who compose the group?' This question is at its most forceful when a group of individuals has to make a decision. Until a decision is called for, the members of the group may have widely different values, but may happily tolerate diversity. Toleration, however, becomes logically impossible when a decision is required. The preferences of some in the group may

lead them to favour one course of action, while those with other preferences may favour another course of action. The problem is that, at most, only one of these courses may be taken. Someone, it appears, must win and someone lose. Any collective decision, therefore, brings to the fore the problem of how a group of individuals can rationally agree about a common set of social values.

Many attempts have been made to establish acceptable relationships between a society's values and the values of the individuals who compose the society, but all these attempts are built around the same central idea; that social values are, in some way or another, an aggregation of individual values. The reasons behind this are not hard to divine; of all possible social values only those are acceptable which manage to reflect, in some way or other, the private, individual values of society's members. The justification of social values is only possible if they can be shown to be a fair reflection of the values of individuals. The political sentiment behind this view is wholly commendable. What I wish to argue, however, is that the technical development of this sentiment by contemporary theories of decision making has no hope of fruition, so that an entirely different account of the relationship between individual and social values is needed. To be more definite, I shall try to show that any aggregation of values like that traditionally supposed involves arbitrary elements which cannot be eliminated.

The following discussion covers quite a lot of ground and so it will be helpful if all the points made can be directed to one, central problem. We may suppose that there exists some way of determining the utility enjoyed by individual members of society and that social utility (or social welfare) is a function of the utility of society's members. For our present purposes, no loss of generality will result from imagining that social utility is simply a weighted sum of individual utilities. All of

the following discussion may then be directed at the question: how can the weights in such a sum be assigned in a rational manner? To those becoming hardened by exposure to scepticism, it will come as no surprise that the discussion will reach a negative conclusion.

Again, it will be useful to begin with welfare economics, which studies the distribution of wealth between consumers in an attempt to evaluate the social desirability of various economic states, or arrangements of economic activity and resources. It might, therefore, be hoped that welfare economics can shed light on the problem of forging social from individual values. Modern welfare theory has given up all hope of directly comparing the enjoyment or utility which one consumer receives from his consumption with that received by another consumer. Instead, recourse is made to ordinal scales of utility. There are two reasons for this; the great difficulty which was found in making interpersonal comparisons of utility, and the realisation that the ability to make such comparisons would add only very little to the explanatory power of welfare economics.[13] It is now an orthodoxy of welfare economics that it needs only ordinal measures of utility. The idea of such a measure is simplicity itself. The number x is applied quite arbitrarily to the utility of an object X for consumer Q and numbers assigned to the utility of other objects such that if Z is preferred to X, then the number is greater than x; if X is preferred to Z, then the number is less than x; and if Q is indifferent between Z and X, then the number is x. In this way the order of the numbers reflects Q's utility; the absolute magnitude of the numbers, the difference between them and their ratios being of no significance. Surprise is often expressed about the progress which is possible in welfare economics using this simple ordinal measure.

In seeing how social values can be based upon individual ones, the great restriction imposed by the use of ordinal measures of utility is that

no interpersonal comparison of utility is possible. It is, therefore, impossible to make a comparison between the gains of the gainers and the losses of the losers. For this reason, a central idea in welfare economics is that of a Pareto improvement. If a change in economic state increases the utility of at least one person, but reduces the utility of nobody, then the change is said to be a Pareto improvement.[14] If the value judgement is made that the welfare of society is a function of the welfare of the individuals who compose the society, then the economic state after any Pareto improvement will be judged better than the earlier state - hence, of course, the term 'improvement'. It is very important to recognise that even at this early stage, welfare economics has to resort to value judgements in order to say anything about social values. The value judgement about the desirability of Pareto improvements may seem harmless enough, but there are imaginable conditions where it does not seem so acceptable; when, for example, the community contains some group who we wish to see punished such as a conquered nation or ordinary criminals. We might well think that a reduction in the welfare of these individuals makes a positive contribution to social welfare. Sen has also pointed to the potentially extremely illiberal consequences of making all possible Pareto improvements.[15]

Unfortunately, very little progress is possible if all we have is the idea of Pareto improvement. When all possible Pareto improvements have been effected, the economy is said to be Pareto optimal (or Pareto efficient). In such a state no individual can be made better off without somebody else being made worse off. If there were only one optimum state, the original value judgement about the desirability of Pareto improvements would enable us to isolate one best economic state which we could struggle to achieve. Unhappily, however, there are a vast number, and on some assumptions an infinte number, of Pareto optimal states. Each state

has a different allocation of economic resources to individuals, but once we have denied interpersonal comparison of utility, there is no way of selecting one allocation over others.[16] Moreover, a non-optimal state cannot even be judged inferior to all optimal states. It is essential, therefore, for welfare economics to provide some other, less universally satisfied, criterion for judging between economic states.

Before leaving the discussion of Pareto optimality, it will be useful to consider it in the light of the question posed above of how weights are to be attached to the utility of individuals in estimating social welfare. If it is assumed that social welfare is the weighted sum of individual utilities, then the principle of making all possible Pareto improvements can be restated as the principle of positive responsiveness. This says that weights must be attached in such a way as to make social welfare increase with an increase in the utility of any individual. Having made all possible Pareto improvements, optimal states can only be compared when we know what weights to apply. This, of course, necessarily takes us beyond Pareto.

There are two principal responses to this need to go beyond Pareto comparisons. One, originally championed by Kaldor, seeks to measure total social welfare independently of its distribution, and the other, proposed by Bergson involves a social welfare function.[17] We shall return to Kaldor's suggestion later. Bergson postulates the existence of a social welfare function which arranges all economic states in order of social value. It is generally imagined, for purposes of exposition, that the social welfare function is laid down by a person, call him Superman. For all pairs of economic states X and Y, Superman dictates that either X is better than Y, Y better than X or X and Y are indifferent, and his judgements are transitive so that a weak ordering of economic states is achieved. If desired, each state may be assigned an ordinal number just

as in the construction of an individual's utility function. It is customary, though by no means essential, to restrict Superman's judgements in various ways. In particular, Superman is expected to accept that an individual is the best judge of his own values, and that social welfare is solely a function of the utility of society's members. Moreover, social welfare increases as the utility of any individual increases (principle of positive responsiveness), so that any Pareto improvement leads to greater social welfare. The aim of economic policy is to arrive at the highest possible social welfare, but this obviously involves questions about the distribution of wealth, for the social welfare stemming from the utility of one pound's worth of consumption varies with income level. Thus, Superman is also required to define an ideal distribution of wealth among the individuals who make up the society. A popular objection to this approach is that, whatever its theoretical adequacy, it cannot possibly be applied as a guide to real decision making. In real decision making, Superman would be some decision maker with authority over the community, although Hypothetical Supermen might also be used to explore the consequences of various ethical views about the distribution of wealth. But, in practice, it is impossible to use this model of a single decision maker with a well defined objective.[18]

The problem is that welfare economics can only pretend that the construction of such a welfare function is an account of rational decision making if reasons can be given for accepting the relationships subsumed under the function. In the special case we are considering, for expository reasons, where social welfare is a weighted sum of individual utilities, this reduces to the problem of providing reasons for assigning particular weights to individuals or groups (most significantly, income groups). Any but the most innocent social change leaves some better off at the expense of others, and so involves a redistribution of utility.

If such a change is to be justified, then the reasons must be given for the gainers' advantage overbalancing the losers' losses, and this inevitably involves attaching weights to the utilities of the affected individuals. Even if all individuals are assigned the same weight, reasons are still required for such egalitarianism. The Bergsonian social welfare function may be a useful theoretical device for deriving various theorems of welfare economics, but reasons are necessary before any real decision can be made. The problem, therefore, is how weights may be attached to individual utilities for the estimation of social welfare in a way which is not arbitrary.

Providing an answer to this question takes us to the roots of social philosophy. Historically, there are two traditions which attempt to give an answer; utilitarianism,[19] according to which all are to count equally in the assessment of social welfare, and social contract theories[20] which see social welfare as founded upon some original contract between society's members. The traditional theory of utilitarianism, developed by Bentham and Mill was eventually abandoned, largely because of difficulties in making interpersonal comparisons of utility, whilst doubts about the binding nature of a hypothetical social contract led to the gradual abandonment of social contract theories. Both traditions have, however, made a startling recovery in recent years.

Harsanyi has restated utilitarianism, and we shall begin by considering his arguments.[21] Harsanyi imagines a person who has the choice of joining several societies. In order that the choice be fair, Harsanyi uses the well known device of assuming that the choser does not know what social position he will occupy in his chosen society. He does, however, know the utility shared by people who occupy each social position in each society.[22] According to Harsanyi, the choser in this situation is faced with a fairly straightforward decision under risk, so he must maximize

his expected utility. If a society has n social positions, his chance of finding himself in any one of them, given his existing knowledge, is $\frac{1}{n}$. If the utility to be acquired from the occupancy of the jth social position is Uj, then the choser's expected utility for this society is simply $\sum_j \frac{1}{n} U_j$. Thus the utility of a society for the choser is simply the arithmetic mean of the utilities of all individuals in the society. He will, of course, chose the society with the highest utility so defined.[23] Moreover, his choice will be objective in the sense that anyone else in the same position would arrive at the same evaluation of the various societies.

To return to our central question about how individual utilities are to be weighted in the calculation of social welfare, Harsanyi's argument leads to the assignment of equal weights to all (the actual weight is irrelevant and functions only as a proportionality constant). It is, of course, a central tenet of utilitarianism that what determines social welfare is only the total amount of utility or happiness in the society, and not its distribution, so that increasing the utility of one person in the society increases social welfare by exactly the same amount as conferring that utility on some other person. I have four criticisms to make of Harsanyi's argument.

(i) The first criticism concerns the determination of individual utilities by Harsanyi's hypothetical choser. The choser must determine the utility achieved by each individual in the society and, according to Harsanyi, this can be done by thinking himself into the life of each individual in turn; imagining that he has someone else's talents, tastes, liabilities, tasks, rewards, likes and dislikes. Harsanyi seems to see no difficulty here; the utility enjoyed by the other person apparently just emerges once this imaginative achievement has been accomplished.[24] Underlying this belief seems to be a disturbingly naive view of our knowledge of evaluation. We

are aware of our own values because we know our own tastes, likes and dislikes, tasks and so on, and this gives us privileged access to our values. To discover another's values, all that is necessary is to imaginatively acquire his tastes etc., and we will have the same privileged access as before, but this time to his values. This is, of course, spurious. My estimation of another's values can, at best, be an intelligent guess, although one which I might be able to improve with greater knowledge. Things are not as simple as Harsanyi thinks.

(ii) The second problem for Harsanyi's account concerns the comparison of the utilities of different people. Harsanyi is quite correct when he says that we make such comparisons constantly, but the question is whether they can be made on the scale and with the precision required by Harsanyi.[25] The chief problem here is that even if cardinal measures for utility are used, such as those of von Neumann and Morgenstern, the measure is unique only up to a positive linear transform. The utility function of a person is, therefore, not unique, but can be represented by an infinite set of functions. For an individual decision maker this is no problem since all functions lead to the same decision, but if interpersonal comparisons are to be made, there must be a way of selecting just one function from the infinite set of each individual. When the utility functions of all the individuals in the society have been scaled in this way, they may be compared.

Vickrey has explored ways of scaling utilities.[26] The zero point is not needed, since what is being measured are the differences in the utility of individuals. All that is required, therefore, is a suitable interval. Vickrey suggests finding two social states A and B such that everyone prefers B to A, and employing those scales which assign a utility of zero to A and 100 to B. It is easily seen, however, that this leads to contradiction. Imagine that 1 and 2 form a two person society and that

their preference for four possible social states is as follows:

1 B pref D pref C pref A

2 D pref B indiff C pref A

where 'pref' and 'indiff' stand for strongly preferred and indifferent. This suggests two ways of scaling the individual utilities since both people prefer B to A and D to C. Scale S will use the first pair and scale S^1 the second, in the manner proposed by Vickrey. Suppose that the utilities of 1 and 2, as measured by, say, a von Neumann-Morgenstern scale are as follows; then the scaled utilities are as shown.

		UNSCALED UTILITY	SCALE S	SCALE S1
1	A	40	0	-100
	B	80	100	300
	C	50	25	0
	D	60	50	100
2	A	10	0	-106
	B	20	100	0
	C	20	100	0
	D	30	200	100

If we know aggregate utility in the way proposed by Harsanyi, scale S tells us to select social state D, with an aggregate utility of 250, whilst scale S^1 tells us to chose B, with an aggregate utility of 300. The scaling proposed by Vickrey will, therefore, be disastrously affected by the choice of calibration points. Indeed, any way of scaling will suffer from the same defect.[27] The problem might disappear if some other way of measuring utility were used, but none of these seem to hold out any promise.[28]

(iii) The third problem facing Harsanyi is the demands his method of measuring social welfare makes on factual information. Harsanyi seems to think that his account can be used for relatively small problems and not just those affecting the structure of the society, such as tax rates, rates

of saving and so on (this will be seen to be important later when we discuss Rawls). But this is to overlook the vast amount of information his method requires. First of all, representatives of each social position must be chosen and their utilities measured. But this only gives their present utilities. Any estimation of social welfare must obviously take into account the future levels of utility likely to be enjoyed by the society's members, and so this must be estimated in addition. Harsanyi's assumption of equiprobability for all social positions is obviously a considerable simplification, and in using his method to chose between two social states better estimates of these probabilities would be required. In the light of these problems it seems a little rash to suggest that the method might be used in real world decision taking.

Harsanyi does, however, suggest a remedy when the above information is not available. He tells us to treat social welfare as a weighted sum of individual utilities, the weights being arbitrary and the utilities being our best guesses. This is, however, a counsel of despair, for there is no reason at all to suppose that decisions made in this way will lead to genuine improvements in social welfare, any more than decisions made in a completely intuitive way.

Problems of information become pressing when one tries to apply Harsanyi's utilitarianism to the problem of setting an optimal tax rate. Here the theory is under two opposing influences. Decreasing marginal returns on consumption imply that a shift of consumption from the wealthy to the poor will increase total welfare, but the loss of incentives for the rich to produce results in a decline of total consumption. There is, therefore, a tension between justice and efficiency, a tension which has been known for many years.[29] Harsanyi tells us nothing about the resolution of this tension, and this is due to ignorance about the effects of incentives

on total production and the sharpness of the decline in marginal utility. Without a great deal of additional knowledge, Harsanyi's theory is impotent here.

(iv) Several commentators have criticised Harsanyi on the grounds that the estimation of social welfare he proposes reflects the attitude to risk of the original choser.[30] If the choser likes taking risks, then he will give greater weight to those individuals with high utility than those with low, and vice versa if he is adverse to taking risks. In each case his estimate of social welfare will be different, and the difference may be sufficiently large to alter his choices of social state. Harsanyi assumes that his choser is risk neutral, so that he weights all utilities equally. In reply, Harsanyi has argued that attitudes to risk are a perfectly proper component of social welfare, any social welfare function reflecting the balance between acquiring high utility and the risk of ending up with a low level of satisfaction.[31]

Despite Harsanyi's defence, the influence of attitudes to risk is enough to demolish his version of utilitarianism. This occurs in two ways. Firstly, Harsanyi supposes that his choser's estimate of social welfare is objective because anyone in the same position of ignorance would make the same estimate. Everyone will, therefore, agree on the ranking of social states and the final choice of one of them. But this will only be the case if everyone shares the same attitude to risk. This, of course, is a quite unrealistic possibility, and, in general, different people have markedly different attitudes to risk. If this is so, then as indicated above, there will be no unanimity about estimates of social welfare or the ranking of social states, or even the selection of the best social state.[32] There may be as many different opinions about what society to chose as there are individuals in the society, and so Harsanyi has the unenviable task of aggregating these opinions in a rational way.

To see the force of the second problem associated with attitudes to risk, the unifying problem posed earlier may be recalled; how are individual utilities to be weighted in a rational way in the estimation of social welfare? The standard view of attitudes to risk is that they are a subjective, personal quirk of the individual decision maker, and are, therefore, not open to criticism. If a decision maker likes taking risks, nobody can argue that he is wrong, nor can his decisions be defended by anything but his own personal tastes for risk taking. This means that anyone's estimation of social welfare using Harsanyi's method, depending as it does on his particular attitudes to risk, will be arbitrary. Harsanyi, therefore, fails to provide a rational, objective way in which individual utilities may be weighted to arrive at social welfare.

After this long discussion of Harsanyi's version of utilitarianism, we may now ask whether Rawls' social contract theory provides a more satisfying answer to our original question about the way in which individual utilities are to be weighted in calculating social welfare.[33] Like Harsanyi, Rawls exploits the idea of a choice of society behind a veil of ignorance, but whereas Harsanyi considered one person making such a choice, Rawls thinks of a group of agents who must agree among themselves on what society to adopt. The function of the assumption is, however, exactly the same in both cases. Social arrangements are unfair because people can exploit inequalities in the society, so that a hypothetical fair society can be constructed by imagining away such inequalities. In any society principles of social justice are required to determine the division of wealth and position, and so to resolve the unavoidable conflicts of interest which arise under this head. Rawls, therefore, tries to imagine what principles of social justice would be agreed upon by a group unable to exploit personal inequalities

because these are hidden behind the veil of ignorance. Such principles will lay down a fair structure for society, covering its major institutions and the distribution of rights and duties which they govern.

In Rawls' 'original position' a group of rational, and mutually disinterested agents must chose principles of social justice to govern a society in which they will live, although they are all in ignorance of the place they will finally occupy in the society, and of their own natural assets, abilities and disabilities, and those of their fellows. Rawls argues that they will chose two principles of social justice. The first of these is relatively straightforward, and states that each person has an equal right to basic liberties compatible with the same right to others. The second principle requires the concept of primary social goods. There are some things which a person needs in order to satisfy whatever wants he may have, and these are primary social goods; the powers and prerogatives of authority, income and wealth. In the original position, Rawls argues, it is rational to adopt the maximin principle to govern the distribution of primary social goods within the group. This leads directly to the second principle of social justice (the difference or maximin principle); social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged and (b) attached to offices and positions open to all under conditions of fair equality of opportunity.

What light does Rawls manage to throw on our original question about how individual utilities are to be weighted in the estimation of social welfare? Assuming that the first principle of social justice is met and part (b) of the second principle, Rawls tells us that social welfare increases when the least advantaged in the society is made better off, whatever happens to those enjoying a greater advantage. In measuring

changes in social welfare, therefore, provided that the above conditions are satisfied, the utility of everyone except the least advantaged is to be weighted zero. The actual weight given to the utility of the least advantaged individual is, of course, of no significance, merely affecting the scale by which social welfare is judged. This is in marked contrast to Harsanyi's utilitarian principle according to which all are to be weighted equally.[34]

Harsanyi's theory was criticised earlier under four heads and we may now ask whether or not Rawls' falls to the same comments. Whereas Harsanyi requires to know the utility of all the individuals in each social state before deciding which state is best, Rawls needs only to know to what extent the lot of the least advantaged has been changed. This requires no interpersonal comparisons of utility, and the utility of the least advantaged individual may be measured purely ordinally. Thus Rawls escapes the first two criticisms which were made of Harsanyi's account. The remaining two criticisms are, however, just as pressing as before.

We saw that Harsanyi seems to think, mistakenly that his account is applicable to relatively micro problems of social choice. Rawls, on the other hand, denies this and sees his theory as providing a rational structure for society in which micro problems may be resolved. But even this modest claim is to be seriously doubted. Taxation is part of society's structure, according to Rawls, but despite a few starts, no suggestion about types and rates of tax deriveable from Rawls' theory came anywhere near being practically applicable.[35] Like Harsanyi, Rawls seems to offer an account of social decision making which, for all its theoretical niceties and subtlties, is impotent in the face of the complexities of the real world.

The final criticism of Rawls is theoretical and concerns attitude to

risk. Someone who employs a maximin criterion in making a decision under risk is taking the least possible risk. Whatever the potential gains, they will chose so as to maximize the benefit received from the worst outcome of each option.[36] In discussing Harsanyi, it was observed that the attitude to risk of the choser will influence the weighting of individual utilities in the summation which produces social welfare. If he is prepared to take risks he will favour a society where some enjoy a utility higher than others, and so he will weight the utility of such an advantaged person more than that of someone not so advantaged. If, on the other hand, he is adverse to risk taking, he will seek to protect himself by adjusting the weights of individual utilities in the opposite direction. We can now see a greater similarity between Harsanyi and Rawls than at first appears. As we have seen, Harsanyi assumes his single choser to be neutral towards risk, so that the utility of all is weighted equally; but Rawls assumes each of his chosers to be infinitely adverse to risk, so that they will employ the maximin principle for the distribution of primary social goods.[37]

This point was used to attack Harsanyi's suggestion, and the same criticism applies to Rawls. Rawls' individuals in the original position must come to an agreement about what principles of social justice to employ, but agreement is only possible if all of the individuals are infinitely adverse to risk and so willing to employ maximin in their choice. Thus, Rawls requires unanimity in attitude to risk just as much as Harsanyi. If some of Rawls' original chosers are not infinitely risk averse, but are willing to take a gamble on getting a larger share of unequally distributed social goods, then there can be no agreement about principles of social justice. Moreover, on standard accounts of decision making, attitude to risk is a matter of subjective taste, and there can be no rational argument capable of showing that someone's attitude is

wrong. If attitudes to risk in Rawls' original position vary, therefore, this is a conflict for which there is no rational solution, and so no rational principles of justice can emerge. Both Harsanyi and Rawls attempt to overcome this problem by elevating some chosen risk attitude - neutrality and infinite aversion respectively - but this will not do.[38] Different attitudes to risk are possible and cannot, at least on traditional accounts of decision making, be stigmatised irrational or unreasonable.

In conclusion we may return to the original question of this section: how are individual utilities to be weighted before their aggregation into social welfare? What was required here were reasons for a particular distribution of weights for in the lack of such reasons the weighting can only be arbitrary. We have seen, however, that attempts to justify dfiferent schemes of weighting have come to grief. Finally, it must be pointed out again that the original question is really an expository device and none of the arguments developed above depend upon the assumption that social welfare is a sum of individual utilities. We may conclude, therefore, that, as yet, the choice of social welfare function (of any kind) can only be arbitrary.

3 The Measurement of Total Income

If the method of social welfare function discussed in the previous section does not offer any promise in the search for the transition from individual to social values, there is an alternative which we shall discuss here. We saw earlier that welfare economics now finds no room for interpersonal comparisons of utility, and that ordinal measures are now used more or less exclusively. Scepticism about interpersonal comparisons became serious in the 1930s where its advocacy by Robbins and others[39] led to something of a crisis within the subject. If one man's utility could not be compared with another's, then what hope was there

for any comparison of different economic states? A way out of the problem was suggested by Kaldor, who proposed that questions of total consumption should be isolated from questions about the distribution of consumption. Thus, welfare economics was to consist of a scientific, objective part concerned with deciding whether one economic state could maintain a greater total consumption than another; and a second, political or ethical part concerned with how this total might be distributed.[40]

Central to the first part of welfare economics was Kaldor's principle that an economic change produces an increase in consumption if those who gain from the change could compensate the losers and still be better off. It is easy to see that this is equivalent to saying that consumption increases if it is possible to make a Pareto improvement. This principle has great intuitive appeal, but cracks soon appeared. Scitovsky revealed an ambiguity in Kaldor's idea of compensation. For Kaldor, this was a flow of consumption from gainers to losers, but Scitovsky showed that it could also take the form of a bribe from the losers to the gainers, for the latter to forego their benefits.[41] It is only possible to speak unambiguously of an increase in total consumption if the Kaldor criterion is satisfied and if this other form of compensation is not possible (the Scitovsky criterion).

This proposal soon became problematic, however. The problem is best seen by using the utility possibility curves below.[42] Such a curve shows the maximum utility which can be enjoyed by one consumer, A, given a fixed utility for another consumer, B, (utilities being measured ordinally) for a fixed quantity of commodities. Each point on the curve represents a different distribution of commodities between the consumers. The double Scitovsky-Kaldor criterion applied to the shift from Q_1 to Q_2 in fig. 1 requires that Q_2 represents a greater consumption than Q_1 only if the utility possibility curve through Q_1 is <u>everywhere</u> inside that through Q_2, as shown.

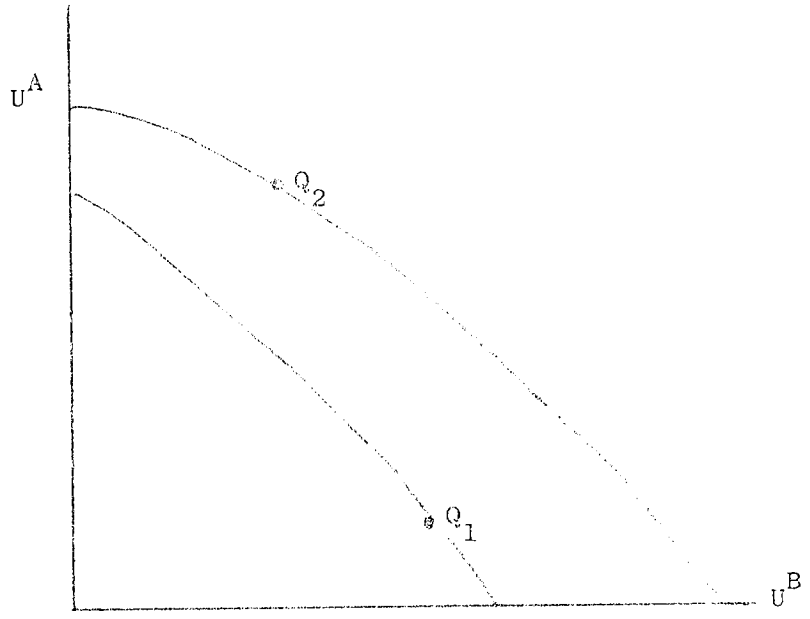
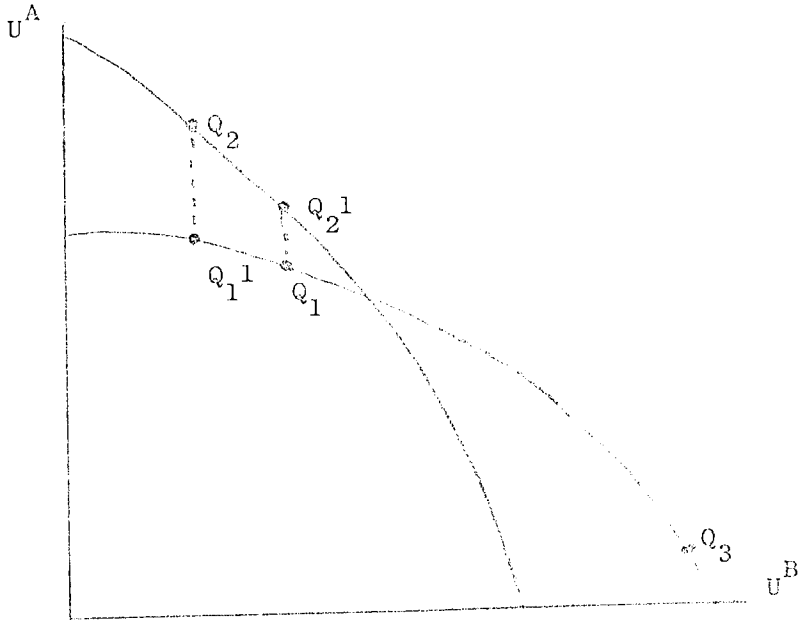


fig. 1 - Satisfaction of Scitovsky-Kaldor Criterion

If the two utility possibility curves cross, as in fig. 2, then matters are not so simple.



In the shift from Q_1 to Q_2, if the gainer, A, moves from Q_2 to Q_2^1 he can compensate B and still be better off than he was at Q_1. Hence the Kaldor criterion is satisfied. The maximum bribe that B can offer to A is by moving to Q_1^1, but this gives A less than he has at Q_2, and so the Scitovsky criterion is met. The transition from Q_1 to Q_2 would, therefore, seem to produce greater consumption. But now consider the move from Q_3 to

to Q_2. At Q_3 there is exactly the same quantity of commodities as at Q_1, the only difference being in their distribution. Nevertheless, there is no way in which A could compensate B. If A gives up all his consumption to B, B is still worse off than he is at Q_3. The Kaldor criterion is not satisfied, and so the transition from Q_3 to Q_2 does not represent an increase in total consumption.

It can be seen from this, that if the utility possibility curves intersect, then no meaning can be attached to increases in total consumption. Total consumption is relative to income distribution, and so the only unambiguous meaning which can be given to an increase in consumption is that there is an increase for all possible income distributions. This requires the non-intersection of utility possibility curves discussed earlier. Unfortunately, such non-intersection cannot be concluded from the usual index data of the economist. It can only be known to exist when the physical quantity of all goods shows an increase, something which is, of course, a trivial case. Attempts to sidestep difficult questions about the distribution of income and concentrate on measuring simply its total, therefore fail. Any judgement that one economic arrangement is better than another can only be justified if the distributional differences can be shown to increase social welfare. This, of course, takes us right back to the problems of the previous section.

This is well illustrated by the use which is made of cost benefit analysis. As we have seen measures of total consumption are relative to income distribution, and so it is customary to assume that its present distribution is optimal. This, of course, is a value judgement. Once this is made, policies which increase total consumption, relative to this optimal distribution can be identified and adopted. Some other distribution could, it is true, be regarded as optimal, but calculations of changes in consumption can generally only be made on the basis of present

income distribution, because the changes in prices resulting from a change in income distribution cannot generally be calculated. This can hardly be seen as a solution to the problem of rationally balancing one man's losses with another's gains, because the acceptance of a particular income distribution as optimal is totally arbitrary. No reason is generally given for the acceptance of present distribution as optimal, except that this distribution is the result of political decisions reflecting the preferences of legitimately appointed social decision makers. Apart from the element of mythology here, the question is whether these decision makers have acted reasonably or not. Can they be criticised for their views on income distribution, or can they justify their views by public reasons which go beyond their own private tastes? We come, therefore, full circle, for these are essentially the same questions as those asked at the beginning of the previous section.

4 Arrow's Theorem

Throughout this chapter we have discussed the problems which beset attempts to base social values upon the values held by the individuals who make up the society. The transition from private to public values calls for some value judgements, and the traditional aim has been to find some set of such judgements which are weak enough to be acceptable by everybody. Hence, of course, Pareto, Harsanyi and Rawls. The possibility of finding a set of value judgements which is sufficiently strong to allow the transition from private to public values, sufficiently weak to be generally acceptable, and consistent has been investigated by Arrow.[43] Arrow makes the following assumptions about a group of individuals.

(a) No interpersonal comparison of utility is possible.

(b) No individual receives enjoyment from the decision process *per se*.

(c) There is no strategic misrepresentation of individual values.

(d) The method of social choice does not influence individual values.

(e) All individuals are rational.

(f) All individuals have a complete knowledge of all possible social states.

(g) Each individual can order these social states rationally.

By the term 'rational' in (e) and (g), Arrow means that an individual's preference judgements about social states x,y etc. are logically consistent, in accordance with the following axioms, where R is the relation "preferred or indifferent':

Axiom I For all x and y, either xRy or yRx (connectivity)

(It follows that for all x, xRx (reflexivity))

Axiom 2 For all x, y and z, xRy and yRz imply xRz (transitivity)

If these axioms are met, all individuals will be able to place all possible social states in order of preference. Arrow then defines a social welfare function;[44]

A social welfare function is a rule which, for every set of individual orderings of social states (one for each individual), states a corresponding social ordering of social states.

Arrow then considers what value judgements are likely to govern the choice of a social welfare function. He suggests four conditions which any social welfare function must meet, and so which are necessary (but not sufficient) conditions for any such function to be acceptable.[45]

Condition U (unrestricted domain): The domain of the social welfare function includes all logically possible individual orderings of social states.

Condition P (Pareto principle): For any x, y if everyone strictly prefers x to y, then x is socially strictly preferred to y.

Condition D (non-dictatorship): There is no individual whose strict preference for any pair x over y is always reflected in social strict preference.

Condition I (independence of irrelevant alternatives): If for any subset of all possible social states every individual's preference remains the same for every pair of social states from the subset, then the choice set of the subset should remain the same too.

Arrow then goes on to prove his famous result, the general impossibility theorem, which states that there is no social welfare function satisfying conditions U, I, P and D. This means that the very weak set of evaluative principles which seem to be a necessary part of any acceptable transition from individual to social values cannot be satisfied. It would seem to follow that the liberal ideal of making social choices in the light of the preference of society's members is an impossibility.

Needless to say, Arrow's result is very disturbing and has generated a whole cottage industry directed towards a remedy. The literature is

immense and cannot possibly be reviewed here, except to observe that no acceptable solution has emerged after a quarter of a century's effort.

CHAPTER 8 - CONTEMPORARY VIEWS OF DECISION MAKING I - INDIVIDUAL AND SOCIAL VALUES

FOOTNOTES

1. The originator of this idea was certainly not Bayes. Credit is normally given to D Bernoulli, Exposition of a New Theory on the Measurement of Risk (1738), translated L Sommer, Econometrica, 22, 1954, 23-36.

2. Starting with J von Neumann and O Morgenstern, Theory of Games and Economic Behaviour, Princeton University Press, 1947.

3. This is partly an extension of the theory of games, where the individuals of the group have incompatible interests, and partly a field of its own. For the latter, see below.

4. See, for example A Downs, An Economic Theory of Democracy, Harper and Row, 1957 and N Howard, Paradoxes of Rationality, MIT Press, 1971.

5. Experimental results on the third condition, transitivity of preference, are reported by W Edwards, The Theory of Decision Making, Psychological Bulletin, 51, 1954 and Behavioural Decision Theory, Annual Review of Psychology, 5, 1960, 473-98; K May, Intransitivity, Utility and the Aggregation of Preference Patterns, Econometrica, 1954, 1-13 and A Rose, A Study of Irrational Judgement, Journal of Political Economy, 1957, 394-402. W McCullock, A Recapitulation of the Theory, Teleological Mechanisms, Annals of New York Academy of Sciences, 50, 1948, 263, reports the discovery of intransitive rats who prefer food to sex, sex to the avoidance of pain and the avoidance of pain to food! For theoretical discussions of these reports see also G Tullock, The Irrationality of Transitivity, Oxford Economic Papers, 3, 1964, 401-6 and A Weinstein, Individual Preference Transitivy, Southern Economic Journal, 34, 1968, 335-43. See also, T Scitovsky, Welfare and Competition, 2nd en., 1971, chapter 11.

6. Roskill Commission on the 3rd London Airport, Papers and Proceedings, Vol. 7, parts I and II, HMSO, 1970.

7. See, for example, M Frost, Value for Money, Gower, 1971, p.27.

8. R Ridker, Economic Costs of Air Pollution, Praeger, 1967.

9. P Self, Econocrats and the Policy Process, Macmillan 1975, p.83. See also p.31.

10. For a review of examples see A Harrison and D Quarnby, The Value of Time in Transport Planning, in R Layard (ed) Cost Benefit Analysis, Penguin, 1972, 173-208.

11. J von Neumann and O Morgenstern, Theory of Games and Economic Behaviour, 2nd edn., Princeton University Press, 1947.

12. For a review of methods for the determination of the utility function of a decision maker see J Hull et. al., Utility and Its Measurement, Journal of the Royal Statistical Society, A 136, 1973, 226-47, reprinted

in G Kaufman and H Thomas (ed), Modern Decision Analysis, Penguin, 1977, 62-95.

13. For brief histories see D Ellsberg, Classical and Current Notions of Measurable Utility, Economic Journal, 64, 1954, 528-56; M Dobb, Welfare Economics and the Economics of Socialism, Cambridge University Press, 1969, chapter 45; S Nath, A Perspective of Welfare Economics, Macmillan, 1973.

14. V Pareto, Manuel d'économie politique, Paris 1909, chapter 6, section 53, and appendix, section 89. A translation of these passages is in A N Page (ed), Utility Theory : A Book of Readings, Wiley, 1968, p38 ff.

15. A Sen, The Impossible Paretian Liberal, Journal of Political Economy, 78, 1970, 156-7; Collective Choice and Social Welfare, Holden Day, 1970, chapters 6 and 6*. See also, B Fine, Individual Liberalism in a Paretian Society, Journal of Political Economy, 86, 1978.

16. K Arrow provides an extremely clear discussion of this problem in his Social Choice and Individual Values, Wiley, 2nd edn., 1963, p 34-7 and 63-4.

17. A Bergson, A Reformulation of Certain Aspects of Welfare Economics, Quarterly Journal of Economics, 52, 1938, 310-34. See also P Samuelson, Foundations of Economic Analysis, Harvard University Press, 1947, chapter 8.

18. I Little, Critique of Welfare Economics, Oxford University Press, 2nd edn., 1957, chapter 7.

19. According to H Sidgwick, Outlines of the History of Ethics, 5th edn., London, 1902, utilitarianism begins with Shaftesbury's, An Inquiry Concerning Virtue and Merit, 1711, though its clearest statement is in J Bentham, The Principles of Morals and Legislation, 1789; J S Mill, Utilitarianism, 1863, and F Edgeworth, Mathematical Psychics, 1888. For modern statements see, J S Smart, An Outline of a System of Utilitarian Ethics, Cambridge University Press, 1961 and D Lyons, Forms and Limits of Utilitarianism, Oxford University Press, 1965.

20. For a historical survey see J Gough, The Social Contract, 2nd edn., Oxford University Press, 1957 and O Gierke, Natural Law and the Theory of Society, translated by E Barker, Cambridge University Press, 1934.

21. J Harsanyi, Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking, Journal of Political Economy, 61, 1953, 434-5; Cardinal Welfare, Individualistic Ethics and Interpersonal Comparison of Utility, Journal of Political Economy, 63, 1955, reprinted E Phelps, Economic Justice, Penguin, 1973, 266-85, and Non-Linear Social Welfare Functions, Theory and Decision, 6, 1975, 311-30. It is interesting to compare Harsanyi's argument with that of Lerner from the equiprobability of individuals possessing a particular utility function, see A Lerner, The Economics of Control, Macmillan 1944. For two extensions see P Samuelson, A P Lerner at 60, Review of Economic Studies, 31, 1964, and A Sen, On Economic Inequality, Oxford University Press, 1973, 83-5. For another similar argument see H Leibenstein, Long Run Welfare Criteria, in J Margolis (ed), Public Economy of Urban Communities, Proceedings of 2nd Conference on Urban Public Expenditure, Resources for the Future, 1965.

22. A very similar argument is due to W Vickrey, Utility, Strategy and Social Decision Rules, Quarterly Journal of Economics, 74, 1960, 507-35, except that he assumes that all the chosers have the same utility function, thus circumventing the problem of interpersonal comparison. The restriction is, however, highly artificial and for this reason Harsanyi's exposition is superior.

23. Harsanyi has a second argument. First, social preferences and individual preferences are assumed to satisfy the usual axioms for utility (e.g. those of J von Neumann and O Morgenstern, Theory of Games and Economic Behaviour, Princeton University Press, 1947 or J Marschak, Rational Behaviour, Uncertain Prospects and Measurable Utility, Econometrica, 18, 1950, 11-41). Secondly, the evaluative principle that if two social states P and Q are indifferent from the standpoint of every individual, then P and Q are socially indifferent is accepted. It follows from this that social welfare is a weighted sum of individual utilities. This is weaker than the first argument, which also tells us that the weights should be equal, and so I will not discuss it in detail. See also P Pattanaik, Risk, Impersonality and the Social Welfare Function, Journal of Political Economy, 76, 1968, reprinted S Phelps, Economic Justice, Penguin, 1973, 298-318.

24. See also A Sen, On Economic Inequality, Oxford University Press, 1973, 14-15.

25. See I Little, Critique of Welfare Eonomics, 2nd end., Oxford University Press, 1957, chapter 4.

26. W Vickrey, Utility, Strategy and Social Decision Rules, Quarterly Journal of Economics, 74, 1960, 507-35. See also C Hildreth, Alternative Conditions for Social Orderings, Econometrica, 21, 1953, 81-94.

27. See P Pattanaik, Risk, Impersonality and the Social Welfare Function, Journal of Political Economy, 76, 1968, reprinted S Phelps, Economic Justice, Penguin, 1973, 298-318; K Arrow, Social Choice and Individual Values, 2nd edn., Wiley, 1963, 31-3; A Sen, Planner's Preferences: Optimality, Distribution, and Social Welfare, in J Margolis and H Guitton (ed) Public Economics, St Martin's, 1969 and P Pattanaik, Voting and Collective Choice, Cambridge University Press, 1971, chapter 9.

It might appear from this that interpersonal utility comparisons are impossible, but this would be too sweeping. Sen suggests that interpersonal comparability is a matter of degree - see A Sen, Interpersonal Aggregation and Partial Comparability, Econometrica, 38, 1970, 393-409 and 40, 1972, 959; Collective Choice and Social Welfare, Holden Day 1970, chapters 7 and 7*.

28. These are based on just discernable differences in utility. See W Armstrong, Utility and the Theory of Welfare, Oxford Economic Papers, 3, 1951, 259-71; M Fleming, A Cardinal Concept of Welfare, Quarterly Journal of Economics, 66, 1952, reprinted S Phelps (ed), Economic Justice, Penguin 1973, 245-265; M Kemp and A Asimakopulos, Social Welfare Functions and Cardinal Utility, Canadian Journal of Economics and Political Science, 18, 1952, 195-200; L Goodman and H Markowitz, Social Welfare Functions Based on Individual Rankings, American Journal of Sociology, 58, 1952, 257-62 and J Rothenberg, Marginal Preference and the Theory of Welfare, Oxford Economic Papers, 5, 1953, 248-63 and

Reconsideration of a Group Welfare Index, Oxford Economic Papers, 6, 1954, 164-80.

Objections to these measures are well known. See K Arrow, Social Choice and Individual Values, 2nd edn., Wiley, 1963, 115-8; P Pattanaik, Voting and Collective Choice, Cambridge University Press, 1971, chapter 9; A Sen, Collective Choice and Social Welfare, Holden Day, 1970, 92-4; W Vickrey, Utility, Strategy and Social Decision Rules, Quarterly Journal of Economics, 74, 1960, 507-35, and Measuring Marginal Utility by Reactions to Risk, Econometrica, 13, 1945, 319-333.

29. At least since H Sidgwick, The Method of Ethics, Macmillan, 1893. The first utilitarian treatment of the taxation problem in E Edgeworth, The Pure Theory of Taxation, Economic Journal, 1897, reprinted in F Edgeworth, Papers Relating to Political Economy, Vol. 2, Macmillan, 1925, partially reprinted in S Phelps (ed), Economic Justice, Penguin, 1973, 371-85.

30. K Arrow, Social Change and Individual Values, 2nd edn., Wiley, 1963, 10; P Pattanaik, Risk, Impersonality and Social Welfare Functions, Journal of Political Economy, 76, 1968, 1152-69, reprinted S Phelps, Economic Justice, Penguin, 1973, 298-318; J Rawls, A Theory of Justice, Oxford University Press, 1972, sections 27-8; and A Sen, Collective Choice and Social Welfare, Holden Day, 1970, chapters 7 and 7*.

31. Arrow changes his mind in Some Ordinalist-Utilitarian Notes on Rawls' Theory of Justice, Journal of Philosophy, 70, 1973, 245-75. See also, J Harsanyi, A Critique of John Rawls' Theory of Justice, American Political Science Review, 69, 1975, 594-606 and W Vickrey, Measuring Marginal Utility by Reactions to Risk, Econometrica, 13, 1945, 319-33.

32. This is recognised by P Pattanaik, Voting and Collective Choice, Cambridge University Press, 1971, chapter 1.

33. J Rawls, A Theory of Justice, Oxford University Press, 1972.

34. The literature on Rawls' theory is now very large and cannot be reviewed here. Of particular interest, however, is the debate between Rawls and Harsanyi. See J Rawls, A Theory of Justice, Oxford University Press, 1972, 25-34, 90-92, 150-175; Some Reasons for the Maximin Criterion, American Economic Review, 64 (Papers and Proceedings), 1974, 141-6 and J Harsanyi, Non-Linear Social Welfare Functions, Theory and Decision, 6, 1975, 311-20; Critique of John Rawls' Theory of Justice, American Political Science Review, 69, 1975, 594-606. Sen's Weak Equity Axiom may provide a sort of half way house between the extremes of Harsanyi and Rawls, A Sen, On Economic Inequality, Oxford University Press, 1973, 18-23.

35. See J Rawls, A Theory of Justice, Oxford University Press, 1973, 277-84; Y Itsumi, Distributional Effects of Income Tax Schedules, Review of Economic Studies, 41, 1974, 371-381; and S Phelps, Taxation of Wage Income for Economic Justice, Quarterly Journal of Economics, 1973, reprinted in S Phelps (ed), Economic Justice, Penguin, 1973, 417-38.

36. This insensitivity to potential profit is, of course, a standard criticism of the maximum criteria.

For criticism of Rawls use of it see references to the Harsanyi-Rawls debate above (ftnt. 34) and A Sen, Collective Choice and Social Welfare, Holden Day, 1970, 135-241.

37. For a very neat formalization of this point see K Arrow, Some Ordinalist-Utilitarian Notes on Rawls' Theory of Justice, Journal of Philosophy, 70, 1973, 245-75. The close similarity of the two theories is also revealed by their reformulation in terms of Suppes' grading principles. See P Suppes, Some Formal Models of Grading Principles, Synthese, 6, 1966, 149-158; and A Sen, Collective Choice and Social Welfare, Holden Day, 1970, 146-151 and chapter 9*.

38. To complete the spectrum we can imagine a group of chosers in the original position who adopt maximax and try to maximize the benefits of their choice, ignoring the likely bad consequences. These would be infinitely risk seeking agents and social welfare would depend solely upon the utility of the best off person in the society.

39. L Robbins, Interpersonal Comparisons of Utility, Economic Journal, 1938, 635-41.

40. N Kaldor, Welfare Propositions and Interpersonal Comparisons of Utility, Economic Journal, 1939, 549-52, reprinted in N Kaldor, Essays in Value and Distribution, London, 1960, 143-6. See also J Hicks, The Foundations of Welfare Economics, Economic Journal, 1939, 696-712.

41. T Scitovsky, A Note on Welfare Propositions in Economics, Review of Economic Studies, 9, 1941, 77-88; reprinted in T Scitovsky, Papers on Welfare and Growth, London, 1964, 123-38.

42. The argument below was developed by P Samuelson, The Evaluation of Real National Income, Oxford Economic Papers, 2, 1950, 1-29. reprinted in E Phelps, Economic Justice, Penguin, 1973, 65-91.

43. K Arrow, Social Choice and Individual Values, 2nd edn., Wiley, 1963.

44. Not to be confused with the Bergson social welfare function discussed earlier which is real valued.

45. What follows is not Arrow's original formulation, but that of A Sen, Social Choice Theory; A Re-examination, Econometrica, 45, 1977, 53-89.

46. For recent reviews see A Sen, Social Choice Theory: A Reconsideration, Econometrica, 45, 1977, 53-89; Collective Choice and Social Welfare, Holden Day, 1970; P Pattanaik, Voting and Collective Choice, Cambridge University Press, 1971 and P Fiskburn, The Theory of Social Choice, Princeton University Press, 1973.

CHAPTER 9

Contemporary Views of Decision Making II - Fact, Forecast and Decision

In the previous chapter contemporary views of decision making, by which is meant all those theories falling under the general headings of welfare economics and Bayesian decision theory, were criticised for the treatment which value receives at their hands. It was argued that all these views incorporated the traditional, but false, conception that an individual has privileged access to his own values, and that individual values are autonomous from factual considerations. It was also maintained that no contemporary theory of decision making could give a satisfactory account of the forging of social values from the values of the individuals composing society. The present chapter will continue this criticism of contemporary decision theories by considering the relationship which these theories claim to hold between facts and decisions. The conclusion I hope to reach is that contemporary theories require such great quantities of factual information for a reasonable course of action to be determined, that they can only apply to decisions which are relatively trivial. Decisions which are of any importance must always be taken with very incomplete information, but this is just where contemporary views of decision making offer no assistance, and can never hope to do so.

1. Fact and Decision

All contemporary theories of decision making claim that at least some decisions can be justified. It is claimed that there is some function which is a measure of what the decision maker wants, and some decisions can be shown to be correct in the sense that they optimize the value of this function. Complications arise when the decision maker has multiple objectives, i.e. when he seeks two or more distinct things, because there is generally no decision which will provide him with the maximum of each of these, so that some kind of trading between objectives is needed. As we have seen, additional complications arise when the decision is to be taken by a group of people who may have different values. In order to focus more clearly on the relationship between facts and decisions, it is best to avoid both of these problems and to consider one decision maker with a single objective.

The simplest case of decision making, according to contemporary views, is a decision taken under certainty. Here the decision maker has a number of options open to him and he knows what consequences will follow from the adoption of each option. The decision maker considers each set of consequences and places them in order of preference, assigning to each a number which reflects the utility of the set of consequences to the decision maker. The magnitude and the differences and ratios

of the numbers assigned are not significant; all that matters is their order. Any assignment will do provided that A is assigned a higher number than B if and only if A is preferred to B, and that A and B are assigned the same number if and only if the decision maker is indifferent between them. In other words, the utility of the sets of consequences is measured on an ordinal scale. The decision maker then adopts that course of action which maximizes his utility, i.e. he acts to obtain the most preferred set of consequences. The use of utility here may seem superfluous, but it greatly assists the development of decision theory, even where it is concerned with decisions under certainty, as in the economic theory of consumers' behaviour.

The first extension from this highly artificial restriction of perfect knowledge covers decisions which are taken under <u>risk.</u> Here, the decision maker is faced with a finite number of options but, for at least one of these options, he does not know for certain what consequences will follow from its adoption. He knows, however, what consequences are possible, and he knows the probability of each consequence which may occur. We may think of these probabilities as objective, i.e. as measures of the limit outcomes of repeated trials or of propensities within the objects exhibiting chance like behaviour. The decision maker may assign utilities to all of the possible outcomes of an option in the same way as before, and regard the utility of the option as the

expected utility of the set of possible outcomes. He may then compare the utility of all the options open to him and decide upon that one which maximizes utility just as before.

In real decision making there is rarely adequate information upon which to judge the probabilities of all the possible outcomes of all options. A less restrictive situation exists when the decision maker is faced with a choice between a finite number of options where he knows the possible consequences of each option and where, for at least one option, he cannot assign a probability to every possible consequence. Here we may talk of decision making under uncertainty[1]. Here, subjective probability, usually elicited from the decision maker by asking him whether he would risk a series of gambles, is used as a surrogate for objective probability[2]. Once subjective probabilities have been assigned, the procedure is as for risk. Utility is again measured on an ordinal scale and the utility for any uncertain option is reckoned as the expected utility of its possible consequences. That option with the highest utility is then chosen.

The relationship between decision and fact is very straightforward on this kind of account. The value of an option before the decision maker is determined by the value he places upon the option's consequences, and so it is necessary to determine what these consequences will be, or at least how likely they are to

follow from the option. There is, however, a class of decision which is not given any explicit treatment in the literature on decision making. These are what I shall call decisions made under ignorance. A state of ignorance exists when, even though the decision maker may be faced with the choice between a finite number of options, for at least one of these options he is unaware of some possible consequence[3].

There is a tendency in the literature to use the term 'uncertainty' to cover both what I call 'uncertainty' and 'ignorance'. This is very unfortunate because there are very great differences between decisions made under these two heads. Even if the standard account of the justification of decisions taken under uncertainty is accepted, there can be no way in which decisions under ignorance can be justified. On all traditional accounts, justification proceeds through the estimation of utilities or expected utilities, and in either case these can be calculated only if all of the possible consequences of all the options open to the decision maker are known[4]. If at least one possible consequence of one option remains unidentified, it can hardly be attributed a probability. Moreover, there can be no way of telling that knowledge of the possibility of this consequence will lead to a different option being favoured. No matter how good or bad the identified consequences of some option, the utility of the option may be lowered or elevated to any degree by the discovery of some previously

unidentified possible consequence. It follows that no decision under ignorance can be justified, unless some radically different account of decision making is adopted.

This is the first problem facing contemporary views of decision making - the impossibility of justifying decisions taken under ignorance. Just how serious a problem this is will be investigated below, but before considering this a second problem, following from the first, should be noted. It is unquestionably true that facts are relevant to decision making, including, of course, decision making under ignorance. Indeed, where our ignorance is profound, those facts which can be gathered are given particular importance. It falls to any theory of decision making to explain the relevance of facts to the making of decisions. Contemporary theories meet this challenge very elegantly for decisions under certainty, risk and uncertainty. The justification of any such decision calls for the identification of the possible outcomes of each option and the assignment to each of a probability, all of which are factual matters. So far, so good, but can this sort of account be extended to explain the relevance of factual information to decisions made under ignorance?

Consider a decision maker faced with the problem of making some decision under ignorance. Let I be a set of factual information about the probabilities of various outcomes of the options open to the decision maker, where I is not sufficiently large to eliminate ignorance and reduce the problem to one of

decision making under uncertainty. Thus, even if he acquires I, the decision maker will still be faced with a situation of ignorance; he will still not know all the possible outcomes for each option before him. The question is, is I worth acquiring, i.e. will knowing I improve the decision maker's choice?

According to all contemporary views of decision making, the answer must be no. We have seen that, on such views, no decision under ignorance can be justified. It follows that the accumulation of factual information can never enable such a decision to be justified. Acquiring I is, therefore, of no value whatsoever. Without I, the decision maker is unable to justify whatever decision he takes, and so his decision can only be seen as arbitrary. With I, he is in exactly the same position; he has in no way advanced. He may, it is true, make a different decision after acquiring I, but this cannot be said to represent an improvement on the decision he would have made without I. His decision is just as arbitrary as before, and the change can only be seen as reflecting the decision maker's particular psychology. According to contemporary views, therefore, factual information is relevant to making a decision under ignorance only if it is sufficient to eliminate ignorance, reducing the problem to, at least, decision making under uncertainty. On these views, the value of information is quantized. If the information is not enough to dispel ignorance, i.e. not enough for the decision maker to be able to identify all possible consequences of each

option, then its value is zero. If ignorance can be dispelled, then the information has a positive value and may be worth acquiring.

This, of course, is a highly paradoxical conclusion. It can hardly be disputed that information can be of the greatest value in making decisions under ignorance, even where it is not adequate to reduce ignorance to uncertainty. Indeed, in many cases very large resources are spent on acquiring just such information. To give one example; the decision which is now pending concerning whether to build a fast breeder reactor has undoubtedly to be taken in a state of ignorance - not all consequences of all of the options which are open are known. Despite this, a huge R and D budget is disbursed for the acquisition of knowledge thought to make for a better decision, yet with no hope of reducing ignorance to uncertainty. Why this should be so, contemporary views of decision making are totally silent.

Contemporary views of decision making set out to state how decisions may be justified, and starting from simple cases, such as decisions under certainty and risk, they hope to extend their account to those taken under uncertainty and, finally, ignorance. We have seen that, whatever their success with simpler cases, there is no hope of these views ever providing a satisfactory account of how decisions made under ignorance can be justified.

The impossibility is manifest when it is remembered that, on all these views, a decision is justified by its possible consequences, and that ignorance exists where the decision maker is unaware of some of the possible consequences of the options before him. Nor, as we have seen, can contemporary theories explain why factual information is relevant to making decisions under ignorance. On these theories, such decisions must be regarded as arbitrary, and so immune from improvement by any amount of information less than that needed to eliminate ignorance. The next step is to inquire as to the importance of decisions under ignorance. If most real decisions are, at worst, taken under uncertainty, then the problems raised here may be relegated to academic squabbling. If, however, there are important decisions which need to be taken under ignorance, it will be as well to search for a more satisfactory account than that offered by contemporary theories of decision making.

It is fair to say that all important real decisions, as against their text book surrogates, are ones which have to be taken under ignorance. In real decision making, it is usual for the decision maker to be unaware of all the options which are, in fact, open to him. Even if he is aware of all the options, he will nearly always be unaware of some of the possible outcomes of at least one of these options. In either case, the decision maker is operating in a state of ignorance. Here, contemporary

discussions of decision making consistently misrepresent the situation by making no distinction between uncertainty and ignorance. It is, indeed, almost universal practice to conflate both kinds of decision problem under the single heading of decision making under uncertainty. Needless to say, this conflation is disastrous and yet, without it, no contemporary theory of decision making can make a convincing stand.

Let us enquire a little further into how the trick is done. Textbooks and research reports on contemporary decision theory abound with descriptions of decisions which are, in reality, ones which must be taken in a state of ignorance, and which are represented as being under uncertainty. It is easy to see how the disguise is effected. Any decision is taken against a background of facts and institutions and ways of doing things which are regarded as more or less fixed, at least as far as the decision maker is presently concerned. In theory, the decision maker has a huge, and sometimes an infinite, number of options which he might take, but the human mind is so limited that they cannot all be considered. Indeed, it would be irrational to consider too many options, since the cost of the investigation might well outweigh any likely benefits from good decision making. Any decision maker, therefore, considers a tiny subset of all possible options before him, rejecting the others out of hand as obviously inferior or else beyond his concern at the moment. The problem in this way is reduced to picking the best option from half a dozen or so. Many of the options which are not considered are not consciously

rejected by the decision maker, because they never emerge into his consciousness; through lack of imagination or a shrewd intuitive eye for the impossible. Regarded strictly, the decision before the decision maker is one under ignorance, but when the number of options are reduced in the way discussed, the decision appears to be one under uncertainty or even risk. The transition is never explicitly discussed in the literature on formal decision making, but the misrepresentation of decisions under ignorance as ones under uncertainty is absolutely central to all of contemporary decision theories.

An example might be helpful here. A typical text book simplification is from P. Moore et al, Case Studies in Decision Analysis, (Penguin, 1976). The example concerns a firm who have to decide between launching a new product and not launching it. Best estimates yield the usual decision matrix below, according to which expected monetary value is maximized by launching the new product.

		Payoff (£000)	
	Decision	10% market share (0.7 chance)	2% market share (0.3 chance)
1.	Launch product	100	- 50
2.	Drop product	0	0

The decision is obviously represented as one under uncertainty (or even risk, depending upon how the probabilities were estimated). A little reflection, however, soon reveals that any decision like the one of the example must be taken under ignorance. Any decision maker considering a problem similar to the example will have before him an enormous number of options and not just the convenient pair of launch product: drop product. The whole firm could cease to market anything, sell its equipment and invest the resulting capital in an enormous number of different ways. The firm could also be sold as a going concern, at a variety of prices, and the money invested as before. Or, again, the plant needed for the new product could be burned down and the fire insurance embezzled by the managing director. And why look at the possibility of launching only one product; why not wait until a pair, or trio, of products can be launched, perhaps to save on advertising. The calculations of the marketing director presumably rest upon assumptions about pricing, distribution, packaging, advertising, retailers and wholesaler margins, and a host of other things. Even altering these in combination opens up a host of alternatives. In the example we are offered, however, all these options, and the many more which may be imagined, are suppressed, leaving us with a simple pair. What is really a situation where there are so many options open that a decision maker can hardly be aware of them all is transformed to a surrogate with only two options. What is a decision under ignorance has become one under uncertainty.

This is, as I have said, a perfectly typical example of how decisions under ignorance are simplified to easily managed ones under uncertainty. The justification for this, of course, is that the human mind cannot hope to compare any more than a handful of options and that the expense of doing more is likely to exceed the benefits derived from any decision analysis. In the hypothetical marketing example, for instance, nobody has suggested liquidating the firm, or selling it; these, and many other options are simply not under scrutiny. The firm has restricted the problem to make it amenable to discussion and eventual resolution at modest cost. The need for such restrictions is a truism. The limitations of the human mind are matched only by the limitations of the human pocket. It is quite impossible and crazily expensive to consider a huge number of options in detail, this no-one can deny. What can be denied, however, is the usefulness of applying the techniques of contemporary decision theories to the necessarily limited number of options which can be considered.

This is best seen by expanding the example of the product launch. If a decision theorist examines the pay-off matrix above he will declare that the decision to launch the product is justified. The firm aims to maximize its revenue and this can be achieved by maximizing the expected monetary value of its projects. But, given the problem discussed above, what can this claim of justification amount to? The most natural sense, and the sense clearly intended by the author of the example, is that launching

the product is justified because no other action by the firm leads to a higher expected monetary value. This, of course, is a very strong claim indeed. It cannot possibly be justified by considering just some of the options open to the decision maker. If restrictions on time, effort and brainpower mean that some options must go unconsidered, then the claim that launching the product is justified cannot be supported; it stands forever in jeopardy from the identification of some option which has a greater expected monetary value.

A weaker claim which might be more defensible is that launching the product is the best option of those considered in the analysis. Even this begins to look a little threadbare on examination. Presumably, what is meant is that when the complete set of options is ordered by expected monetary value the launching of the product will appear superior to dropping the product. Considering more options than the pair first looked at, however, may well lead to the reversal of the original preference. Suppose that another pair of options, sell the firm: keep the firm, are imposed on the first pair. We now have four options, launch the product and keep the firm, drop the product and keep the firm, launch the product and sell the firm, drop the product and sell the firm. The first two of these are equivalent to the original options. Imagine that the net present monetary value of the firm's activities in the future are £9 million and that this is

independent of the fate of the new product. The firm can be sold for £10 million if the product launch is successful, but if it fails then inefficiencies in the organization will be uncovered which will reduce the price to £8 million. If the product is dropped, the sale will realise £9.5 million. The matrix is now:

	Decision	Payoff (£000)	
		10% market share (0.7 chance)	2% market share (0.3 chance)
1.	Launch product - keep firm	100 + 9,000	- 50 + 9,000
2.	Drop product - keep firm	9,000	9,000
3.	Launch product - sell firm	10,000	8,000
4.	Drop product - sell firm	9,500	9,500

Expected monetary values for options 1-4 are, 9,055, 9000, 9,400 and 9,500 (£000) respectively. Option 4 should, therefore, be favoured; the product being dropped and the firm sold. This, of course, is quite contrary to the conclusion reached when only 1 and 2 are considered, when launching the product is favoured. The example is, of course, generalizable, and constitutes a grave embarrassment to much of contemporary decision theory. When a restricted number of options is considered, choosing the best of them may preclude choosing the best from an expanded set of options, or the best from the complete set of options. In general, adopting the best of the considered options may effectively prevent the

attainment of the best possible option. There is no way of knowing that this will not happen, except by considering the complete set of options. This is very reminiscent of the general problem of the second best in economics[5].

So much, then, for the way in which many decisions are disguised[6] as ones under uncertainty and the dangers of such dissembling. But so far, my claim that most real decisions are ones which must be taken in a state of ignorance is just assertion. The remainder of the chapter will attempt to place this on a more solid foundation by considering the necessity of forecasting in decision making and the perils associated with trying to glimpse the future.

2. Problems of Forecasting

Any decision is an attempt to bring about some favoured event in the future, and so all decisions involve hypothetical (or provisional) forecasts of the form 'if decision d is made, then c will happen (with probability p)'. If a decision is to be justified, as is supposed by all contemporary theories of decision making, then all the hypothetical forecasts which it involves must, themselves, be justified. It is necessary, therefore, to ask when such forecasts may be regarded as justified. It is best to begin with a hypothetical forecast of a physical event such as 'if stone s is unsupported at time t, then it will fall to the ground with a constant acceleration a' (H). This claim is justified by the

physical law 'all unsupported bodies close to the Earth's surface fall at a constant acceleration a' (T), plus the initial condition 'at time t, stone S will be close to the Earth's surface' (I). The conjunction T.I is able to justify H because H is entailed by the conjunction. Note also that since the antecedent of H refers to some future time t, the initial condition I also refers to a future event. In other words, I is a (non-hypothetical) forecast.

If this account is general, and there is no reason for thinking otherwise, then we can say that any hypothetical forecast is justified if and only if it follows from some general law, which we have reasons for accepting, and some set of initial conditions referring to the future which we have reasons to think true. The justification of any hypothetical forecast, therefore, depends upon knowledge of some general law and a forecast of future initial conditions which is known to be reliable. From what has been said above, the justification of any decision will also be dependent upon the same two factors; known general laws and reliably forecasted initial conditions.

The justification of a decision is, on this account, a very difficult enterprise. It is not enough that the hypothetical forecasts involved in the decision are correct; nor is it even enough that they are supposed correct for the right reasons. Justification requires that the hypothetical forecasts are supposed correct for the correct reasons, and that these reasons are known

to be correct at the time the decision is made. Failure to notice this last condition can enable hindsight to greatly exaggerate our powers of making justified decisions. In an article expressing great faith in such powers, Gabor gives the example of a silt bar which began to grow in the Rangoon River in 1910 and which, by 1931, was threatening to close the port of Rangoon[7]. The port authorities consulted the distinguished engineer Sir Alexander Gibb. Gibb had the inspiration of employing a scale model to simulate changes in the estuary. Using this, he made the hypothetical forecast that if no dredging was undertaken, the bar would begin to disappear of its own accord. The port authorities, therefore, decided to do nothing, a decision which Gabor regards as justified by Sir Alexander's work, and happily the bar behaved exactly as predicted and the port was saved.

Is Gabor's claim that the port authority's decision was justified reasonable? The problem is that when Sir Alexander performed his experiments the theory he employed about scale models was not known to be correct, although we now know this to be the case. His hypothetical forecast was, therefore, based upon a clever guess, and no more. It cannot by any stretch of the imagination be regarded as justified. Nor, of course, can the decision based upon this forecast be regarded as justified. What is necessary for justification is that the theory employed in the making of any hypothetical forecast should be known to be correct at the time the forecast is made. In claiming a forecasting success,

Gabor is using hindsight. All kinds of hypothetical forecast are made every day and are based upon speculations, guesses and theories of all varieties. Some of these forecasts happen to be correct, and some of these happen to be based upon guesses which turn out correct, like Sir Alexander's. But this is not enough to show that the forecasts in question, and decisions based upon them, were justified. For this, the theories involved must be known to be correct, not with hindsight, but at the time the decision or forecast is made.

Having seen how any decision involves forecasting, let us turn to some of the problems of forecasting. The first of these concerns the general laws used in the justification of a hypothetical forecast. In dealing with simple, isolated physical systems, there is little difficulty here. The difficulties become notorious, however, when we seek to make hypothetical forecasts about the behaviour of social systems. The laws which govern even relatively simple social systems are, as yet, hidden from us. As we shall see below, even economics, the most studied and developed of the social sciences, provides us with only the sketchiest theoretical understanding; one quite insufficient for the justification of any hypothetical forecast. The reasons for this startling difference between the natural and the social sciences need not concern us here, as it is universally recognised[8].

Hypothetical forecasts about physical systems can be very hard to make if the system is complex, even though the laws which govern each part of the system are well understood. In meteorology, for example, the laws underlying the behaviour of the atmosphere are more or less completely worked out, but the system is so complex as to defy the making of any really precise hypothetical forecasts or straightforward predictions[9]. Here, it may be appropriate to remind ourselves of one of the several definitions of the term 'system'. According to Ackoff[10], a system is a set of elements where:

1. The properties of each element in the set has an effect on the properties of the set as a whole.
2. The properties of each part and the way they affect the whole depends upon the properties of at least one other element in the set.
3. Every possible subgroup of elements in the set has the first two properties; that is, each has an effect, and none has an independent effect, on the whole. Therefore, the elements cannot be divided into independent subsystems.

When we consider that groups of people form systems of which the third condition is particularly true, the limitations on our theoretical understanding of social systems is at least partially explained. Any forecaster is in a dilemma in dealing

with social systems. Any forecast has three dimensions; topic, geographical limit and time horizon. If a forecast of the behaviour of any part of the system is to be justified, then understanding must extend to all parts of the system and to the system as a whole. The three dimensions of the forecast, therefore, grow to such an extent that impossibly large quantities of data are required. If the problem is to be kept to manageable proportions, then the three dimensions must be severely limited, in which case interactions with unaccounted parts of the system may always throw forecasts awry. This is clearly seen in the various modelling techniques used in forecasting. If the model attempts to deal with all aspects of the system, it soon becomes enormously complex and acquires an insatiable appetite for data. If, however, only part of the system is modelled, no forecast made by the model is trustworthy because interactions with the unmodelled part of the system may always generate surprises[11].

An independent problem for the justification of a hypothetical forecast about a social system stems from the difficulty of forecasting initial conditions. Even if all the laws governing the system were known, a hypothetical forecast could only be justified if these could be coupled with a reliable forecast of future initial conditions. The first problem here is the waywardness of individual tastes and behaviour. Suppose, for example, we are trying to find the consequences of insulating houses on a

large scale; i.e. we wish to know which hypothetical forecasts of the sort 'if houses were now insulated then in y years time .....' are true. It is a law of economics that if people spend less on heating their houses to present temperatures, then their consumption of other things will increase. But just where this increase will come is a function of private tastes which form the initial conditions for our hypothetical forecasts. Determining what these tastes are on the scale and with the precision required is an impossible task.

A second, related, problem concerning the determination of future initial conditions is that they often depend upon decisions made by others. Many factors relevant to a decision are beyond the decision maker's control, but within the control of somebody else. For example, in deciding on a business investment, a businessman may need to forecast tax rates in the future, over which he has no control. Government ministers, however, have some control over these rates, so that the businessman is effectively forced to forecast future ministerial decisions[12].

To sum up the discussion so far. A decision is justified only if the hypothetical forecasts upon which it is based are justified. A hypothetical forecast is, in turn, justified if it follows from some known general law and a forecast of future initial conditions which is known to be reliable. In the case of decisions about social systems, it is typically impossible to employ general laws and forecasts of future conditions in this way.

It would seem, therefore, that as far as social systems are concerned, decisions about them cannot be justified.

It might be objected that this scepticism is the result of a puritanical view of justifying forecasts, and that things become much less gloomy when a more relaxed attitude is adopted. It may be true that our knowledge of social systems is inadequate to justify hypothetical and non-hypothetical forecasts; but there are ways of forecasting which make little demands on knowledge and yet produce reliable predictions. Indeed, it will be objected, this is how forecasts are actually made in the social sciences. Very little appeal is made to any theoretical understanding of the system in question, and forecasts are developed from various ad hoc devices. An enormous number of these devices are to be found in the forecaster's toolbox of all varieties of sophistication. The toolbox is, however, divided into three fairly watertight compartments; economic forecasting, technological forecasting and business forecasting[13]. A typical technique, usually applied to technological forecasting is the Delphi method. Here a panel of experts is assembled and each expert is asked in private when he thinks some particular technological development will occur if certain decisions, for example, on levels of R and D funding, are made. The results are then placed in a statistical format and circulated to the panel. Each expert then has the opportunity to

change his mind in the light of the opinion of others, but free from the psychological pressures of an ordinary meeting. These moves may be repeated several times until a consensus emerges, which is taken as a reasonably good hypothetical forecast[14]. It may also be used for non-hypothetical forecasts. Other devices are much more mathematical and depend on powerful statistical devices for detecting regularities in a series of data.

It is appropriate to call all these devices ad hoc, because they do not pretend to give any insight into why the forecasted event is more likely to occur than any other, or, in other words, why the results of the device are to be given any greater credence than guesswork. This is true even of those devices which extrapolate some detected trend in historical data. With no knowledge of the mechanism which generated such a trend in the past, the extrapolation of the trend is ad hoc in the sense I am here using the term. We obtain no insight into why the trend should continue into the future; we merely hope that it will.

The question before us, then, is whether the use of ad hoc devices can justify any hypothetical forecasts, and so assist in the justification of a decision based on such forecasts. Readers of Hume will find no surprise in a negative answer. If ad hoc forecasting devices are to be employed, some criteria for placing reliance upon them is obviously needed[15]. It might be thought

that all that is required for an ad hoc device to be reliable is a good track record, i.e. a high rate of forecasts in the past which have been found correct with the passage of time. But, without knowing why the forecasts generated by the device have been so successful, what justifies the extrapolation from its past to its future success? We are making a forecast about the future success of the ad hoc device, and without knowledge of the reasons for its past success, we will need some ad hoc device to assess this forecast. If the device used is the one in question, we have a vicious circle, whilst if it is some other ad hoc device, we have succeeded in starting an infinite regress. The problem evaporates once an understanding of the success of some device is granted to us, but then the device at once ceases to be ad hoc. It would appear, therefore, that ad hoc devices cannot justify hypothetical predictions or non-hypothetical ones. They may be the only tools at our disposal when we are dealing with social systems, but we generally do well to recognize the limitations of our tools[16]. Justification of decisions involves justification of hypothetical forecasts, and for this there is no substitute for the understanding of general laws governing the system. For this reason, of course, our ability to justify decisions which affect social systems must be counted as limited in the extreme.

3. The Performance of Forecasts

In this section I propose to briefly look at the track record of predictions made about social systems by what I call *ad hoc* devices. I argued in the previous section that past predictive success for these devices cannot be extrapolated to their success in future predictions, but the temptation to make this naive error will persist so long as it is thought that the predicters' toolboxes contain some devices with an excellent record of achievement. Once the poverty of the toolbox is exposed, there can be no temptation to look to it for future predictive success.

The Delphi technique mentioned earlier depends on questioning experts, and some work would seem to make this a promising avenue. Gilfillan, in a famous study[17], considered the success rate of the experts in predicting developments within their own field of competence and came to the promising conclusion that 65% of such predictions came true. A later study suggests 50%, but even this holds out hope after the scepticism of the previous section[18]. When the list of predictions is considered, however, they seem to follow the advice of Lilley[19] in being extremely imprecise. Obviously, a very imprecise forecast, such as that new motive power sources for motor cars will be available after 1995, has a greater chance of being correct than a precise forecast, such as that motor cars will be powered by hydrogen gas from January 1995.

This offers no comfort, however, for, as Cole points out[20], forecasts are used by decision makers to distinguish between rival options, so that they must be precise enough to enable options to be distinguished. The kind of vague conjectures considered by Gilfillan may turn out correct surprisingly often, but are so feeble that they offer no guidance about what decisions to make now.

Shifting to more precise, and so to more relevant, forecasts, the picture is far less sanguine. From the dismally long list of failures of predictions based on ad hoc predicting devices which appear in many works[21], we need only point to some of the more recent and outstanding ones. The annual revisions to the UK population in 2000 A.D., upon which so many present decisions rest, show how difficult it is to forecast birthrate, even for a highly stable society[22]. A not untypical forecasting story is that concerning higher education in the UK. Estimates of the total number of students in higher education depends upon the birthrate eighteen years before and the proportion of that age group wishing to remain in education. Whilst the former is easy to estimate, forecasts of the latter have been seriously inaccurate[23]. The Robbins Committee, 1961-63, attempted to forecast student demand for the whole of higher education, but its forecasts were seriously on the low side[24]. By 1971/72 student numbers in higher education (463,000) and in universities (236,000) had already surpassed Robbins forecasts for 1973/74 (392,000 and 219,000 respectively). Anyone planning for the provision of higher education in the long term must have been dismayed when Robbin's forecasts for 1981 total

higher education students and university students (560,000 and 350,000) were replaced in 1970 by a new forecast[25] of 835,000 and 460,000 respectively. The 1972 White Paper[26] reduced the forecast of 1981 total numbers to 750,000, which was further reduced to 700,000 and reached 640,000 by the end of 1974. If these forecasts were really needed for planning, the chaos which their rapid change must have caused can only be imagined.

Manpower planning gives us a similar picture. The Willink Committee in 1957 forecasted a decline in the demand for doctors[27] which was accepted by many planners. As a consequence, there was a huge import of qualified doctors from overseas to meet an unexpectedly high demand and a crash programme of new medical schools. In 1961 the Zuckerman Committee forecast an adequate supply, or possible surplus, of qualified scientists by 1965[28]. By 1967, so great were the fears of a serious shortage of scientists that the University Grants Committee took emergency measures to increase the ratio of science: arts places in higher education to 2:1.

There is, I hope, no need to labour the point by piling on gloomy example after gloomy example. Even if the Humean problem of the previous section could be overcome, we simply do not seem to have any worthwhile ad hoc forecasting devices whose past record might justify us in accepting the forecasts they now offer.

We may now draw conclusions from our discussion of forecasting. Many proponents of forecasting would agree that it is extremely difficult, but argue that it is, for all that, something which cannot be avoided. de Jouvenal, for instance, tells us that[29]:

> the reason why we give forecasts is not that we know how to predict ..... We do not make forecasts out of presumption, but because we recognize that they are a necessity of modern society ..... I would willingly say that forecasting would be an absurd enterprise were it not inevitable. We have to make wagers about the future; we have no choice in the matter.

Why is forecasting inescapable? Because we have to make decisions and, as we have seen, the justification of any decision rests upon some set of hypothetical forecasts which rest, in turn, upon forecasts of future initial conditions. Our conclusion can only be that forecasting is such a difficult enterprise where it concerns social systems, that decisions about these systems cannot be justified. It is not, as is sometimes said, that great efforts are needed to produce good forecasts[30], but that the difficulties facing forecasting are so profound as to make forecasts precise and general enough to be of interest to policy makers quite impossible. In a word, all but trivial decisions about social systems must be made in a state of ignorance. None

but the trivial can be based on a set of reliable forecasts.

What is needed, and what the remainder of this work hopes to provide, is a view of decision making which can cope with this sceptical conclusion, i.e. which can show how one decision may be better than another even though neither can be justified. What such a view should also provide is a more realistic view of the function of forecasting. On contemporary views of decision making, forecasting functions to justify decisions, but we have seen the errors which are embodied here. Schlesinger suggests that there are two different points of view in dealing with lack of knowledge in decision making[31]:

> ..... A first group, whom we might call the contingency planners, has felt some confidence in our ability to chart in advance successful policies for the unknown future. Their method has been to designate a system which can deal adequately with each of them. A second group, whom we might describe as contingency planners, has tended to emphasize the uncertainties and our limited ability to predict the future. Those who hold this view have consequently stressed the need for sequential decision-making, for improvisation, for hedging, and for adaptability. Heightening awareness of inevitable change should tend to make us more sympathetic to the latter approach.

Contemporary accounts of decision making embody the first of Schlesinger's two points of view. What is needed is a much more articulated account of the second point of view, which should include the role of forecasting in improvisation, hedging and adaptability.

CHAPTER 9

Contemporary Views of Decision Making II - Fact, Forecast and Decision - Footnotes

1. My terminology here differs from the conventional one which uses the term 'uncertainty' to cover what I divide into two groups, uncertainty and ignorance. See below.

2. Occasionally, probabilities may be avoided altogether by the use of the Laplace equiprobability rule or the maximin rule or some other rule. On this see R. Radner and J. Merschak, Notes on Some Proposed Decision Criteria, in R. Thrall et al (eds), Decision Processes, Wiley, 1954.

3. The conditions for a state of ignorance to exist are actually somewhat weaker than this. Ignorance exists if the decision maker does not know that he has complete knowledge of all possible outcomes for all options. Where he has such complete knowledge, but does not know of its completeness, it is appropriate to speak of the decision being under ignorance. This is, however, a point of very little significance.

4. If probabilities are avoided by the use of some decision rule such as Laplace or maximin, the problem of identifying all the possible outcomes persists.

5. First formulated in a fully general way by R. Lipsey and K. Lancaster, The General Theory of Second Best, Review of Economic Studies, 24, 1956/7, 11-32.

6. Another aspect of this disguise is the selection of possible outcomes for each option. Generally, a list of such outcomes will be extremely large, too large to handle. In the same way as options are chosen, some subset of possible outcomes is selected for each option. This raises exactly the same problems as before.

7. D. Gabor, Predicting Machines, Cambridge Opinion, 27, 1950.

8. On this see S. Cole, Long Term Forecasting Methods, Futures, 8, 1976, 305-19; I. Miles, The Poverty of Prediction, Saxon House, 1975; D. Phillips, Forty Years On: Anti Naturalism, and Problems of Social Experiment and Piecemeal Social Reform, Inquiry, 19, 19, 403-25; K. Popper, The Poverty of Historicism, Routledge and Kegan Paul, 1961; The Open Society and Its Enemies, Routledge and Kegan Paul, 1966.

9. The example of meteorology is from S. Coles, Long Term Forecasting Methods, Futures, 8, 1976, 305-19.

10. R. Ackoff, The Systems Revolution, in C. Freeman et al (eds) Progress and Problems in Social Forecasting, Social Science Research Council, 1976, 66-71.

11. S. Encel et al, The Art of Anticipation, Martin Robertson, 1975, p. 31 ff; S. Coles, Limitations of Large-Scale Models in Forecasting, The Planner, 60, 1974, 646-9.

12. See S. Encel et al, The Art of Anticipation, Martin Robertson, 1975, 24-8.

13. Several guidebooks to the available tools exist, each with a different ordering of the tools. For these see: E. Adam, Individual Item Forecasting Model Evaluation, Decision Sciences, 4, 1973; F. Chambers et al, How to Choose the Right Forecasting Technique, Harvard Business Review, 13, 1971; S. Makridakis and W. Wheelwright, Forecasting Methods for Managers, Wiley, 1973; D. Edwards, International Political Analysis, Holt Reinhart and Winston, 1969; R. Brown, Smoothing, Forecasting and Predicting, Prentice Hall, 1963; G. Box and G. Jenkins, Time Series Analysis, Prediction and Control, Holden Day, 1970; E. Jantsch, Technological Forecasting in Perspective, OECD, 1967; Technological Planning and Social Futures, Associated Business Programmes, 1972; G. Wills, Technological Forecasting, Penguin, 1972.

14. The method is described in all books on technological forecasting. It has lately come under some criticism. See: R. Ament, Comparison of Delphi Forecasting Studies, Futures, 2, 1970, 35-44; R. Amora and G. Salancik, Forecasting: From Conjectural Art to Science, Technological Forecasting and Social Change, 3, 1972, 415-26; E. Grabb and D. Pyke, An Evaluation of the Forecasting of Information Processing Technology, Technological Forecasting and Social Change, 4, 1973, 143-50; H. Sackman, Delphi Critique - Expert Opinion, Forecasting and Group Process, Lexington and Heath, 1975. See also, H. Linstone and M. Taroff, Delphi Method, Addison Wesley, 1975 and Technological Forecasting and

Social Change, 7 (2), 1975.

15. See, for example, R. Amora and G. Salancik, Forecasting: From Conjectural Art to Science, Technological Forecasting and Social Change, 3, 1972, 415-26.

16. Essentially the same point may be found in K. Popper, The Poverty of Historicism, Routledge and Kegan Paul, p. 115 ff. See also S. Encel et al, The Art of Anticipation, Martin Robertson, 1975, p. 80.

17. S. Gilfillan, The Prediction of Invention, U.S. National Resources Committee, Technological Trends and National Policy, U.S. Government Printing Office, 1937, 15-23.

18. G. Wise, The Accuracy of Technological Forecasts 1890-1940, Futures, 8, 1976, 411-9.

19. S. Lilley, Can Prediction Become A Science, Discovery, November 1946, reprinted in B. Barber and W. Hirsch (eds), Sociology of Science, Glencoe Free Press, 1962, 142-52.

20. S. Cole, Accuracy in the Long Run, Omega, 5, 1977, 529-42.

21. S. Cole, Accuracy in the Long Run, Omega, 5, 1977, 529-42; The Structure of World Models, in H. Cole et al, Thinking About the Future, Chatto & Windus, 1974, 14-32. S. Encel et al, The Art of Anticipation, Martin Robertson, 1975; I. Miles, The Poverty of Prediction, Saxon House, 1975; C. Freeman et al (eds), Progress and Problems in Social Forecasting, Social Science Research Council, 1976; T. Hutchison, Knowledge and Ignorance in Economics, Blackwell, 1977.

22. See The Growth of Population to the End of the Century, Social Trends, HMSO, from No. 1, 1970 onwards. See also, R. Page, Population Forecasting in S. Cole et al, Thinking About the Future, Chatto & Windus, 1974, 159-74.

23. H. Parkin, Innovation in Higher Education: New Universities in the U.K., OECD, 1969, pp. 62-4.

24. Robbins' Report on Higher Education, HMSO, Cmnd. 2154, 1963.

25. DES Planning Paper No. 2: Student Numbers in Higher Education, HMSO, 1970.

26. Education: A Framework for Expansion, HMSO, Cmnd. 5174, 1972.

27. Committee to Consider the Future Numbers of Medical Practitioners and the Appropriate Intake of Medical Students, HMSO, 1957.

28. The Long Term Demand for Scientific Manpower, HMSO, 1961.

29. B. de Jouvenal, The Art of Conjecture, Weidenfeld and Nicolson, 1967, p. 277. See also R. Bauer, Detection and Anticipation of Impact, in R. Bauer (ed) Social Indicators, MIT Press, 1975.

30. Thus S. Makridakis, Forecasting, its uses and limitations, European Business, 43, 1975, 48-62, tells us that 'the process of forecasting could be made very accurate if we are willing to increase the amount spent on achieving forecasts'.

31. J. Schlesinger, The Changing Environment for Systems Analysis, in E. Quade and W. Boucher (eds), Systems Analysis and Policy Planning, Elsevier, 1968, p. 359.

CHAPTER 10

TOWARDS A FALLIBILIST THEORY OF DECISION MAKING

1. What May Be Learned From the Fallibilist Theory of Value?

Having looked at some of the problems facing contemporary theories of decision making, it is now time to see what might be learned about decisions from the fallibilist theory of value developed in Part 2, and to what extent an account of decision making based upon this theory can overcome the problems facing contemporary theories of decision making discussed earlier. I do not pretend, however, to offer a final, complete and polished fallibilist account of decision making. What I can present at the moment is more like the skeleton of such an account, but I hope to show that even this gives us the opportunity of viewing decision making in a different, and perhaps more powerful, perspective than the one we are accustomed to. My reason for exposing these ideas before they have received their final burnishing is simply to encourage others to explore this novel perspective.

Before developing the ideas of Part 2 into an account of decision making, it is first necessary to be clear about the status of the theory we are dealing with. The fallibilist theory of value provides a normative account of evaluation. It tells us, that is, how value judgements should be assessed if we are concerned to discover the truth. The theory does not set out to provide a theoretical description of how people actually evaluate their value judgements. If a theory of decision making can be extracted from this theory of value, it too will be normative. It will concern how decisions should be made, not how they are, in fact, made. In this, the theory resembles Popper's account of scientific method discussed in chapter 4, which tells us how scientific theories should be tested if we are interested in the truth. These normative theories cannot, however, be entirely isolated from contingent fact. Suppose, for instance, that Popper's views of how scientific theories should be tested is found to be at odds with how scientists actually test their theories. A staunch Popperian could always insist that this means that

scientists are mistaken and irrational in their assessment of theories, but such a cavalier move comes close to being ad hoc. If a rival view of scientific method holds that the scientific practice labelled irrational by the Popperian is really perfectly correct, then this must be counted a point in favour of the rival view and against Popper's. The same is true of the fallibilist theory of value, and any application of this theory which can be made to decision making. If people never assess their value judgements in the way prescribed by this theory, then this counts against the theory. It may not amount to a fatal objection, since it may be held that people are always mistaken in the assessments they make, but the theory is weakened because a rival theory of value has a chance of winning support for itself by giving an account of how value judgements should be assessed which is closer to actual practice. The more closely the actual behaviour of people fits the behaviour prescribed by the fallibilist theory of value, the stronger the theory. Exactly the same consideration applies to any fallibilist account of decision making which we may arrive at.

As I remarked earlier, my reason for investigating the possibility of a fallibilist theory of value was that it might throw illumination upon a fallibilist theory of decision making which might be capable of dealing with the sort of decisions concerning large scale technologies, taken under great uncertainty, or, as we should now say, ignorance, which are of enormous and increasing importance. To decide which of a number of courses of action to adopt is a special case of picking the most preferred item from a list, where the items are courses of action. Hence, a general fallibilist theory of value should lead naturally to a theory of decision making, if not in its entirety, at least in skeleton form. Traditional theories of decision making attempt to show how decisions may be justified, and, therefore, deny both sceptical claims below.

1. No decision can be justified.

2. No reason can be given for preferring one course of action over another.

The acceptance of 1., and so the denial of all traditional theories of decision making, follows from the sceptical arguments of Part 1. Taking a decision is a particular kind of value judgement, so that if no value judgements can be justified, no decision can be justified. A totally sceptical view would stem from the additional acceptance of 2, for then there would be no rational way of selecting one course of action from others. A middle way is provided by a fallibilist view of decision making which accepts 1 but denies 2, maintaining that different courses of action may be compared by critical discussion which provides reasons for taking one course of action rather than another.

Because selecting one course of action from a list of possible courses is to determine the most preferred course of action, all the apparatus of the fallibilist theory of value described in chapter 6 applies to decision making. Indeed, the case studies used to illustrate the theory of value in chapter 7 all concerned decisions; whether to introduce corporal punishment, whether to employ Buddhist economics in favour of conventional economics, and whether to remove lead from petrol. A course of action is proposed as preferable to all others and this proposal is then open to criticism. If it fails to withstand criticism, then some rival course of action must be suggested and submitted to the same critical scrutiny. If the proposal withstands criticism, this can never be enough to justify the proposal, no matter how intense the critical barrage which it survives, but the proposal becomes corroborated. Criticism is the discovery of facts which, together with background values shared by all parties to the debate, falsify the proposal.

For this reason, debates about what to do are entirely dominated by factual issues, the values involved in such debates being shared because of their mundaneness, rarely receiving any explicit consideration. As a proposal to act in a particular way is to be criticised by searching for falsifying factual sentences, it is easier to criticise proposals which forbid many factual sentences, i.e. proposals with a high factual content. To have a high factual content, the proposal should be highly universal and precise, requirements which also favour simplicity. Background values may, of course, become questioned as the debate proceeds, and when this happens they may be subject to the same sort of critical scrutiny given the proposal itself.

All of these points were illustrated in chapter 7, but to refresh the memory a brief example might help. It will also tie up with the discussion which follows on the reversibility of decisions.

Nuclear power stations built now will be providing energy in 20 years time. A comparison of the economic costs of nuclear power and conventional power must, therefore, stretch 20 years into the future. Predicting prices for fossil fuels and uranium over such a period is, however, quite impossible. Studies such as the U.S. Atomic Energy Agency's cost benefit analysis of the U.S. breeder reactor programme are rendered nonsense on this point alone. It is not surprising, therefore, to find narrow economic questions playing a very minor role in the continuing debate about nuclear energy. One cost which is, at this point in the debate, much more important is the risk to which the ordinary population will be exposed by the development of nuclear energy. We shall take this as an example of a critical debate about costs and benefits. Even this debate is now too extensive for a complete treatment here, so it is best to consider just one aspect of it, the safety of nuclear reactors.

For simplicity, we may imagine the debate to be between two parties, one upholding V_N, and one seeking to deny V_N.

V_N (x) $\exists$y(y is a nuclear energy programme and if x is a non-nuclear energy programme, then y pref x)

What a supporter of nuclear power must do is to point to facts, which when coupled with background values common to both participants, corroborate V_N. Similarly, an antagonist of nuclear power must seek facts which falsify V_N when coupled with background values. For our limited purposes, we may concentrate on just one background value, V_B.

V_B (x)(y)(If x is an energy programme posing normally accepted risks and if y is an energy programme posing more than normally accepted risks, then x pref y)

Because V_B is common to both sides in the debate, it can be expected that it will not receive any explicit mention. Instead, the debate will centre on the factual claims used by both sides. This, of course, is in great contrast to traditional views of value and decision making. On these views, there are ways of resolving factual disagreements but not disputes about values. Any disagreement between the parties in a debate which is persistent must, therefore, be due to their adoption of different values. The debate between the parties should, therefore, revolve around values and not facts. In this case, there should be agreement about how many lives are at risk from nuclear energy, the disagreement being about the value placed on these lives. A study of actual debates, as I hope I have shown in chapter 7, shows them to be much more like my account than the account given by traditional theories.

The first public report on reactor safety was published in 1957 by the U.S. Atomic Energy Agency.[1] The 'maximum credible accident' which the report foresaw involved a breach in the containment of a 200 MWe[2] reactor 50 km from a large city. The accident led to 3,400 deaths, 43,000 injuries and damage to property of $7,000 million. The A.E.C. were, however, unruffled by these results, since the report estimated the probability of the accident to lie between 1 in 10^5 and 1 in 10^9 chances per reactor per year, although little operating experience then existed on which to base such estimates. A similar study[3] regarded the maximum credible accident as even more serious, causing 133,000 deaths and inflicting immediate or long term injury on 500,000 people. The A.E.C. followed their 1957 report with a similar study in 1964 whose results were more dramatic, but this report was never made public.[4]

The debate over the safety of reactors centred on whether the risks involved were comparable to those normally accepted. The consequences of the maximum credible accident were horrendous, but the probability of such an accident was reckoned as extremely small. Opponents of nuclear power claimed that the hazard was quite beyond that normally accepted. Supporters of the new power source, on the other hand, argued that the risk involved was of the same order as that normally accepted without concern. Research revealed a pattern in the acceptance of risk which is often quoted. An annual risk of death to an individual of around 1 in 1,000 is not acceptable; at a risk of 1 in 10,000 public money may be spent to reduce the risk and below 1 in 100,000 risks are generally treated as problems for the individual although warnings may be issued. Annual risks of 1 in a million do not cause public concern and are generally ignored (being struck by lightening is in this category).[5]

We can see that opponents of nuclear power were arguing from the following premises:

F_1 (y)(If y is a nuclear power programme then y poses more than normally accepted risks)

F_2 $\exists$ x (x is a non-nuclear power programme and x poses normally accepted risks)

Since

$V_B.F_1.F_2 \longrightarrow$ $\exists$x(y)(x is a non-nuclear power programme and if y is a nuclear power programme then x pref y)

$V_B.F_1.F_2 \longrightarrow \bar{V}_N$

V_B is a background value and F_2 seems equally admitted by both parties, at least for the time being, so that if F_1 can be established, V_N must be seen as falsified. If, on the other hand F_1 is shown to be false, then V_N is corroborated. This is best seen by noting that:

$$V_N \longrightarrow \overline{F_1.F_2.V_B}$$

If F_2 and V_B are accepted, then the discovery of $\bar{F}_1$ corroborates V_N, and the more novel $\bar{F}_1$, the greater the degree of corroboration conferred on V_N, i.e. the more severe the test survived by V_N. For this reason, the crucial point in the debate became the truth or falsity of F_1.

A second criticism from opponents of nuclear energy concerned the emergency core cooling systems. Most of the energy produced from a power reactor is from the fission of uranium atoms, but this fission produces radioactive isotopes of other elements. It is relatively easy to shut off

the fission reaction in a reactor, but the radioactive decay of these isotopes continues, releasing heat. If the reactor core is uncooled, this heat could build up and eventually melt the reactor's steel pressure vessel and concrete containment, leading to a very serious so-called meltdown accident. To prevent this all large U.S. water reactors have an emergency core cooling system which cools the core in the event of the normal coolant being lost. These systems were, however, designed using very crude computer models and full-scale tests have never been conducted. What tests there have been have been worse than expected.

In response to this criticism, the A.E.C. held a large national meeting on the emergency systems, but failed to convince their opponents that the systems are satisfactory. The A.E.C.'s next move was the commissioning of a new study of reactor safety under the direction of Norman Rasmussen. The study was instructed to look at a broad spectrum of accidents, from common minor accidents to extremely uncommon disasters, in the hope of providing a less alarming picture than that obtained from looking at the maximum credible accident alone. The draft report of Rasmussen's group[6] concluded that the average risk to the population from nuclear accidents was very low; about 6 persons per year were expected to be injured and 0.3 persons per year killed by a reactor accident in any of America's first 100 reactors. There were two major surprises in the report; meltdown accidents were expected to be far more common than previously thought, but almost all such accidents would do very little harm outside the nuclear plant itself. The likelihood of a meltdown was put at 1 chance in 10,000 per reactor per year, but in nine out of ten cases the dangerous radioactivity would merely melt quietly into the ground. A meltdown followed by a major release of radioactivity was assigned a probability of 1 chance in 100,000 per reactor per year. The chance of a really catastrophic accident was put at 1 in 1,000 million per reactor per year.

On a fallibilist view, the Rasmussen report must be regarded as lending considerable support to the defenders of nuclear energy. Two highly unexpected facts combine to show that the dangers of nuclear energy are of the same order of the dangers normally accepted, so corroborating V_N by showing $\bar{F}_1$.

As might be expected, the Rasmussen conclusions were hotly contested by opponents of nuclear power. Some of these criticisms have, in turn, been answered. One objection is that the technique used by Rasmussen's team ignores the human factor in the control of sophisticated technology, and cases of accidents resulting from human failing are quoted in support.[7] In reply to this it is argued that 'the human fallibility argument is one that, pressed too far, would set an arbitrary and unduly restrictive limit on technological development'.[8]

A second objection points to failures in the Apollo programme which a technique very similar to that employed by the Rasmussen team failed to identify.[9] In reply to this it is argued that the Apollo programme was highly innovative and so had to use components whose reliability could not be adequately judged. Whilst nuclear energy is also a considerable innovation, it uses components which have a long record of use and hence are of a known reliability.[10]

A third objection is that the Rasmussen group did not consider the possibility of rupture of the reactor's containment which could lead to a very serious accident indeed.[11] In reply, defenders of the group point out that if the group's task is to be finite a limit must be drawn somewhere and some possible failures simply regarded as too incredible to merit attention. The rupture of a containment vessel is such a failure.

A more general criticism is that the methodology used in the study is erroneous. A detailed survey of this criticism is beyond the scope of our present discussion, but we may pick on just one part of it for purposes of illustration. Where failure rates for a particular component are unknown the Rasmussen team at times treated the rate as a random variable. This move is open to the objection that the failure rate is not a random variable; that it is a determinate, if unknown value. If this value is such as to produce higher levels of risk than those generated by other values in the random distribution, then a serious underestimation of risk will occur.[12]

So far, I have concentrated on the American scene, but developments in Britain have followed the same pattern. It has been estimated that a major reactor accident with release of 10% of the most harmful fission product, iodine-131 (some 10^7 curies), should occur less frequently than once in 1,000 years or, assuming 500 reactors, with two chances in a million per reactor per year. Farmer has, therefore, suggested that an event releasing 10^6 curies should have a probability below 1 in 10^6 per year per reactor; an event releasing 10^5 curies should have a maximum probability of 10^{-5}, and so on.[13] Proponents of nuclear energy are keen to stress that nuclear reactors satisfy the Farmer criterion, whilst opponents seek to deny this. As before, the debate rages around factual and not evaluative claims.

A second argument by supporters of nuclear energy argue that nuclear energy, even on the massive scale they envisage, will kill and injure far fewer people than providing the same quantity of energy with fossil fuels. The deaths and injuries from the latter are made up very largely of two classes, coal miners and those affected by sulphur dioxide emissions which accompany much fossil fuel combustion. This amounts to the rejection of F_1 since nuclear power is held to pose less risks than those normally accepted from the combustion of fossil fuels. This, of course, corroborates

V_N. In reply, critics of nuclear energy argue that improved technology can both reduce the accident rate in mines and lower sulphur dioxide in the air to perfectly safe levels. As before, then, the debate centres on factual and not evaluative claims. That nuclear energy, involving such potentially dangerous reactions, can be made less harmful than conventional fossil fuel burning is a surprising claim. If it is true, it would, therefore, seem to strongly corroborate V_N.

2. Special Features of Decisions

The previous section explored how the fallibilist theory of value gives rise to a theory of how decisions should be taken, since thechoice of a course of action is simply a special case of selecting the most preferred item from a list. What can be said of values in general can also be said of decisions, this much is clear, but there are important ways in which decisions differ from other values. We have so far ignored these special features of decisions, particularly in the various case studies which have been considered, and it is now time to remedy this.

If a value judgement is thought sufficiently important to be tested, then it should be framed in a highly testable way, i.e. it should be formulated to be highly universal and precise, so that it has a high factual content. The more facts forbidden by a value judgement, the easier it is to find some factual assertion which falsifies the value judgement. To test the evaluation is to search for falsifying factual sentences. For all value judgements, V, which do not result in a decision, if V is falsified by the factual sentence F, then the facts which make F true are not generally a consequence of the adoption of V. Adopting a value judgement which does not lead to a decision to act in some way has very little influence on what happens in the world. If, on the other hand, V^1 is a value judgement which leads to action, and if it is falsified by F^1, then the facts which make F^1 true are often a direct result of V^1's adoption. To give a simple example; consider an agent who adopts V_1 and V_2, which together entail F. Assume that he knows that apples do not make him sick, but that he is unaware of the effect of bananas on his digestive system. If the agent acts on V_1, and selects a banana to eat in preferene to an apple, let us suppose that he will be sick. This shows him that F is true, and that V_1 is, therefore, false. But if he had not acted on V_1,

he would not have been sick, and so would not have falsified V_1. The fact which makes F true is a direct consequence of the agent acting on his value judgement.

V_1	Banana B pref apple A
V_2	(x)(y)(If x makes me sick and y does not, then y pref x)
F	Banana B makes me sick and apple A does not make me sick

The testing of most valuations upon which major decisions rest involves this kind of testing. The search for falsifying factual sentences becomes the search for unexpected consequences of the decision already taken. For this reason, it is best to identify this kind of testing, so I shall speak of it as monitoring. The monitoring of a decision is the search for unexpected consequences of the decision which, when coupled with suitable background values, falsify the value judgement upon which the decision is based. When such a falsification occurs, I shall talk of the decision as being in error, or mistaken, notwithstanding that it might have been the best possible decision given the state of knowledge of the agent at the time of its taking.

There are two kinds of problem encountered in monitoring a decision's consequences, what I shall call system and institutional problems. Consider a decision to expand the University output of engineers in order to better the performance of British manufacturing industry. It is virtually impossible to monitor the consequences of such a decision since so many other factors are involved in how well or badly industry performs. The

problem of monitoring is due to the complexity of the system which one is dealing with, and so it is a system problem. Another example in the same area concerns decisions to invest in pure scientific research in the hope that economic benefits will eventually accrue to society as new discoveries are found to be useful. The pathways from pure research to practical application, however, are so complex that no estimate of the effect of the investment in research on the rate of technological change is possible, so that the decision to make this investment cannot be monitored. Again, this is a system problem.

Institutional problems of monitoring can arise in the absence of system problems. There may, for example, be no body able to monitor a decision, even though this monitoring is perfectly possible, or else there may be such a body but it may do its job badly. What is required of an institution performing monitoring will, of course, vary with the nature of the decision, but great care must be taken to consider this question, and to try to foresee the institutional problems involved in the particular monitoring.[14]

The second special feature of decisions not shared by other kinds of evaluation is that the revision of a decision, once taken, is always associated with a cost. If I used to rate Hardy above Dickens and have now reversed this rating, my revision imposes no costs on me or anyone else. It may, it is true, be accompanied by all kinds of mental strains as fixed opinions are challenged and changed, and these might be thought of as constituting some kind of subjective cost, but the point still holds for it can be said that the change of mind involves no objective costs. If, on the other hand, a decision is taken, then its revision is always associated with some objective cost, although 'subjective costs' may also be incurred. Any decision involves the investment of resources (though not necessarily monetary or even material ones, e.g. the writing of a book) and not all of these can be rescued if the decision is found to be

wrong and has to be revised. Since any decision may prove wrong, we should favour decisions which are highly reversible or flexible, i.e. decisions much of whose invested resources can be recovered and used for some other purpose.

This consideration immediately leads to three questions: what exactly is meant by a decision being flexible?; how can a decision's flexibility be measured?; and how can considerations of flexibility be combined with those about costs and benefits in the making of a decision? The following section deals with the first two questions, and section 4 with the final one.

3. Flexibility and Its Measurement

(a) Flexibility and the Control of Systems

In this section it will be shown that flexibility may be viewed in two equivalent ways: as a property of systems which reflects the ease by which a system may be controlled, and as a property of a sequence of decisions which reflects how decisions taken at one time destroy options at a later time. As will be seen from the next section, flexibility is usually seen in the second of these two ways; a decision is usually seen as inflexible to the extent that it closes off options which would otherwise be open to the decision maker in the future. There is nothing wrong in this conception, except that it tends to mask a crucial factor in estimations of flexibility; the time which separates one decision point from the next. This factor, however, becomes extremely clear when flexibility is viewed as the ease with which a system may be controlled. For this reason, this conception, it is suggested, is generally superior to its more customary alternative.

The task of showing the equivalence between the two views of flexibility is a double one. It must be shown that a system which is easy to control can be seen as a sequential decision of high flexibility, and then the reverse of this must be established. Taking the tasks in this order, we may begin by considering, at the highest level of abstraction so as to ensure generality, how systems may be controlled.

Fig. 1 represents the general problem of controlling a system. Control is called for only if the system is not performing adequately, and its poor performance may be regarded as a cost (though not necessarily, of course, a simple monetary one) which we may call a misbehaviour cost. The curve A represents the uncontrolled misbehaviour cost of a system which is going

wrong. Before action to rectify the system can be taken, the system's controller must know that is it misbehaving, and for this there needs to be some recognisable signal indicating unacceptable performance. This may give perfect knowledge of bad performance, or else indicate the probability of it. Generally, the signal will only be monitored when the system's behaviour reaches a particular threshold level, T, at time t_1, after its behaviour has ceased to be satisfactory. In the following discussion it will be assumed that monitoring the system has insignificant costs, although the way of handling these costs if they are significant should be fairly clear by the end of the section. At t_1 the controller may leave the system uncontrolled in the hope that it will quickly reform itself (in general he will not know that the system will generate costs following curve A), or he may control the system by initiating some action. Control will not generally be immediate; the system will instead slowly reform, following a controlled misbehaviour cost curve B which reaches zero at t_2. t_2 may be called the (gross) response time of the system to control. This may be broken into t_1, the controller's response time and $t_2 - t_1$ the system's response time. Because control is not immediate, it will involve some control cost, represented by the shaded area under B. The difference between this cost and the cost represented by the area under A is the benefit of controlling the system. At t_2 the controller may have a number of options open to him; he may be able to place the system in a number of states, all of which are acceptable. If his choice leads to unexpectedly bad behaviour from the system, then this may be controlled according to the same schema as before.

At an intuitive level it seems fairly clear that a system is easy to control if it has a low response time and low controlled misbehaviour cost, i.e. if it can be rectified quickly and cheaply. These two features may be related to flexibility as normally conceived, that is as a function of

the number of options open to the decision maker, in a straightforward way. Consider the two systems represented in fig. 2. The controlled misbehaviour costs of both systems is the same (the shaded areas are equal), but system S* is rectified by t_2*, which is sooner than t_2 when S is rectified. This is the only difference between the systems. At t_2* the controller will be able to select one of a number of possible acceptable states of S*, but as time goes on some of these options may become closed to him. The power of time alone to close options is, of course, notorious. The controller, therefore, has at least as many, and possibly more, options at t_2* than he has at t_2. Indeed, in general, it will be possible for the controller to postpone his decision from t_2* to t_2, so that if there is so much as one other option at t_2*, then there are more options at t_2* than at t_2.

Now consider the two systems represented in fig. 3. They are exactly alike, except that S^+ has lower controlled misbehaviour costs than S. The response times of the systems are equal. If each system is supposed to make a profit (though not, of course, a simply monetary one) over the long term, then at t_2 the controller may only choose states of the system which yield an adquate profit. The higher the system's controlled misbehaviour costs, the larger this profit must be in order to meet it. Hence, there will be no more acceptable states of S to choose from at t_2 than there are acceptable states of S^+ to chooose from at t_2. The number of options open to the controller of S^+ at t_2 is as least as great, and possibly greater, than the number of options open to the controller of S at the same time. In general, of course, the controller will not know the profit derivable from each of the options open to him at t_2, but he will be able to rule out some options as clearly inferior. Any option about the state of S^+ at t_2 which is so ruled out will have a corresponding option about the state of S at t_2 which will be likewise ruled out. But because of the difference in controlled mis-

behaviour costs, the reverse will not be true. There are, therefore, at least as many acceptable states of S^+ at t_2 than there are acceptable states of S at t_2.

It can be see, therefore, that a system which is easy to control is one which leaves the system's controller with many options to choose from. Investing in an easily controlled system is to keep one's options open.

The reverse of this, that to keep one's options open is to invest in an easily controlled system, must now be shown. In a sequential decision:

(i) The payoff to the decision maker is a function of decisions made at various times and states of the world over which he has no control.

(ii) For at least one time, what courses of action are open to the decision maker at that time is a function of his earlier decisions.

(iii) The decision maker has less than perfect knowledge about states of the world.

Imagine a decision maker faced with a sequence of decisions $d_1 \ldots\ldots d_n$ where each decision involves setting a number of decision variables from $\checkmark_1 \ldots\ldots \checkmark_m$. Let the decision maker set $\checkmark_i$ to the value $\checkmark_i^1$ at time t_j. Why should he want to alter this value to some other, $\checkmark_i^{11}$ at some later time t_k? Unless from sheer capriciousness, this can only be because what the decision maker has learned in the time shows him that he will be more likely to receive a greater payoff if $\checkmark_i$ is $\checkmark_i^{11}$ than if it is $\checkmark_i^1$. We can, therefore, see the decision maker as being in control of a system of decision variables, the aim of control being to maximize his payoff. The decision sequence is flexible (at least on $\checkmark_i$) if the setting of $\checkmark_i$ to a particular value at one time restricts what values it may have at later times only little, or not at all. If this is so, then the system of decision variables must have a low response time for changes in the value of $\checkmark_i$.

If it is expensive to change the value of v_i, and if the aim is a long term profit, then some combinations of decision variable settings will be ruled out as insufficiently profitable if v_i is changed. This, of course, means that setting v_i restricts later decisions. Flexibility is, therefore, achieved if changing the values of decision variables is inexpensive. Hence, flexibility in a sequential decision is attained if decision variables can be revised quickly and cheaply. This is to say that the system of decision variables has a low response time and low controlled misbehaviour cost, but these are exactly the requirements for a system which is easy to control.

It may be concluded, therefore, that flexibility in decision making may be viewed in two equivalent ways; as a reflection of the ease of controlling a system, and as an indication of the number of options open to a decision maker. Any problem concerning flexibility may, therefore, be cast in either mould, in theory without loss. In practice, however, conceiving it in the traditional way tends to hide what the control view brings into prominence - the crucial importance for real world decision making of how quickly change may be made. When considering the number of options open to a decision maker, one naturally conceives of a sequence of decisions which have to be made at some previously determined time. When considering control, however, a previously hidden question immediately presents itself - what gains in control can be obtained by reducing the time interval between decisions, and what will these gains cost? This should become clearer from the discussion which follows.

(b) The Measurement of Flexibility

In this section the various measures for flexibility which have emerged in the literature will be discussed. It will first be shown that

the measures proposed to date all agree on giving decisions about the control of an easily controlled system a high flexibility. This may be done very briefly, as it should come as no surprise after the discussion of the previous section.

The simplest measure of flexibility is proposed by Merkhoffer[15] and also by Marschak and Nelson[16] and is based on the size of the choice set. If D and D^1 are two alternative sets of options for a decision, and if D is a proper subset of D^1, then the flexibility of D^1 is greater than the flexibility of D. It was observed in the previous section that the number of options open to a decision maker who controls a system is increased as the response time and controlled misbehaviour cost for the rectification of the system decrease, or as the system becomes easier to control. Using the subset measure, it can be said that the flexibility of the decision sequence increases as the system becomes easier to control.

Pye[17] suggest that flexibility is determined by the uncertainty of the decision maker about what his future moves will be, this uncertainty being measured by the entropy of his probability function over his future options, U, given by

$$U(p_i) = \sum_i p_i \log_e p_i$$

In the decision represented by fig. 4, the flexibility retained if a is adopted is the uncertainty in the decision maker's estimate of the probability distribution

$$(\text{prob}(c_i \text{ is chosen}/a))_{i=1}^{M}$$

Flexibility, measured by the entropy of this distribution is maximal if

$$\text{prob}(c_i \text{ is chosen/a}) = \frac{1}{M} \quad \text{for all i.}$$

i.e. if, when chosing a, the decision maker believes that he is equally likely to chose any one of his next possible moves. The maximal value is $\log_e M$. If b is chosen, maximal flexibility of $\log_e N$ occurs when

$$\text{prob}(d_i \text{ is chosen/b}) = \frac{1}{N} \quad \text{for all i.}$$

If $N > M$, then b represents the more flexible move.

In fig. 2, it should be clear that the maximum flexibility, so measured, is greater for S^* than for S since there are more options open at t_2^* than t_2. Pye accommodates questions of cost by assuming the decision maker to attribute zero probability to his choice of any option which leads to an inadequate payoff. In fig. 3, the controlled misbehaviour cost of S^+ is less than that of S, and if these costs must be recouped by the choice of profitable options at t_2, then there will be at least as many options yielding an adequate profit for the controller of S^+ as there are for the controller of S. The first controller will have to give no more of his options a zero probability than does the second controller. Using Pye's measure, it follows that the maximum flexibility attainable for S^+ is not less than the maximum flexibility attainable for S.

Rosenhead et al[18] have suggested robustness as a measure of flexibility. One initial decision d_i is to be selected from the set of possible initial decisions D, with the aim of realizing one of a set S of alternative plans in the long run. Any initial decision will restrict the set of attainable plans to S_i, some subset of S. If some subset $\bar{S}$ of S is considered

acceptable according to some criterion (usually concerning the profitability of plans), some subset $\bar{S}_i$ will be attainable after the initial decision d_i. The robustness of the initial decision is then defined simply by

$$r_i = \frac{n(\bar{S}_i)}{n(\bar{S})}$$

where n(S) denotes the number of elements in S. An argument may be constructed along exactly the same lines as the one for Pye's measure to show that the robustness of the decision to control S^* in fig. 2 is at least as great as that of the decision to control S, and that the same is true of S^+ and S in fig. 3.

Dissatisfied by the subset measure discussed first, Marschak and Nelson[19] have proposed two other measures. Both have boundedness conditions which the authors admit to be extremely restrictive and, indeed, it does not seem possible for them to be applied in the same straightforward way as the other measures to the problem of controlling systems. Nevertheless, the idea underlying the measures fits with the view of flexibility as ease of control. The intuitive idea which the measures try to formalize is that the flexibility of a sequence of decisions increases as the payoff becomes less and less sensitive to future signals. An example discussed by the authors, due to Stigler,[20] is the choice between two technologies having average cost curves shown in fig. 5. Plant 1 has the advantage over plant 2 within the region $X_1 - X_2$, but outside this area this is reversed and plant 2 gives the best performance. Since the profit from plant 2 is less sensitive to future price signals than the profit from plant 1, we arrive at the intuitively satisfactory conclusion that plant 2 is more flexible than plant 1.

If a system is easy to control it can respond to signals which reveal change to be desirable quickly and cheaply. The long-run payoff from such a system will, therefore, be less sensitive to such signals than the payoff from a system which is more difficult to control. It may be useful to apply

this to Stigler's example. If the aim of the production system in each case is to maximize profit, then if the output of plant1is between X and X_2 it bears no misbehaviour cost. The misbehaviour cost of plant 2 in this region is the lower profit than that obtainable from plant 1. Outside X_1 - X_2, however, the situation is reversed. If output is likely to fall into a wide region beyond X_1 - X_2, then plant 2 must be counted the more easily controlled plant (assuming the plants to have the same response time for changes in output).

It can be seen, then, that the various measures for flexibility proposed in the literature agree well with the claim that to invest in easily controlled systems is to keep options open, and that to keep options open is to invest in easily controlled systems. Viewing flexibility as the ease with which a system may be controlled does, however, suggest a new measure which may prove of greater practical utility than those discussed above. Before developing this suggestion it will be helpful to consider an example of control, so that an over naive view may be guarded against. This is the view that any change occurs according to one fixed response time and controlled misbehaviour cost. The example concerns the removal of lead from petrol and altering engines to accommodate the change.

Most existing motor car engines are designed with a high compression ratio so that they demand petrol of high octane number. This is presently achieved by adding lead compounds to the petrol, but growing concern about the health effects of lead in the environment may make it desirable to severely reduce lead levels in petrol. One way to respond to the call for such a reduction is to alter engines so that they will run on lower octane fuel, which may be produced with very little or no lead. The country's car population represents a very large fixed asset which is depreciating only relatively slowly; the number of new cars in one year being about 10% of the total car population. A natural step would, therefore, be to require all new cars to be fitted with engines which can run on low lead or lead

free petrol. In this way, after 10 years the problem of lead emissions from cars will have been overcome.

It is possible to speed up this process, but only at great cost. If we wished to overcome the problem in 5 years, car engine output would have to be doubled for this period, after which it would return to its previous level. This would, of course, require enormous investment in new plant, which would, moreover, have to spread its capital costs over only five years. This is such a large undertaking that considerable time is required to plan and build the new plant, so that there is a limit to the speed at which existing engines can be replaced. Beyond this limit we may count the cost of replacement as infinity. This would seem to be a typical case of change, and so indicates that in general, the rate of change may be increased, but only by incurring costs which eventually rise to infinity (see fig. 6).[21]

What does this example tell us about controlling a system? The system is the car population of the country, and its unacceptable performance is pollution of the environment by lead. This may be reduced at various rates, i.e. with various response times, but there is a minimum response time beyond which the cost of further reduction is infinite. The situation is summed up in fig. 7. In each case, controlled misbehaviour costs decline with response time, since the lower the response time, the quicker the pollution level falls to acceptable levels. The chief lesson, then, is that many controls are generally available, each with its own controlled misbehaviour cost and response time, and that reducing these involves bearing costs, which we may term control costs. Deciding what control to adopt involves trading off low response time and low controlled misbehaviour costs with high control costs.

With this in mind, let us denote uncontrolled misbehaviour cost as UMC and the j th controlled misbehaviour cost as (CMC)j, letting C_j represent its associated control cost. At any time t_i after the system's gross response time, t_j, the saving due to the imposition of the control is given by

$$\text{SAVING} = \int_o^{t_i} \text{UMC}\, dt - \int_o^{t_j} (\text{CMC})_j\, dt$$

So that

$$\text{SAVING} - \text{COST} = \int_o^{t_i} \text{UMC}\, dt - \int_o^{t_j} (\text{CMC})_j\, dt - C_j$$

To maximize this, we should minimize

$$\int_o^{t_j} (\text{CMC})_j\, dt + C_j$$

i.e. we should impose that control which gives the lowest sum of controlled misbehaviour cost and control cost. In general, as the integral increases, i.e. as the controlled misbehaviour cost increases, then C_j, the control cost, decreases. There is, therefore, a tension between the two terms in our final expression. We may think of this as representing the tension widely recognised to exist between flexibility and control cost. This suggests that we may measure the flexibility of a system by

$$(\text{MIN} \left[\int_o^{t_j} (\text{CMC})_j\, dt \right])^{-1}$$

i.e. by the reciprocal of the minimum controlled misbehaviour cost which the system can show.

(c) Cost, Flexibility and the Scale of Production

Any theoretical account of flexibility should shed some light upon the tension between cost and flexibility. The tension is nowhere clearer than when the decision concerns the scale of production. Economies of scale often favour very large production units, but these can be very difficult to control and manage, especially where there are serious uncertainties about future input prices, processing capacity and costs, and demand. This section considers the opposing pulls of cost and flexibility in such a situation. No claim will be made to deal with a real decision problem, however, for this would introduce many details special to the example which would tend to obscure the general features of flexibility and its role in decision making.

Any production process may be thought of as a three-part system where inputs are processed to give output at a level which satisfies demand (fig. 8). A flexible process is, therefore, one which is flexible with respect to changes in input prices and availability, changes in processing and demand for output. Anything which has an effect on the production process, strike, revolution, war, environmental legislation, fashion, safety considerations and so on, all do so by affecting inputs, processing, and/or demand for output. A process which is flexible can, therefore, cope better with changes from all these causes than an inflexible system. This is useful because rather than looking at flexibility to cope with war, revolution and so on, all that needs be considered is flexibility with respect to input, processing and demand for output. Of these, the last is most easily treated at an abstract level and so it is this on which we shall concentrate.

There are four reasons why a system of small production units is generally more flexible than one of large units.

(i) The Step Effect

If O_S and O_L are the annual capacity of one small and one large production unit, there are more numbers equal to IO_S than there are equal to IO_L where I is an integer. The decision maker, therefore, has more options open to him in deciding the capacity of a system of small units than he has over the capacity of a system of large units. As expected from the previous section, this means that the former system is easier to control. There will generally be a mismatch between capacity and demand, and this may be regarded as the controlled misbehaviour cost of the system. The mismatch arises because demand changes continuously, whereas capacity varies in a stepwise fashion. By reducing the height of the step, the mismatch, and hence the controlled misbehaviour cost of the system, may be reduced (see fig. 9). This is in accord with the measure of flexibility proposed above. The minimum controlled misbehaviour cost of a system of small units will always be less than or equal to the minimum controlled misbehaviour costs of a system of large units.

For simplicity, assume that $O_S = \frac{1}{2} O_L$. For a given demand D;

minimum controlled misbehaviour cost large system = $MIN/IO_L - D/$

where I is an integer. Likewise;

Minimum controlled misbehaviour cost small system = $MIN/IO_S - D/$

$= MIN/\frac{1}{2}IO_L - D/$

But

$$MIN/\tfrac{1}{2}IO_L - D/ \leqslant MIN\ /IO_L - D/$$

i.e. by the measure above;

flexibility of small system $\geqslant$ flexibility of large system.

(ii) Lead Time

In trying to match demand and capacity, overcapacity can often be eliminated with nearly zero response time, whatever the size of the units declared redundant or mothballed. If new capacity is required to remedy undercapacity, however, response time may be considerable and is generally a function of the size of the units of production. Small units generally have a shorter lead time than large ones, so the response time of a system of small units to undercapacity is generally less than that of a system of large units. This generally means that the minimum controlled misbehaviour cost, again measured by the mismatch between capacity and demand, is lower for the system of small units. On the measure proposed earlier, this gives us the expected result that the system of small units is more flexible in dealing with undercapacity than the system of large units.

The above discussion, however, contains a simplification. Instead of talking of one lead time for the commissioning of a unit of production, it is best to recognize that the whole process can be speeded up by incurring larger capital costs (see fig. 6). This is best seen by considering an example. Suppose that demand for a product suddenly increases by an amount D in excess of capacity. To the firm this represents foregone profit, and so we may think of this as the system's misbehaviour cost. Let the annual cost of not meeting this extra demand to the firm be C, supposed constant. New capacity may be added by adding n small units of capacity $\frac{D}{n}$ or by adding $n/2$ large units of capacity $\frac{2D}{n}$. The numbers chosen here eliminate any advantage in terms of flexibility given to the system of small units from the step effect, so that the effects of lead time may be considered in isolation. If capacity and demand are matched only after x years, then the controlled misbehaviour cost of the system is the discounted value of C over x years. The minimum time taken to build the appropriate number of small units will be less than that to

build the large units which are required, so that the minimum controlled misbehaviour cost is less for the system of small units. According to the measure above, therefore, the system of small units is more flexible than that of large units.

For both systems;

Cost of adding capacity D = controlled misbehaviour cost + control cost. The control cost is, of course, the cost of adding extra capacity. For the small units;

$$\text{Cost of adding capacity } D = C^i + \left[K_s^i + V_s^i\right]$$

where C^i is the controlled misbehaviour cost = C discounted over period O to i; K_s^i is the/capital cost of n small units built/over time i; and V_s^i is the variable cost of the small units discounted over their lifetime to time O. Using the same nomenclature for the system of large units;

$$\text{Cost of adding capacity } D = C^j + \left[K_L^j + V_L^j\right]$$

The small plant provides the better system if and only if:

$$\left[K_S^i + V_S^i\right] - \left[K_L^j + V_L^j\right] < C^j - C^i$$

The left hand side of the inequality is positive if there are scale economies of large plant built in j years over small plant built in i years. The right hand side is the difference in controlled misbehaviour costs for the large and small systems. The inequality, therefore, represents the opposing pulls of scale economies and misbehaviour costs. A system of small units will generally bear lower misbehaviour costs for any pattern of changing demand than a system of large units, but the large units may be cheaper to operate, i.e. may have lower control costs. Thus, the essence of decision making under the mind of uncertainty where

flexibility is important, is the trading between control costs and misbehaviour costs. A flexible system will have higher control costs than an inflexible one, but will impose lower misbehaviour costs.

(iii) Forecasting

Because of the lead times involved in controlling many systems, monitoring often involves making forecasts about what will be required of the system at some future time. For present purposes, where we are interested in the problem of matching capacity and demand, it is often necessary to forecast future demand because acquiring new capacity takes time. The better the forecast, then the better the monitoring of the system and the lower its misbehaviour costs, as measured by the mismatch between demand and capacity. If two systems differ only in that the forecasts made about future demand for one system's output are more accurate than those made for the other, then the cumulative minimum controlled misbehaviour cost for the former must be less than that for the latter. According to the suggested measure above, therefore, the former system is the more flexible.

One way of acquiring better forecasts is to employ a system which has a shorter lead time. This, of course, favours systems of small units of production. Here, forecasts need to be made over shorter periods, and so may be expected to be more accurate (or, at least, no less accurate). Small unit production systems, therefore, have a third source of flexibility. As well as the step effect and the primary effect of lead time, lead time also enhances flexibility through improving forecasts.

(iv) Learning

If many small units are constructed instead of a few large ones, learning about the units is likely to be more rapid, even if the lead times for both units are the same. The effect is, of course, enhanced if

the lead time for the small units are less than that for the large units. A system of small units is, therefore, likely to operate more efficiently than a system of large units, resulting in lower misbehaviour costs. For this reason, the system of small units is likely to be more flexible than one consisting only of large units, according to the measure proposed earlier.

The importance of learning will, of course, differ greatly from one system to another; so much so that little can be said at the general level. Learning can, however, be viewed as the very essence of flexibility; for a flexible system, as characterized here, is one whose behaviour can be learned cheaply and quickly. We learn from our mistakes, and a flexible system is one where mistakes can be remedied cheaply and swiftly.

(d) Use of Scenarios in Estimating Flexibility

It is helpful to view scenario analysis in terms of the flexibility of the systems being investigated. A simple scenario may be thought of as a hypothetical forecast of form. If $D_1 \ldots D_k$ and $S_1 \ldots S_\ell$, then $O_1 \ldots O_m$ where $D_1 \ldots D_k$ are decision variables, $S_1 \ldots S_\ell$ state variables and $O_1 \ldots O_m$ outcome variables. A set of scenarios is a set of such hypothetical forecasts having different values for the decision and state variables. In a proper use of scenarios, the analyst is concerned with the sensitivity of outcome variables to changes in decision variables. For convenience, the outcome variables may be represented by a single variable having just two values, acceptable and unacceptable. Letting D_i^j be the j th value of the decision variable D_i, the outcome is more sensitive to D_i^j than to D_i^h if there are fewer satisfactory outcomes given $D_i = D_i^j$ than satisfactory outcomes given $D_i = D_j^h$. If the aim of the decision maker is to achieve an acceptable outcome, then setting D_i to D_i^h gives him at least as great a chance of this than setting it to D_i^j. The point of the exercise, of

course, is that future values of state variables are not known, so that chosing D_i^h in favour of D_i^j is making a decision which is more robust to whatever future events come to pass.

An example may help to clarify the matter, and so we may consider the Marshall report[22] which addresses itself to the problem of what energy R&D programmes the UK should invest in. The problem is succinctly put:

> Unfortunately, new technologies take a long time to develop and apply; the period from the inception of an initial R&D programme to successful commercial application tends to be measured in decades rather than years. It is vital, therefore, to identify those technologies which seem most likely to be important in the future and to consider how best we can ensure that they will be available to us when required. In addition, because the future is so uncertain, we must seek to keep our options open over a wide range of technologies, any of which might become important under particular circumstances.....
>
> Our ultimate objective will be to formulate a strategy sufficiently robust to cope with the immense uncertainty of the future. This calls for scenario analysis: the United Kingdom's future technology requirements in energy are inextricably linked to future events, most of which cannot be forecast with any degree of certainty; yet one of the main purposes of an energy R&D strategy must be to ensure that future policy makers are equipped with appropriate energy technologies to implement their policies whatever shape the energy economy takes. Therefore, it is clear that we need to adopt an analytical approach which assesses technology needs over a spectrum of possible futures, and not just on some 'central' view.
>
> Of the many techniques available, we have adopted an approach in which the contribution of each technology is assessed over a set of narrowly defined 'snap-shot' views of the future - these are termed 'scenarios'.

The exposition of the scenario analysis does not follow the pattern discussed above and, indeed, is somewhat confusing (e.g. no clear division between state and decision variables is made). Nevertheless, it is easy to restructure the analysis along the lines suggested earlier.

The problem is to chose a set of R&D programmes which will be robust in the sense of allowing the future UK energy system to be in an acceptable state. For the sake of simplicity, we may think of an acceptable state as giving a balance between primary energy, energy carriers and useful energy.[23] Each scenario, therefore, has the form

If R&D decisions R .... R_n are made at times $t_1 \ldots . t_k$, and if S_1 .... S_n, then there will be (will not be) a balanced energy system.

S_1 .... S_n are the state variables of the scenario and Marshall presents seven scenarios given in outline in figure 10.

The economic factors in the scenario are first used to calculate useful energy requirements in the three sectors, building, industry and transport. By means of a simple energy model requirements for the heat supplied by various energy carriers (solid, liquid and gaseous fuels, electricity, district and direct heat) are calculated. The primary energy in the form of coal, oil, natural gas, nuclear power and alternatives needed to meet this requirement for carriers is then calculated. Thus calculated, there is obviously a balance between primary energy, energy carriers and useful energy. To actually achieve this balance, however, requires new technologies to be available in the three areas; primary energy exploration and provision, energy conversion and distribution and energy utilization. What technologies are required for the scenario are, therefore, identified and timing of the technologies estimated.

When all the scenarios have been treated in this way, the sensitivity of obtaining a balanced energy system to the R&D decisions is estimated. Some R&D programmes are involved in nearly all the scenarios, so that the outcome is not sensitive to them, whilst others figure in few scenarios, so having a greater sensitivity. The table below (fig. 11) gives the contributions of the technologies and their overall importance to ensuring a balance in the UK energy system. Fig. 12 gives the R&D priorities identified to provide these technologies.

It is now time to consider how this relates to flexibility and its measurement, first considering flexibility in the standard way, as a property of sequential decisions. As stated quite clearly in the Marshall report, the need for scenario analysis arises because the scale of the decision maker's ignorance is so great that he places a high premium on keeping his future options open; i.e. a premium on making decisions which are highly flexible. Engaging in an R&D programme such as those investigated by Marshall is essentially buying the future option to implement the technology involved should it be required. The aim here is to achieve a balanced energy system, and there are many ways in which this balance might be achieved. Many of these involve nuclear technology so that adopting R&D programmes for this technology means that many future ways of balancing the system are possible. Opting for a nuclear energy R&D programme, therefore, places far fewer constraints on what other decisions may be made to maintain a balanced energy system than not adopting such a programme. If, on the other hand, a particular technology figures in only one scenario, then not investing in an associated R&D programme closes only a few future balanced energy systems. A limited R&D budget may mean that the programme competes with a nuclear energy programme, but the latter is clearly more important in keeping open future options.

Putting the matter formally: 'nuclear energy programme' is a two valued decision variable (this, of course, is a simplification but the general point still holds) having values 'yes' and 'no'; there are more balanced energy systems across the scenarios which are possible given the value 'yes' than given the value 'no', hence 'yes' is a less sensitive value for 'nuclear energy programme' than 'no'. In general, insensitivity may be taken as a measure of the decision's flexibility.

Flexibility may also be seen as a property of a system, reflecting the ease with which it may be controlled. In the case considered by Marshall, for example, the system is the UK energy system and control is the maintainence of balance between primary energy, energy carriers and useful energy. Misbehaviour costs arise when this balance is not achieved, and can be lowered by the introduction of technologies. These technologies have, however, very long lead times so that controlled misbehaviour costs would be extremely high if they were developed only when seen to be necessary. Response time may be reduced by investing in R&D programmes before the technologies are known to be required. In this way, controlled misbehaviour costs are reduced so that, according to the measure proposed earlier, the effect of investing in R&D is to increase the flexibility of the energy system. This is accompanied by an increase in the control costs, because R&D programmes must be paid for. The decision maker must, therefore, trade off lowering the energy system's controlled misbehaviour costs with increasing the control costs of the system. Investing in, say, nuclear power R&D is shown by scenario analysis to be effective in reducing misbehaviour costs, whilst investing in a technology which figures in very few scenarios is likely to be much less effective. Scenario analysis, therefore, gives guidance as to the increase in the flexibility of the system purchased in various ways.

4. Flexibility and Costs and Benefits

We have seen how a critical debate can serve to guide us about the likely costs and benefits which will flow from a decision and we have also seen how the need for flexibility arises and how it may be measured. Flexibility must, in general, be paid for - the greater the flexibility of a system the greater its cost or, formally, the lower the system's minimum controlled misbehaviour cost, the greater the control costs incurred. It is, therefore, of the essence of decision making under the kind of ignorance which places a premium on flexibility to trade flexibility with cost. This section considers how this may be done.

The short answer is that the trade off between flexibility and cost can only be achieved after a critical debate. After all, where the balance should be is clearly an evaluative issue, and so must be settled by the kind of evaluative debate we have considered here and previously. To give an example of arguments against a proposal on the grounds of its inflexibility, we may consider the case against nuclear energy made by Walter Patterson. Nuclear power, like the generation of electricity by conventional means, is subject to very marked economies of scale; a station twice as large does not require twice the labour, twice the material, twice the investment, so that the cost per unit of electricity produced is less. Stations now considered economic may take ten years or more to construct, but no reliable forecast of electricity demand can be made so far ahead. Hence, according to Patterson:[24]

> Ordering the station then becomes not an act of foresight but an act of faith, founded on policy. The suppliers must then endeavour to ensure that electricity demand 10 years hence will have increased to the level warranting addition of the new station. The internal objectives of the system planners thus take precedence

over the social and economic role of the system. Planning, as conventionally understood in mixed economies, disappers; the technology takes over and reproduces itself according to its own introverted criteria.

A second factor leading to great inflexibility is that supply of electricity must be assured.[25]

The nature of grid electricity as a commodity - impossible to store, supplied by a monopoly and guaranteed available at all times - imposes severe constraints. 'Security of supply' of grid electricity refers not to the supply of fuel, or of facilities, but to the supply of electricity at the user's power points and switches. The consequent philosophy of planning by nuclear electricity suppliers diverges steadily from 'planning' as commonly understood in a mixed economy: that is, anticipating future developments and attempting to harmonize economic activities with them. Instead, electricity supply planning begins increasingly to resemble the planning-by-edict which takes place in centrally planned economies such as those of eastern Europe. As was discussed earlier, the nature of grid electricity as an essential commodity removes the ultimate sanction of bankruptcy in the event of unfulfilled plans. Public participation in planning is at least an inconvenience; public opposition to particular plans - say for the siting of a new power station - may become such an inconvenience that it may have to be administratively overruled.

A third problem, again stemming from the inflexibility of the electrical supply industry, is that reliance upon a guaranteed supply of electricity gives enormous power to the workers in the industry, who can do immeasurable damage in a matter of minutes by throwing the switches. All in all, then,

the generation and distribution of electricity form a highly inflexible system, whose inflexibility will grow if nuclear generation is ever used to provide a substantial part of the country's energy needs. Patterson would argue that this is so serious a problem, that nuclear power should not be developed.

It is to be hoped that the measures for flexibility and the use of scenario analysis for its estimation will be able to add precision to the kind of arguments used by Patterson, so that the crucial trade off between cost and flexibility may be achieved in a more well considered way in future debates about what projects to invest in.

CHAPTER 10 - FOOTNOTES

U.S. Atomic Energy Agency, Theoretical Consequences of Major Accidents in Large Nuclear Power Plants, WASH-740, Washington D.C. 1957.

A 200 MWe reactor is one which produces 200 million watts of electricity - enough to run 200,000 1 bar electric heaters. To do this it produces about 600 million watts of heat, of which 400 million watts is dumped as waste heat.

Report of the Engineering Research Institute, University of Michigan to the Power Reactor Development Corporation, July 1957.

The minutes and records of the 1964/65 WASH-740 update were made available in 1973. See H. Kendall and S. Moglewer, Preliminary Review of AEC Reactor Safety Study, Sierra Club and Union of Concerned Scientists, 1974.

See, for example, F.R. Farmer, 'Accident Probability Criteria', Journal of the Institution of Nuclear Engineers, 16, 1975 and 'Advances in the Reliability Assessment of Reactor Systems', Atom, 230, 1975, pp. 218-226; B Lister, ' Nuclear Power - the Perspective of Risk', Atom, 223, 1975, pp. 68-75.

U.S. Atomic Energy Commission, Reactor Safety - An Assessment of Accident Risks in U.S. Commercial Power Plants (Draft), WASH-1400 Washington D.C., 1974.

For a brief description of some of the more notable accidents see W. Patterson, Nuclear Energy, Penguin 1975, Chapter 6.

6th Report of the Royal Commission on Environmental Pollution, H.M.S.O., 1976, para. 176.

See U.S. Senate Congressional Record, Dec. 9th 1974 and W. Bryan, Nuclear Week, 4th April 1974.

F. Farmer, 'Advances in the Reliability Assessment of Reactor Systems', Atom, 230, 1975, p. 222.

For example, see G. Foley, The Energy Question, Penguin 1976, p. 178 ff.

For this and a summary of other criticism see A. Lovins, Nuclear Power, 2nd Edn., Friends of the Earth, 1975, Appendix 1.

See footnote 9.

For example, the Department of the Environment has recently been criticised for failing to check on the accuracy of its estimates of road demand by the simple device of counting traffic, see Report of the Advisory Committee on Trunk Road Assessment, Chairman Sir G. Leitch, H.M.S.O., 1978.

M. Merkhoffer, The Value of Information Given Flexibility, Management Science, 23, 1977, 716-27.

T. Marschak and R. Nelson, Flexibility, Uncertainty and Economic Theory, Metroeconomica, 14, 1962, 42-58.

R. Pye, A Formal Decision Theoretic Approach to Robustness and Flexibility, Journal of the Operational Research Society, 29, 1978, 215-229.

J. Rosenhead et al, Robustness and Optimality as a Criterion for Optional Choice, Operational Research Quarterly, 23, 1972, 413-431 and J. Rosenhead and S. Gupta, Robustness in Sequential Investment Decisions, Management Science, 15, 1968, 18-29.

Op. cit.

G. Stigler, Production and Distribution in the Short Run, Journal of Political Economy, 476, 1939, 305-327.

For more on this example see D. Collingridge, The Entrenchment of Technology - the Case of Lead Petrol Additives, Futures, forthcoming.

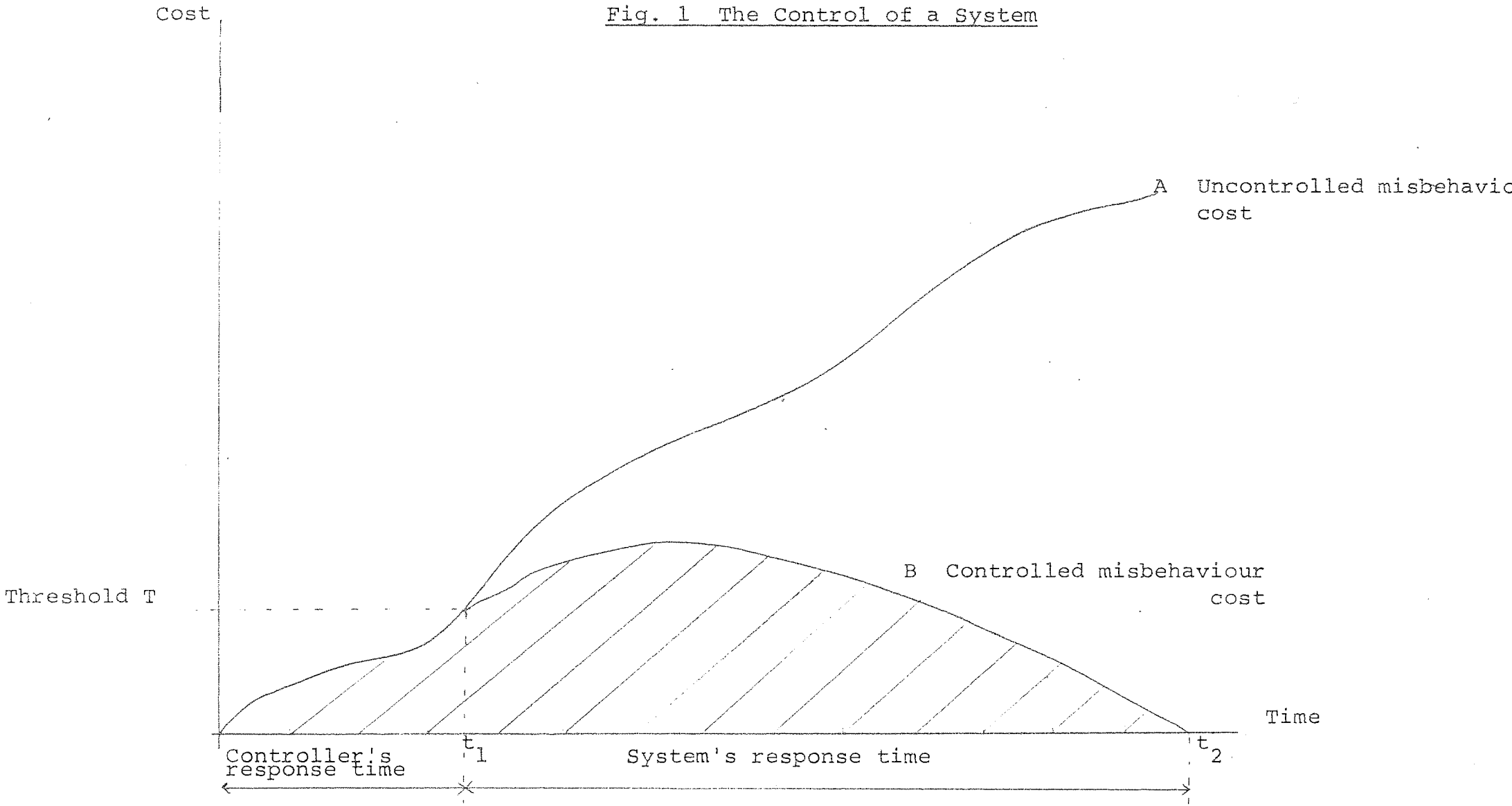
W. Marshall, Energy R&D in the UK, Energy Paper No. 11, Department of Energy, H.M.S.O., 1976.

For the meaning of these terms see Marshall op. cit. pp. 10-13.

W. Patterson, The Fissile Society, Earth Resources Ltd., 1977, p.49.

Op. Cit. pp. 90-91.

Fig. 1 The Control of a System



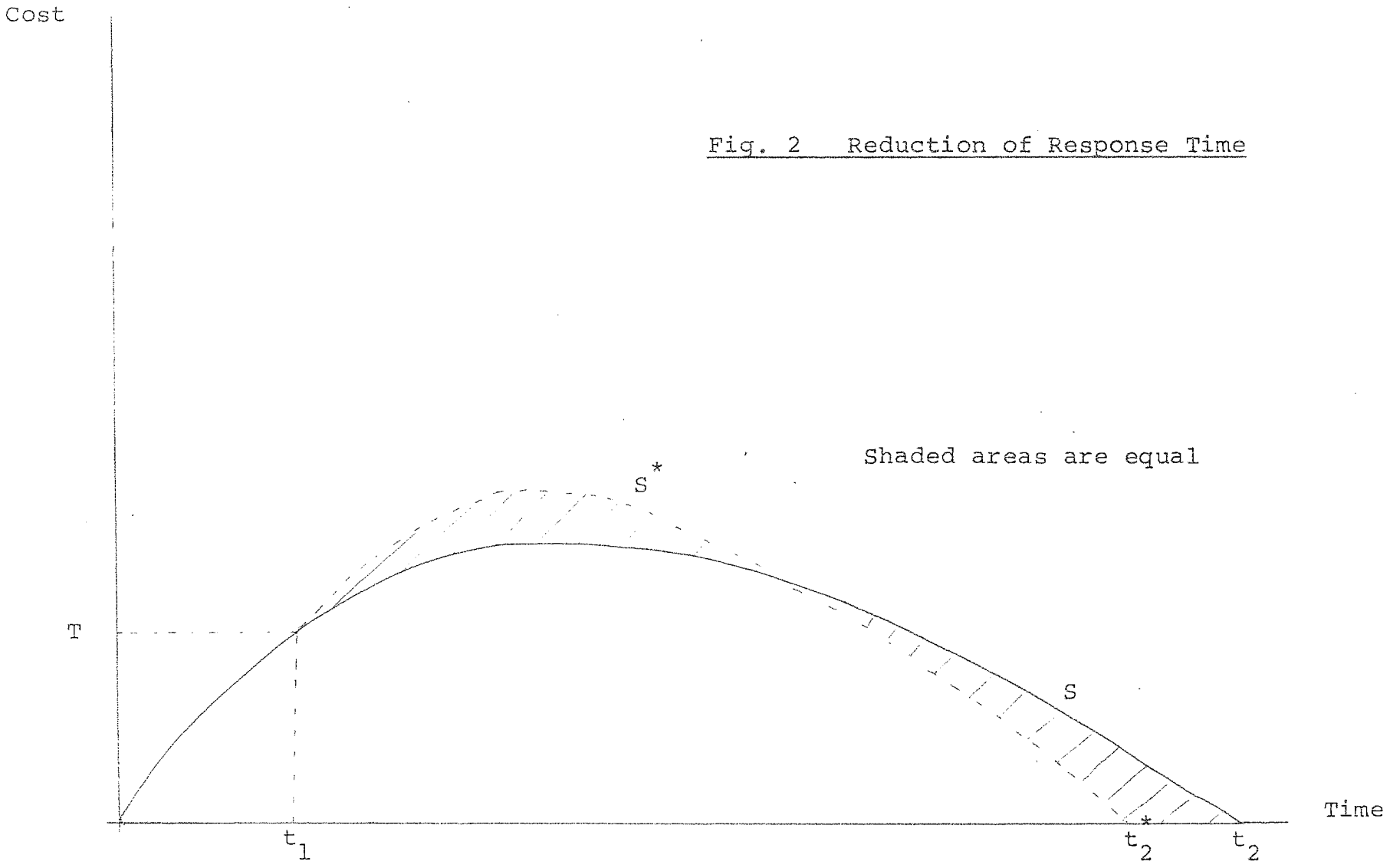


Fig. 2 Reduction of Response Time

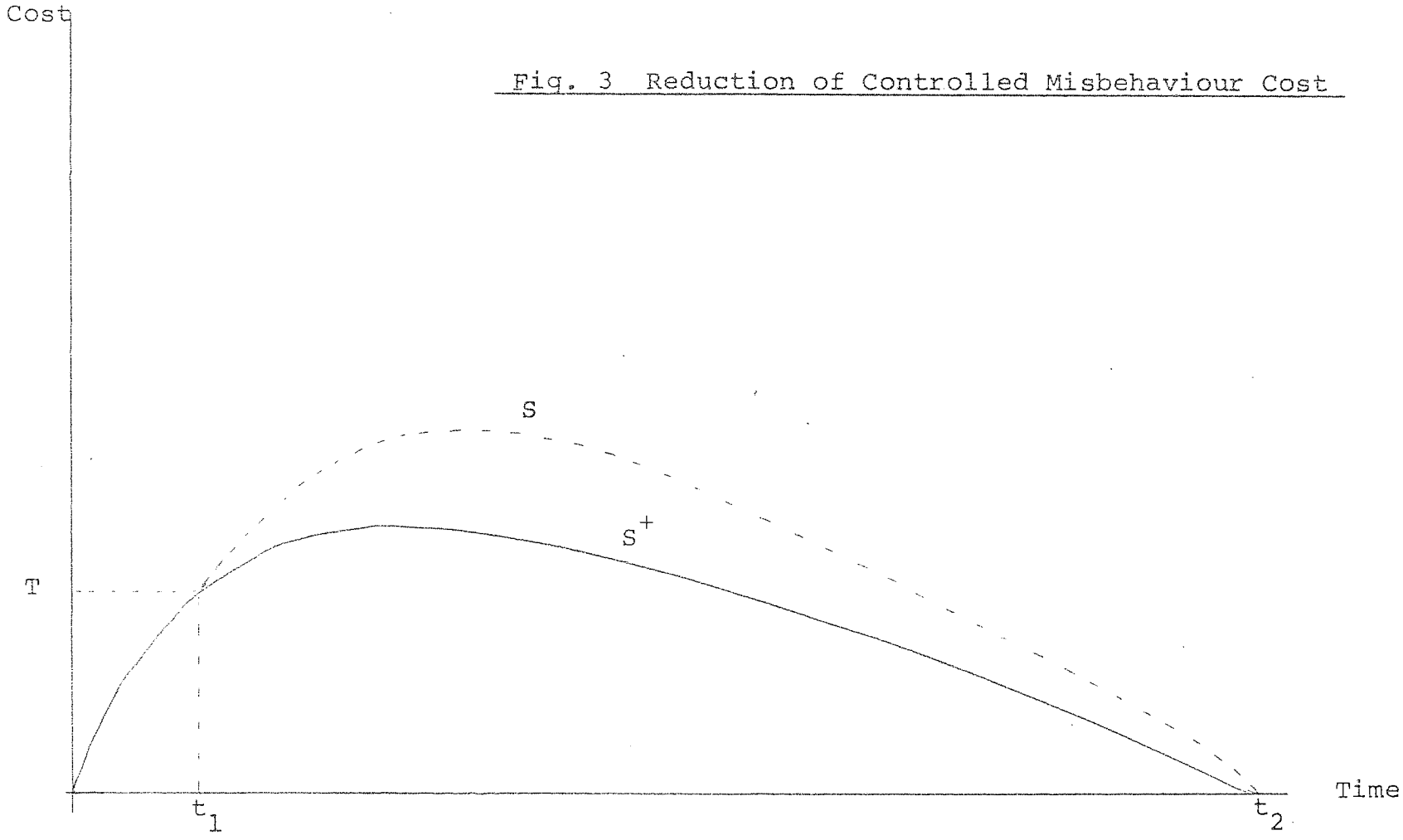


Fig. 3 Reduction of Controlled Misbehaviour Cost

Fig. 4 Decision Tree

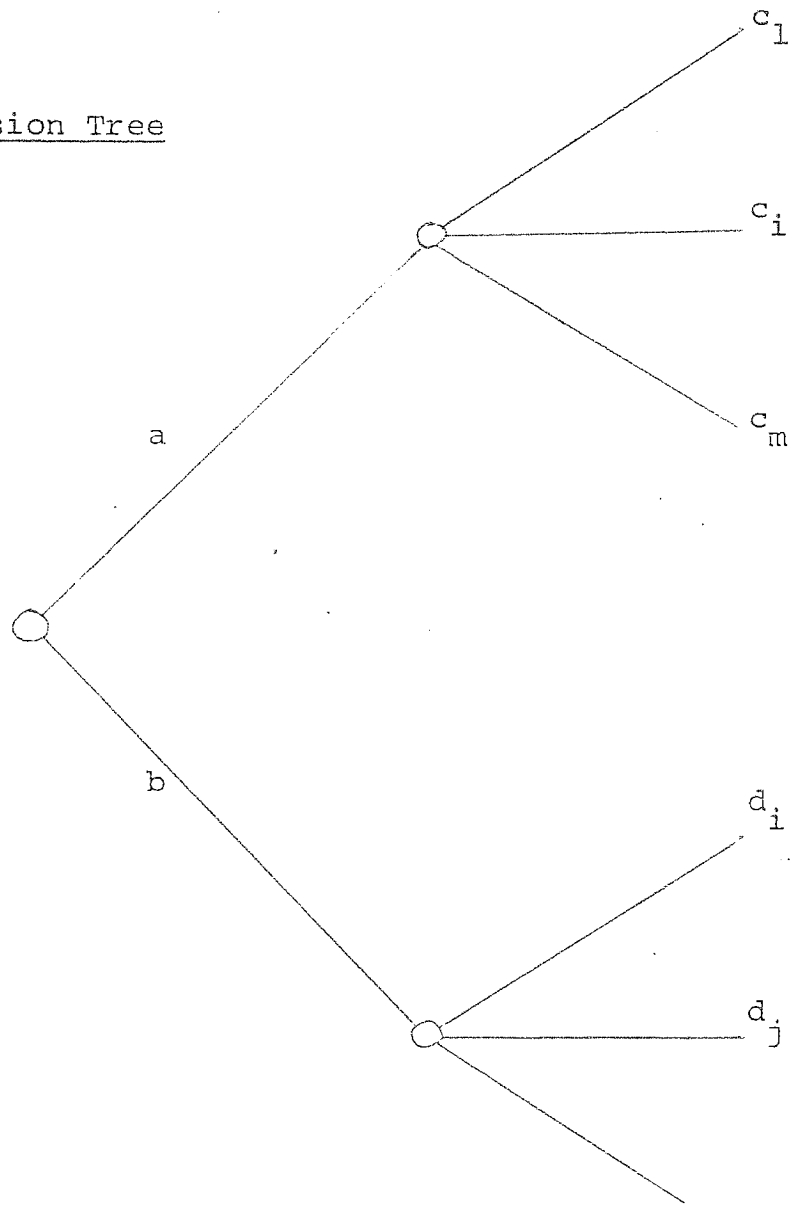
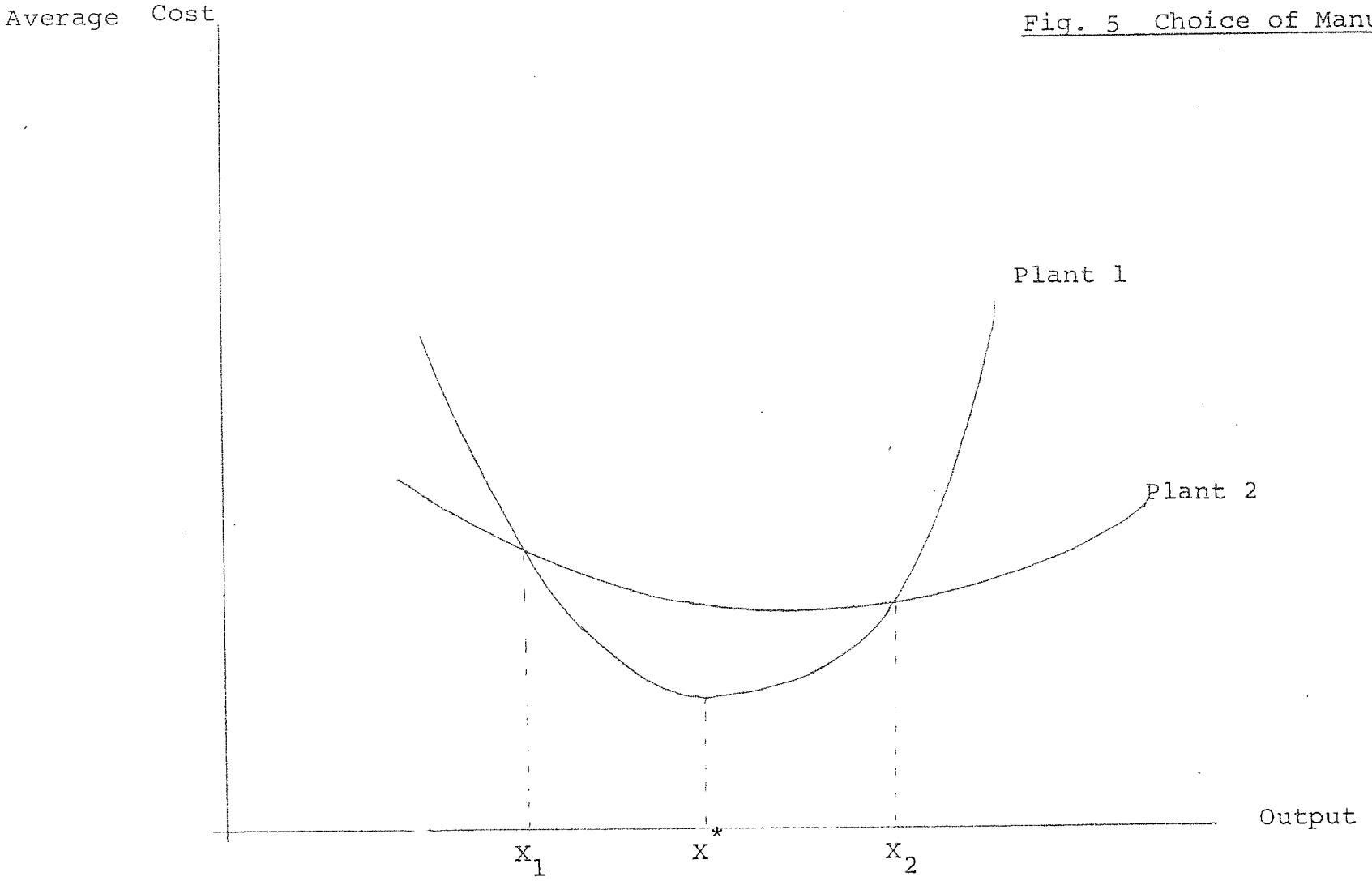


Fig. 5 Choice of Manufacturing Plant



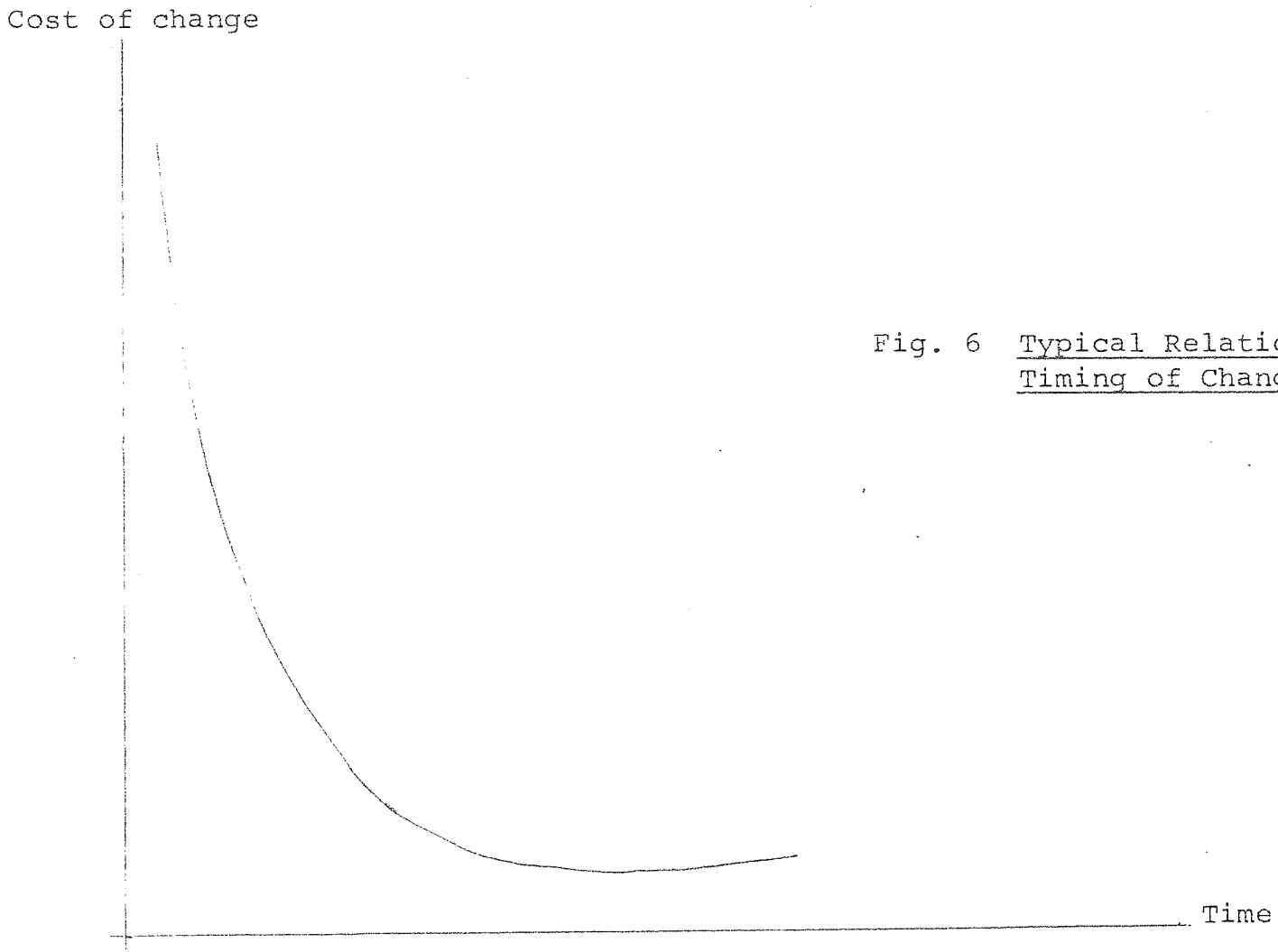
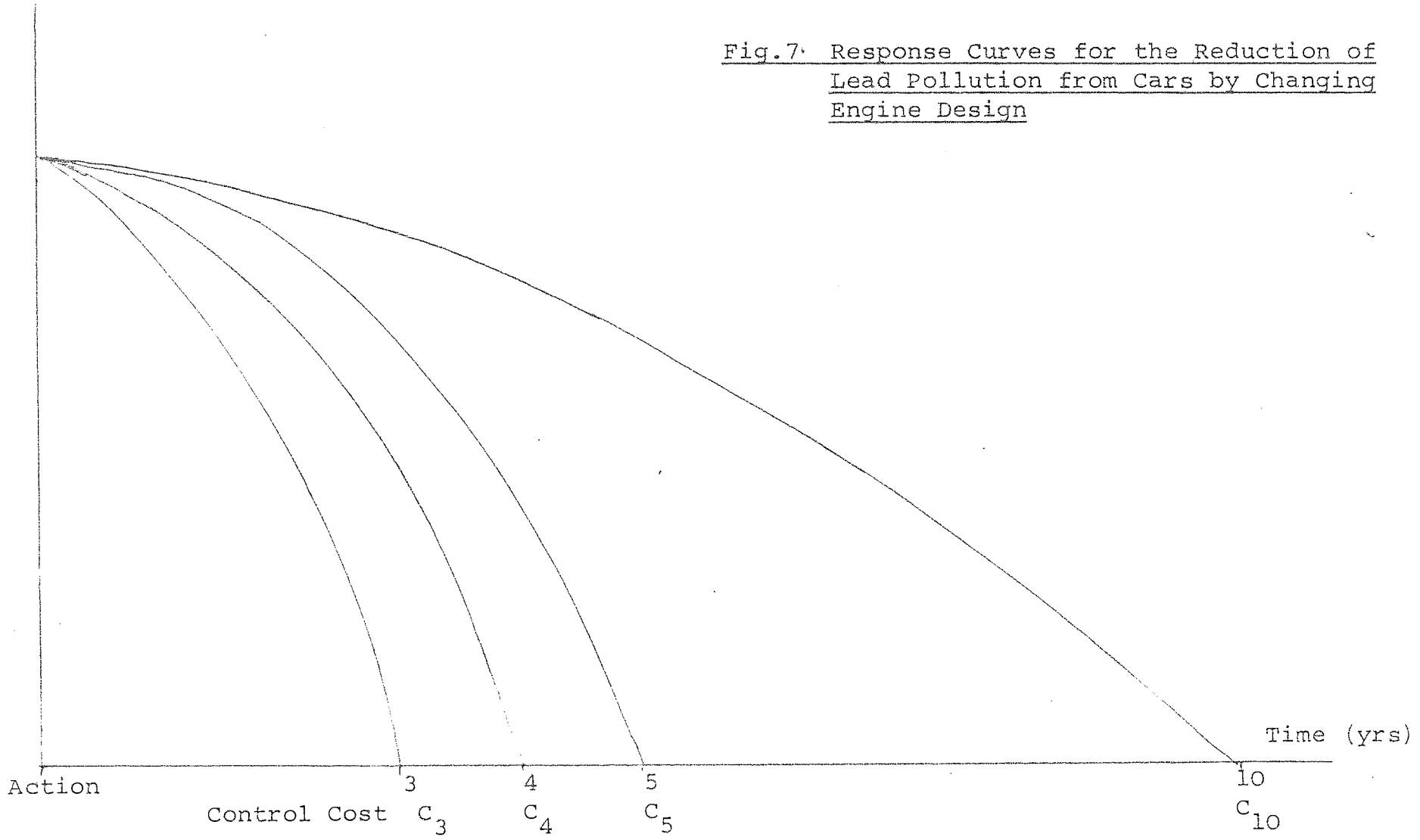


Fig. 6 Typical Relation Between Cost and Timing of Change

Fig.7· Response Curves for the Reduction of Lead Pollution from Cars by Changing Engine Design



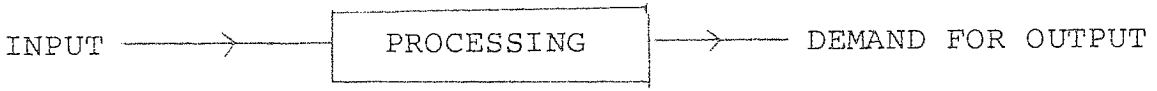
C_{10} = Design of new engines, extra retooling of plant

C_5 = C_{10} + Construction of new plant of equal capacity toexisting plant, lifetime 5y

C_4 = C_{10} + Construction of new plant of 1¼ capacity existing plant, lifetime 4yrs.

C_3 = Infinity

Fig. 8 Three Part Production System



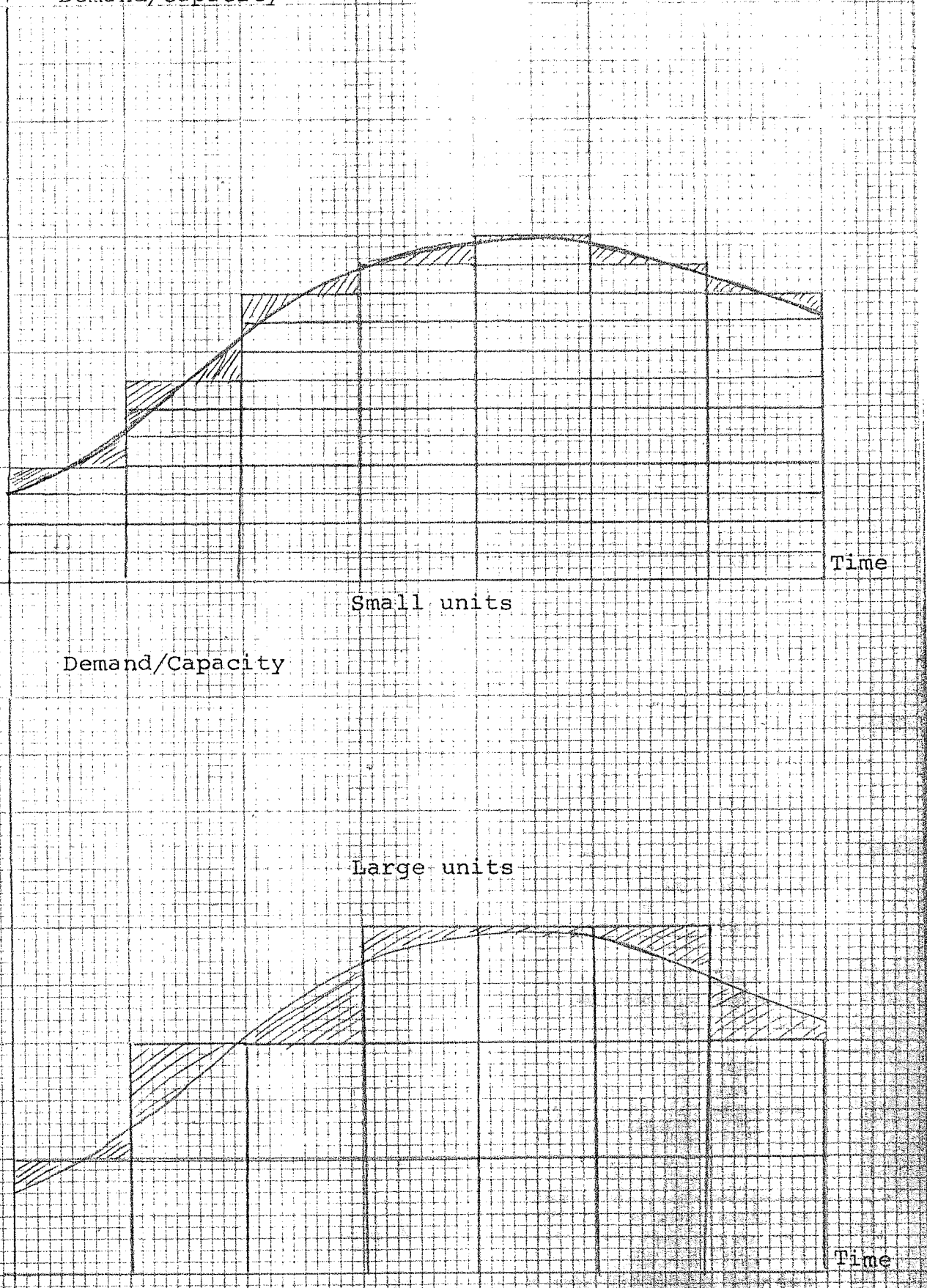


Fig.9 Effect of Size of Production Unit on Ability to Follow Demand (shaded area=mismatch between demand and capacity)

State-of-World Assumptions

e- /orld ables	Scenario 0 Trends-Continued Scenario	1 Low-Growth Scenario	2 Limit-on-Nuclear Scenario	3 High-Energy-Cost Scenario	4 Price-Transition Scenario	5 Self-Sufficiency Scenario	6 High-Growth Scenario
K Economic rowth (GDP erage nual)	Central view	Depressed	Reduced	Reduced Slightly	Central view to 1990, then reduced	Close to Central view	Buoyant
975–1990	3%	2%	2½%	2½%	3%	3%	4½%
990–2000	3%	1½%	2½%	2½%	2%	3%	4%
000–2010	2½%	1%	2%	2½%	2%	2½%	3½%
010–2025	2%	½%	1½%	2%	2%	1½%	3%
orld conomic rowth	Central view, 1–1½% above UK growth	Depressed ½–1% above UK growth	Low, ½–1% above UK growth	Reduced, ½–1% above UK growth	Central view to 1990 then reduced	Central view, 1–1½% above UK growth	Buoyant, 1½–2% above UK growth
orld Prices Energy w materials	Rising slowly until 1990s, then faster, exceeding cost of syncrude soon after 2000	Slightly below Scenario 0 levels	Rising rapidly from early 1980s onwards	Sharply higher, doubling immediately, stabilising until 2000, then as in Scenario 0	Stable until 1990, then jumping well above Scenario 3 levels	Similar to Scenario 0 levels	Rising somewhat faster than Scenario 0
K Internal ergy Prices relation to orld prices)	Present relationship	High	Slightly higher than present relationship	Present relationship	Present relationship until 1990, then high	High	Present relationship
ergy nservation fects	Present expectations	High, but delayed by low economic activity	Very high, well above present expectations	Very high, well above present expectations	Present expectations to 1990, then very high	Higher than on Scenario 0	Present expectations, but accelerated by high economic activity
emium on lf- fficiency	Present policy	Present policy	Present policy	Present policy	Present policy	Very high	Present policy
uclear nt Building ogramme	Up to a maximum of 50–60GW(e) installed in year 2000	Up to a maximum 40–50GW(e) installed in year 2000	Programme halted after SGHWRs commissioned in mid 1980s	Up to a maximum of 50–60GW(e) installed in year 2000	Up to a maximum of 50–60GW(e) installed in year 2000	Up to a maximum of 90–100GW(e) installed in year 2000	Up to a maximum 100–110GW(e) installed in year 2000
ospects of ergy from ternative urces	Present expectations	Below present expectations	Very high, well above present expectations	Above present expectations	Present expectations	Present expectations	Present expectations
al oduction	Present plans and expectations	Slightly above present plans and expectations	Well above present plans and expectations	Above present plans and expectations	Present plans to 1990, then rising above present expectations	Above present plans and expectations	Present plans and expectations

(1) The same demographic assumptions are common to all scenarios, namely: total population would slowly rise but not exceed the 1975 level by more than 10 per cent in 2025; however, the total productive workforce is not expected to increase significantly over this period.

(2) Structural shifts within the economy are assumed to have only a second order effect on energy consumption. Industrial output has been assumed to rise at a rate of 0.25 percentage points above the GDP rate on all scenarios.

Fig. 10

Contributions of the technologies and their overall importance

hnology	Contribution Across the Scenarios: Medium Term (2)	Contribution Across the Scenarios: Long Term (3)	Overall importance of the technology's contribution to the UK (4)
MARY ENERGY TECHNOLOGIES			
l Production			
ing Technologies	Large	Large	x x x x x
Situ Coal Conversion	Nil	Nil*	x
end Natural Gas Production			
ntinental Shelf Technologies	Large	Large/Medium	x x x x x
anced Recovery of Oil	Medium	Medium	x x x x
nulation of Natural Gas Reservoirs	Small	Small	x x
pwater Oil and Gas	Small/Medium	Medium/Large	x x x x
clear Energy			
nium Supply	Medium/Large	Large	x x x x x
l Processing and Reprocessing	Medium/Large	Large	x x x x x
rmal Reactors	Medium/Large	Large	x x x x x
t Reactors	Small/Medium	Medium/Large	x x x x x
dioactive Waste Management	Medium/Large	Large	x x x x x
clear Safety	Medium/Large	Large	x x x x x
clear Process Heat	Nil/Small	Small/Medium	x x
ernative Energy Sources			
ion Power	Nil	Nil*	x x x
othermal Heat	Nil/Small	Nil/Medium	x
ar Heat	Small	Small/Medium	x x
ar-Photovoltaic Power	Nil	Nil*	x
ar-Biomass Fuels	Nil	Nil/Small	x
al Power	Nil/Medium	Nil/Medium	x x
ve Power	Nil/Medium	Nil/Large	x x x
nd Power	Nil/Small	Nil/Small	x
Shales	Nil	Nil/Small	x
stes	Small	Small	x x
RGY CONVERSION AND DISTRIBUTION TECHNOLOGIES			
al Conversion			
al as a Power Station Fuel	Large/Medium	Large/Small	x x x x
G from Coal	Small/Medium	Medium/Large	x x x x
nthetic Hydrocarbon Liquids from Coal	Small	Small/Medium	x x
tallurgical Cokes from Coal	Medium	Medium	x x x x
ctricity Supply			
ctricity Generating Plant	Large	Large	x x x x x
nsmission and Distribution	Large	Large	x x x x x
ctrical Bulk Storage	Small	Small/Large	x x x
mbined Heat and Power Plant	Small/Medium	Medium	x x x
s Supply			
G from Oil	Small/Medium	Small	x x x
nsmission, Storage and Distribution	Large	Medium/Large	x x x x x
her Energy Carriers			
strict Heat	Small	Small/Medium	x x
drogen	Nil	Nil/Small*	x
ERGY UTILISATION TECHNOLOGIES			
ilisation of Fuels			
al as a Domestic and Industrial Fuel	Large	Large/Medium	x x x x x
ctricity Utilisation Technologies	Large	Large	x x x x x
ctric Traction	Small	Small/Medium	x x
s Utilisation Technologies	Large	Large/Medium	x x x x x
at Pumps	Small	Small/Medium*	x x x
ernative Transport Fuels	Small	Small	x
ergy Conservation Technologies			
nservation in Buildings	Large	Large	x x x x x
nservation in Industry	Large	Large	x x x x x
nservation in Transport	Large	Large	x x x x x
PPORTING RESEARCH STUDIES			
sic Research	Small/Medium	Large	x x x x
ergy Systems Studies	Medium	Medium/Large	x x x x
vironmental Studies	Large	Large	x x x x x

otes: (1) Where quantifiable, 'small' refers to a contribution affecting up to 2 per cent of total energy supply, 'medium' up to 10 per cent and 'large' over 10 per cent.

(2) The asterisks refer to technologies whose contributions beyond our 50-year time horizon may considerably exceed those indicated. Their potential has been taken into account in arriving at the judgement in column 4.

Fig. 11

Research priorities and main implementation options

chnology	Overall Importance of the Technology (from Table 3)	Ongoing Phase of R, D & D	Priority For UK Involvement in Ongoing Phase	Appropriate Option for Implementation of Ongoing Phase	Lead UK Organisations for R, D & D in the Technology
)	(2)	(3)	(4)	(5)	(6)
IMARY ENERGY TECHNOLOGIES					
oal Production					
Mining Technologies	x x x x x	R, D & D	High	(a)	NCB
In-Situ Conversion	x	Research	Low	(e)	NCB
il and Natural Gas Production					
Continental Shelf Technologies	x x x x x	D & D	High	(a) & (c)	DEn/Oil Industry
Enhanced Recovery of Oil	x x x x	R, D & D	Medium	(c)	Oil Industry
Stimulation of Gas Reservoirs	x x	R, D & D	Low	(c)	Oil Industry
Deepwater Oil and Gas	x x x x	R & D	High	(a) & (c)	Oil Industry/DEn
luclear Energy					
Uranium Supply	x x x x x	Exploration	Medium	(c)	BNFL/Industry
Fuel Processing & Reprocessing	x x x x x	R, D & D	High	(a)	UKAEA/BNFL
Thermal Reactors	x x x x x		High	(a)	UKAEA/CEGB/Industry
Fast Reactors	x x x x x	Demonstration	High	(a)	UKAEA/CEGB
Radioactive Waste Management	x x x x x	R, D & D	High	(a)	UKAEA/BNFL
Nuclear Safety	x x x x x	R, D & D	High	(a)	UKAEA/H&SE/Industry/CEGB
Nuclear Process Heat	x x	Development	Low	(d)	UKAEA/BSC/DI/CEGB
lternative Energy Sources					
Fusion Power	x x x	Research	Medium	(b)	UKAEA
Geothermal Heat	x	Survey	Medium	(d)	DEn
Solar Heat	x x	D & D	Medium	(a)	DoE/DEn
Solar-Photovoltaic	x	Research	Low	(e)	DI
Solar-Biomass	x	Research	Low	(e)	DEn/MAFF
Tidal Power	x x	Assessment	Medium	(b)	DEn/CEGB
Wave Power	x x x	R & D	High	(a)	DEn/CEGB
Wind Power	x	Assessment	Low	(d)	DEn/CEGB
Oil Shales	x	Survey	Low	(d)	DEn
Wastes	x x	D & D	Low	(d)	DoE/DEn/Industry
NERGY CONVERSION & DISTRIBUTION					
Coal Conversion					
Coal as a Power Station Fuel	x x x x	R, D & D	High	(a)	NCB/CEGB/Industry
SNG from Coal	x x x x	D & D	Medium	(b)	BGC/NCB
Synthetic Hydrocarbon Liquids from coal	x x	R, D & D	Low	(d)	NCB/Oil Industry
Metallurgical Cokes from Coal	x x x x	R, D & D	High	(a)	BSC/DI
Electricity Supply					
Electricity Generating Plant	x x x x x	R, D & D	High	(a)	Industry/CEGB
Transmission and Distribution	x x x x x	R, D & D	High	(a)	Industry/CEGB/EC
Electrical Bulk Storage	x x x	R & D	Low	(d)	CEGB
Combined Heat and Power Plant	x x x	(Commercial)	—	—	CEGB/SSEB/DEn
Gas Supply					
SNG from Oil	x x x	R, D & D	High	(a)	BGC
Transmission, Storage and Distribution	x x x x x	R, D & D	High	(a)	BGC
Other Energy Carriers					
District Heat	x x	(Commercial)	—	—	DEn/DoE/Electricity Supply Industry
Hydrogen	x	Research	Low	(e)	DEn
ENERGY UTILISATION TECHNOLOGIES					
Utilisation of Fuels					
Coal as a Domestic and Industrial Fuel	x x x x x	D & D	High	(a)	NCB/Industry
Electricity Utilisation Technologies	x x x x x	R, D & D	High	(a)	EC/Industry
Electric Traction	x x	R, D & D	Medium	(b)	DI/Industry
Gas Utilisation Technologies	x x x x x	D & D	High	(a)	BGC/Industry
Heat Pumps	x x x	D & D	Medium	(b)	DoE/DI/Energy Industries
Alternative Transport Fuels	x	R & D	Low	(e)	DoE/Oil Industry
Energy Conservation Technologies					
Conservation in Buildings	x x x x x	R, D & D	High	(a)	DoE/Industry/Energy Industries
Conservation in Industry	x x x x x	R, D & D	High	(a)	DI/Industry/Energy Industries
Conservation in Transport	x x x x x	R, D & D	High	(a)	DoE/DI/Industry/Energy Industries
SUPPORTING RESEARCH STUDIES					
Basic Research	x x x x	Research	High	(a)	SRC/DEn/UKAEA
Energy Systems Studies	x x x x	Research	High	(a)	DEn/Energy Industries
Environmental Studies	x x x x x	R & D	High	(a)	DoE/DEn/Energy Industries

Notes: (1) In column 3, the letters 'R, D & D' refer to research, development and demonstration, 'R & D' to research and development, 'D & D' to development and demonstration.

Fig. 12

Key to Fig. 12 column 5

(a) Take a national lead

(b) Maintain a good technical competence

(c) Rely on international commercial interests

(d) Acquire and maintain the status of an informed buyer

(e) Await developments elsewhere

Chapter 11 Fallibilist vs Contemporary Decision Theories

In chapters 8 and 9 several criticisms were made of what I called contemporary theories of decision making, by which is meant welfare economics, including cost benefit analysis, and the various forms of Bayesian decision theory now current. These criticisms, it will be remembered, fall under three main heads. It was argued that contemporary decision theories incorporate a false view of individual values; that their attempts to show how social value is founded upon individual values all fail; and that they can give no account of the most practically important class of decision, those taken under ignorance. We may now add a little flesh to the skeleton fallibilist theory of decision making of the previous chapter by contrasting it with contemporary theories. Of particular interest, of course, will be the extent to which the fallibilist account can overcome the three chief defects we have identified in contemporary theories of decision making.

1. Individual Values

In chapter 8 contemporary decision theories were observed to incorporate the fairly standard view of individual values, according to which an agent has privileged access to his own values. If an agent holds that he attributes a value V to an object X, then, provided he is in his right mind and using language correctly, X has the value V for him. No facts can be appealed to to show that the agent is wrong in his evaluation of the object as evaluation is autonomous from factual considerations. The doctrine of privileged access clearly entails the principle of autonomy. In Part One, however, it was argued that no value judgement can be justified. This conclusion was supported in two ways; by an argument about the impossibility of halting the regress of reasons generated by any attempt to justify a

value judgement (chapter 1), and by arguing that the doctrine of autonomy is false (chapters 3 and 5). The whole purpose of the idea of privileged access to values, however, is to show how value judgements may be justified, so that we may conclude that contemporary theories of decision making incorporate a view of individual values which is false. This is extremely serious for these theories of decision making, for they aim to show how decisions may sometimes be justified. If, however, the values upon which a decision is based are not, and cannot be, justified, it follows that the decision itself cannot be justified. Contemporary decision theories, therefore, depend upon justificationist views of individual values. If these are rejected, then these theories must share the same fate.

In chapter 6, an alternative, fallibilist account of value judgements was developed which could cope with the sceptical conclusions of Part 1. As indicated above, such a theory of value can play no part in a justificationist theory of decision making. It is, however, possible to base a fallibilist theory of decision making upon this theory of value, because making a decision is simply a special case of picking the best item from a list. This is what I attempted in the previous chapter. The theory of decision making outlined there depends upon the fallibilist theory of value. A key element in the theory of decision making is that proposals for action be submitted to a critical debate of the same sort as is appropriate for any proposed value judgement.

2. Social Values

Social values are values which are used in the making of public, or group decisions, and, as we saw in chapter 8, contemporary theories of decision making see them as being based on the values held by the individuals who compose the group or society. The problem is to see how this aggregation can be performed in a rational and non-arbitrary way, and it was argued that

all attempts to tackle this have failed. The view of social values taken by the fallibilist theory of decision making outlined earlier is radically different.

On this account, the transition from disagreement in private values to agreement in public values is by persuasion. Imagine that X supports some high level evaluative claim V, which Y is seeking to attack. We may suppose that X and Y agree on some values, which we may call background ones. If Y is to argue against X, he must use arguments having both factual and evaluative premises, but to what values should he appeal. If he uses values which are his alone, the argument will have no impact on X, whilst if he uses values belonging only to X, he himself will be unconvinced. What he can use, therefore, are background values common to both. What Y must do is seek for facts which, when coupled with background values, falsify V. In pointing to such facts Y shows X's system of values to be factually incorrect, a defect which may be remedied by the rejection of V. If V proves resistant to such criticism, then it is corroborated and both sides must admit that it has a certain success, success which may accumulate sufficiently to convince Y that he should change his mind and adopt V. Starting with different private values, the process of critical debate leads to agreement, agreement which confers a public status on the value claim. This is how we should see the transition from private to social values, not, as on traditional views, as a summation of immutable individual values.

This solution to the problem of the transition from individual to social values has a surprising consequence which has been mentioned before. Because the values involved in the debate about rival courses of action are common to both parties, and generally pretty mundane, the important elements of such debates are all factual. We have already found this in looking at the debates about intermediate technology, the removal of lead from petrol and the return of corporal punishment in chapter 7, and the

case of nuclear energy discussed in chapter 10. Each of these is a debate about what course of action to take, and in each factual issues predominate, evaluative questions rarely being explicitly discussed. This appears to be very difficult to explain on contemporary accounts of decision making.

Thus, a group decision based upon social values is rational, not if the immutable and autonomous values of the individuals in the group are correctly and reasonably balanced (whatever this may mean), but if every individual is given the opportunity of arguing the case he favours, and if the minds of his fellows are sufficiently open to allow his arguments to alter their own evaluations. Rationality does not consist of finding compromise between fixed positions, but in using persuasion to alter positions in the hope of a consensus.[1]

This enables a quite different account of the binding force of social values to be given than that offered by the two traditions of social contract and utilitarianism, exemplified by Rawls and Harsanyi. The problem here is to explain why an agent should pay regard to social values even when these run counter to his own private interests. Both Rawls and Harsanyi argue that what gives certain social values this privileged position is that it would be in the private interests of every agent to adopt these values if he were ignorant of the social advantages and disadvantages which he and his fellows have. This potential agreement confers a degree of objectivity on the social values thus identified not shared by any agent's private values. To act in breach of these objective social values to one's own personal benefit is to exploit some kind of advantage, in physical strength, intelligence, wealth or connectedness, which one has over one's fellows. If these are imagined away by the device of the veil of ignorance, then one's best interests are served by these objective social values.

In the terms of the discussion of objectivist and subjectivist theories of value in chapter 6, the theories of Rawls and Harsanyi are, therefore, objectivist. There it was argued that any theory of value must explain two features of value judgements, their connection with action and their subjection to reason. Objectivist theories of value, it was then argued, can explain the latter but not the former, whilst subjectivist theories suffer from exactly the reverse defect. I have criticised both Rawls and Harsanyi in chapter 8, but it is now time to lodge a further, but no less fundamental, objection. In brief it is that, being objectivist theories, neither Rawls' nor Harsanyi's account of values can explain the connection between value and action.

On both theories, to know that a particular value judgement is to one's own selfish advantage under the veil of ignorance is to know that the value judgement is objectively justified as a social value. For Harsanyi it is in the selfish interests of someone (assumed risk neutral) choosing between social states under a veil of ignorance to value each state as the sum of the utility of the individuals in the society, for he then maximizes his own expected utility from participating in a society. For Rawls, a group of agents (assumed infinitely risk averse) deciding what rules of social justice to adopt under a similar veil of ignorance should protect their own selfish interests by agreeing to value a society by the utility enjoyed by its least well off member. In either case, evaluations made under the veil of ignorance are assumed to be binding even when the veil is lifted. Without this assumption, of course, the whole exercise of worrying about what values to adopt in a state of gross ignorance is quite wihout point.

But we may ask what can motivate an agent to adopt values which appear to be in his own selfish interests under the veil of ignorance, when he is no longer under the veil and has learned that they are now against his selfish interests. Why, in other words, should an agent not exploit the

knowledge granted him by the lifting of the veil to exploit his own advantages, regardless of the values which once served his interests under the veil? What we are looking for here is a connection between the agent's selfish interests and motives and the values which serve them under the veil of ignorance. Rawls and Harsanyi try to connect selfish interests and value judgements by pointing to the kind of value judgements which people would make in order to further their selfish interests under a veil of ignorance, but the problem is to see what motivation an agent can have for continuing to hold such values when the lifting of the veil reveals them to be against his selfish interests. Without such motivation, there is no link between action and value - the vital question 'why act according to these value judgements?' has not been answered. Like other objectivist accounts of value, therefore, Rawls' and Harsanyi's can explain the role of evaluative reasoning, but cannot explain why the objective values they identify should affect action.

A desperate escape from this difficulty might be the claim that what motivates people to adopt the values which serve their own interests under the veil of ignorance even when the veil is lifted is the recognition that such values are fair. This is, however, no more than a sleight of hand. If fair means simply serving selfish intersts under the veil of ignorance, our original question about motivation becomes rephrased as 'why should an agent adopt fair values when the veil is lifted?' If, on the other hand, fair retains its commendatory force, we can ask our original question by asking what can commend values serving an agent's selfish interests under a veil of ignorance when the veil is lifted.

On the theory of value presented in chapter 6 the division between objectivist and subjectivist theories of value is bridged, as was noted in that chapter. Action is affected by a value judgement because a value judgement is about an agent's preferences, but there is room for

reason, because reason is required to test what the agent claims his preferences to be, there being no privileged access to preference as on traditional views of value.

We have seen that the transition from private to public values is mediated, on this view, by persuasion by critical debate. In this way people with very diverse values can come to agree on some particular value, this agreement promoting the value to a social one. But agreeing on some value judgement in this way, is the result of each agent coming to realise that this value favours his own set of private preferences. There is, therefore, no more of a gap between social values and action than between private preferences and action. The example of the soldier used in chapter 6 may be broughtto mind here. He has the choice of protecting himself by staying in his trench or risking his life and health in an attack on the enemy and chooses the latter. Why he wants to stay in the trench calls for no explanation, but why should he prefer doing his duty above this? It may simply be that the evaluation that doing his military duty is prefereable to cowardice is one which has stood up to severe testing and so convinced him.

I have spoken against the use traditionally made of the veil of ignorance, but it must be admitted that there is, at its core, a good deal of common sense. We do find profoundly distasteful the unconstrained exploitation of whatever particular advantages one person has over his fellows, and we wince at the valuations used to justify such actions. What a device like the veil of ignorance can properly do is to prevent ad hoc special pleading for a particular value judgement. If the agent pretends that he is unaware of his own speical advantages and attributes, he cannot exploit this knowledge to construct ad hoc defences of those value judgements which serve his selfish interests. He is then forced to submit his value claim to genuinely critical assessment.

We have seen the need for evaluative claims to be corroborated by novel factual consequences, since corroboration is a measure of success and success is the passing of a test, a severe test being one which is expected to be failed. If an agent merely adapts his evaluations to facts which are already known to him and claims success for his evaluations from the factual consequences, his claim is spurious, since success can only be bought at the risk of failure, and his evaluations have not risked falsification. For success, evaluations need to have facutal consequences expected to be false and yet found to be true on investigation. This much, then, can be saved of the doctrine of the veil of ignorance: the only facts capable of conferring success on an evaluation are those not known at the time the evaluation is proposed.

3. Decision Making Under Ignorance

In chapter 9 it was argued that no contemporary decision theory can throw light on the problem of how to take rational, non-arbitrary decisions under ignorance. At best, decisions under uncertainty, where all the possible consequences of each option can be identified, can be handled by the use of subjective probabilities. Where not all possible consequences of each option can be identified by the decision maker, he is in a state of ignorance and can hope for no assistance from contemporary theories of decision making. Nevertheless, as was seen in that chapter, all important decisions are ones which have to be made under ignorance, even though proponents of contemporary decision theories are well practised in disguising cases of ignorance as ones of uncertainty. Thus, the theoretical and practical scope of contemporary decision theories are extremely limited. Moreover, it is difficult to see how any of them could extend their scope to cover decision making under ignorance. It would, therefore, appear that a radically new point of view is required. Does the fallibilist theory

offer assistance here?

I think that it does. The whole purpose of the theory is to recognize that any decision may be found to be mistaken, e.g. by the discovery of some quite unexpected consequence, so that decisions capable of being monitored should be favoured, and that any important decision should wait upon a critical scrutiny to eliminate what mistakes can be discovered in the proposals beforehand. The essence of good decision making is not optimizing some objective function, but being prepared for the unexpected, and being able to cope with it when it does occur. The theory, therefore, most naturally fits decisions made under ignorance. This, I think, is its greatest strength, for the failure of contemporary theories to accommodate such decisons seems to disqualify them from offering any kind of assistance, theoretical or practical, to real, as opposed to textbook, decision problems.

Since my paradigm is the making of decisions under ignorance, it can be expected that I will have difficulty in giving an account of decisions under more restrictive conditions. This is, indeed, the case, for the theory whose skeleton is given here cannot say how decision making under certainty should proceed. In a critical debate between proponents of rival courses of action it is essential to assess the degree of corroboration of the rivals, but corroboration involves the discovery of unexpected facts. Hence, if all the facts are known before the debate opens, no debate is possible. Under conditions of certainty, the method of critical debate fails.[2] This will, no doubt, sound shocking to those brought up in the justificationist tradition, but to them I would address the following points. Firstly, justifying a decision under certainty is not possible since the values attributed to each outcome cannot themselves be justified. Secondly, even if such decisions were justifiable, no

real decision is of this kind, so that it hardly matters if no theoretical account can be given. If the ability of the fallibilist theory to cover decision making under ignorance is admitted, then there is no longer any need to disguise such knotty decision problems as much simpler ones under uncertainty. The true nature of the problem can be appreciated wihout losing a way of tackling the decision. Decisions under uncertainty are as much textbook inventions as decisions under certainty.[3]

There are two subsidiary points which are worth discussing here, both of which cause serious problems for contemporary decision theories. We saw that these theories could give no account of why information was useful in making decisions under ignorance, where the information was not enough to reduce ignorance to uncertainty. We also saw the problems created for these decision theories by the need they have for reliable forecasts. What, then, is the value of information for making decisions under ignorance and what is the role of forecasting in such decision making according to the fallibilist account of chapter 10?

It is essential to submit any decision under ignorance which requires scrutiny to a critical debate, and we have seen that the central issues in all such debates will be <u>factual</u> ones. Both sides of the debate will make factual claims which, when coupled with shared, background values, support their case, and information is valuable to the extent to which it settles these factual claims and so furthers the debate. The debate tells all those concerned in the decision making process what information is relevant, though, of course, the cost of its acquisition may always be judged too high. In this way, the fallibilist theory of decision making can explain why information is important and relevent to decisions under ignorance, even where it is inadequate to reduce ignorance to uncertainty. Its practical benefit here is that the theory shows decision makers how to discover what information is relevant to the decision problems they face.

This brings us conveniently to the role of forecasting in decision making, as seen from the different perspectives of contemporary decision theory and fallibilist theory. It will be remembered from chapter 9 that the justification of a decision calls for the justification of a set of hypothetical forecasts of the form 'if decision d is made at t_1, then C will happen at t_2', and we saw the difficulties involved in this. To be justified, a hypothetical forecast requires to be a consequence of some universal theory which there are reasons for accepting and initial conditions concerning some future state of affairs. For this reason, justifying a hypothetical forecast requires predicting (or making a non-hypothetical forecast about) some set of future states of affairs. We saw the impossibility of satisfying these two requirements, at least for the kind of complex social systems inevitably involved in large scale decisions. For this reason, forecasters have resorted to all kinds of *ad hoc* devices which generate hypothetical forecasts and predictions, but without affording any insight into the mechanism producing them. We saw that this move runs into Hume's problem of induction, so that there is no hope of justifying forecasts using these devices.

This may be contrasted with the use of scenario analysis (or, misleadingly, scenario forecasting) in the estimation of flexibility discussed in the previous chapter. The function of such analysis is not to predict the future, nor to justify hypothetical forecasts, rather it is to test the flexibility of the system under the control of the decision maker. It is, in short, an explanation of possible futures and responses of the system to them, rather than a prognostication of the real future. The questions posed do not include 'what is going to happen?', but do include: 'if this decision is made now, what restrictions are placed on

future decisions?'; 'if this decision is made now, what consequences has it for our future ability to control the system?'; 'what future states of the system are ruled out by this decision?' and so on.

It might be objected that since a scenario is a hypothetical forecast of form 'if this decision is made at t_1, and if these states of affairs occur, then the state of the system at t_2 will be such and such', then the problems plaguing the use of such forecasts in the justification of a decision should equally afflict their use as scenarios in estimating flexibility. It must be remembered, however, that the hypothetical forecast plays quite different roles in these two exercises. If a hypothetical forecast is used to justify a decision, then it must itself be justified and it is this requirement which leads to all the problems mentioned earlier. On the other hand, estimating the flexibility of a system or set of systems may require hypothetical forecasts of much less precision and accuracy, especially if all that is required is a ranking. The hypothetical forecasts of a scenario may, therefore, quite properly be just our best guesses. It must also be remembered that the analysis can always be extended, for example by including additional scenarios, or by investing extra effort in revising any scenarios which prove to be crucial to the final decision. In either case, the original analysis is improved by force of criticism.

But if best guesses are good enough for scenario analysis, why are they inadequate in selecting the best decision from consideration of a set of hypothetical forecasts? The final decision will not be justified, since these hypothetical forecasts are not justified, but, as far as our efforts can make it, this is the best decision that can be made. This is not a satisfactory suggestion. If the decision is not justified, then it may be wrong and it is, therefore, incumbent upon the decision maker to search for mistakes, even after the decision has been taken, and to

ensure that detected mistakes may be remedied as quickly and as cheaply as possible. There is, of course, no need to treat a decision which is justified in this way. If it is justified, it is correct and no safeguarding against error is called for. If a decision maker treats the best decision he can arrive at as one which is justified, then the consequences may prove disastrous. If error is not detected and remedied, the costs of being wrong may be very large indeed. Once it is admitted that no decision can be justified, decisions must be viewed in quite a different way. Since any decision may be mistaken, all (sufficiently important) decisions must be scrutinized for error and revisable so that error can be remedied. It is not enough to continue as before, only confessing to the use of best guesses in place of justified decisions.

One difficulty in justifying forecasts discussed in chapter 9 is the problem of system boundaries. If only part of a system is studied in detail and understood, and possibly modelled, then any forecasts based on this knowledge are in jeopardy from unforeseen interactions with unconsidered parts of the system. If, on the other hand, forecasts are based on an understanding of the entire system, the forecaster has taken on an impossibly large task. It may now be inquired whether estimates of a system's flexibility encounter the same difficulty. To answer this, recourse to chapter 10 is required. There it was noted that flexibility can generally be treated at a very abstract level by viewing the system in question as comprising three stages; input, processing and output. If the system is flexible with respect to changes in input and processing and changes in demand for output, then it is flexible with respect to any change which affects the system. A revolution, for example, can affect the system, but only by altering the availability or cost of input, the ability and cost of processing or demand for output. If the system is flexible with respect to all three types of change, there is no need to ponder on

whether or not it is flexible enough to cope with revolution. Treating flexibility in this way avoids the necessity of having to prognosticate about revolutions and the like.

It also overcomes the boundary problem, which so plagues traditional attempts at forecasting. As an example, consider the electricity generating system. Optimizing the system calls for a thorough knowledge of everything which seriously affects it, and so demands an understanding of the factors determining electricity demand, the availability and price of fuel inputs such as coal, oil and uranium, the price of rival fuels like domestic coal, natural gas andso on. The boundary problem is obviously very severe here. If the electricity system is treated in isolation, forecasts about its future are very likely to be rendered wrong by interactions with all the other systems involved; and yet to understand the interactions between the whole bundle of systems is an impossibly complex task.

If, however, the electricity system is flexible with respect to changes in the availability and price of fuels, then the system can cope with such changes, irrespective of how they are caused; whether by exhaustion of resources, the public acceptability of using certain fuels, competition from coal or gas, or whatever. It is certainly possible to build a system which is highly flexible in this way. Sufficient capacity to meet maxiumum demand for electricity could be installed three times over, using nuclear, coal and oil stations. Such a system might have a very low response time to changes in fuel prices and availability, and so very low controlled misbehaviour costs. It would, though, be insanely expensive in terms of control costs. In practice, flexibility is enhanced by ensuring a mix of primary fuels and convertibility between oil and coal fired stations, a solution which bears far lower control costs.

It is also possible to construct an electricity system which is flexible with respect to changes in demand over the long term. Simply covering

every other acre of the country with generating plant would ensure the system's ability to respond to increases in demand, no matter now sharp and spectacular, but, again, the control costs would be huge. Luckily, there are more effective ways of buying flexibility. For such a flexible system, there is no need to be concerned with the reasons for changes in demand; it is enough to know that the system can accommodate itself to these changes however caused. Similar remarks might be made about the flexibility of the electricity system with respect to changes affecting the processing, in this case the conversion of primary fuel to electricity. Providing a system which is flexible in all three ways, to changes in inputs, processing and demand for output, avoids the boundary problem which any attempt to optimize the system runs into. If the interactions between the electricity system and surrounding systems is not understood, this does not matter. Any such interaction affects the electricity system by altering fuel prices and availability, conversion of primary fuel to electricity and/or demand for electricity. In each case, the electricity system has the flexibility to adjust to these changes. It does not matter that the origin of these changes is not understood; what does matter is that the system can cope.

What the above discussion also reveals is the tension which exists in any decision making problem between the ability to control the system in question and the need to predict its future behaviour. At one extreme, if it is possible to predict with confidence satisfactory future performance of the system, then there is no need to ensure flexibility. Flexibility, after all, is only required when things go wrong, so if it is known that the system's future behaviour will be satisfactory, there is no call for flexibility. At the other extreme, if a system has the highest possible flexibility, so that it can respond to changes with zero misbehaviour costs, and if its control costs are also zero, then there is no need to predict

how it will behave. It can always be adjusted to an acceptable state at zero cost. All real systems, of course, are between these extremes. What this means is that flexibility must be traded off against the need to forecast the system's future behaviour. This is the old trade off between controlled misbehaviour cost and control cost in a different guise. A flexible system will be less sensitive to errors in forecasting than an inflexible system, since it bears lower controlled misbehaviour costs, but it will generally bear greater control costs. The decision maker must, therefore, compare an inflexible system with high misbehaviour costs, and hence a high sensitivity to forecasting error, but low control cost, with a flexible system of low misbehaviour costs, and hence low sensitivity to forecasting error, but high control costs. Where the balance will be will depend upon the difficulty in forecasting the system's behaviour. Where this is a serious problem, an appropriate degree of flexibility must be purchased.

CHAPTER 11 - FOOTNOTES

1. This is not to imply that every decision demands a complete consensus between all interested parties.

2. The same is true on Popper's account of science. If all empirical facts (basic statements) are known beforehand, theories cannot be compared since comparison involves considering the unexpected predictions of the two theories.

3. One consequence of this which will be welcomed by many is that in tackling real decision problems, i.e. ones under ignorance, no recourse to subjective probabilities is required.