The University of Hull

Paradoxes in Decision-making and their relevance to Education

being a Thesis submitted for the degree of Doctor of Philosophy

by

Steven Norman Davis, B.A.

March 1995
Chapter 1 Paradox and Decision

INTRODUCTION

DECISION-MAKING

PARADOX

This chapter investigates the nature of decision-making and paradox. An Introduction looks at possible situations where decision-making might founder. These not only provide a starting point for discussion, but form the basis of practical considerations of paradox in Chapter 4. The idea of an organisation having individual members, a structure, a purpose, and usually being managed, is introduced. The importance of decision-making as a means by which organisations progress is discussed, and I contend that decision-making can be affected by paradox. The nature of paradox is next considered. Paradoxes are seen to be of three types: self-referential; additions to a collection which change the collection; additions to a collection which do not change the collection; with level shifting as a product of their interaction. I conclude that decision-making can only be understood within the framework of an associated definition of rationality. Only when we have evolved such a definition can we decide if a particular decision makes sense, or whether paradox will manifest itself.
Chapter 2 Paradox and Rationality

THE SEARCH FOR CONSISTENT AND GENERAL PATTERNS IN DATA

THE TREATMENT OF THE FUTURE AS A PARAMETER

THE BELIEF IN THE PERFECTIBILITY OF INFORMATION

THE USE OF FORMAL SYSTEMS FOR MAKING DECISIONS

In order to illustrate the relationship between rationality, decision-making and paradox, this chapter examines a commonly held view of rationality. This view I call the *Game Theoretic* view, since game theory in particular rests on this narrow, mathematical, algorithmic style. I contend further that this style is particularly prone to paradox, and illustrate this with reference to certain features of game theoretic thinking: the search for consistent and general patterns in data; the treatment of the future as a parameter; the belief in the perfectibility of information; the use of formal systems for making decisions.

I conclude that if this narrow type of rationality causes problems with paradox, a broader definition of rationality might be appropriate.
Chapter 3  Paradox Lost and Paradox Regained

RATIONAL BEHAVIOUR

RATIONALLITY AND META-RATIONALITY

THE BOUNDARIES OF RATIONAL BEHAVIOUR

WILLING WHAT CANNOT BE WILLEd

MEANS AND ENDS

THE PROBLEMS OF AGGREGATION AND GROUP DECISION-MAKING

HOW THE PARADOXES INTERACT

Following the conclusions of the previous chapter, this chapter examines the implications of a broader definition of rationality, to see if this can eliminate paradox. I start with such a definition, and then analyse its efficacy with reference to certain key questions or areas, which I contend should be dealt with by any adequate view of rationality. These questions are:

- can one rationally choose to be irrational?
- what are the limits or boundaries of rationality?
- is it rational to want something which cannot be obtained by a direct act of will?
- is it rational simply to intend to be rational even though outcomes are irrational?
- in a group of individuals whose actions define the group action, is rationality judged with reference to the individual or the group?

I conclude that the particular definition under consideration does not adequately deal with these issues. Further, any definition will founder, because the need to be rational in one area contradicts a supposed rational stance in another. So although a broader view of rationality might eliminate some paradoxes, it only strengthens others.

Chapter 4 Paradox in Practice

A CLASSROOM APPROACH TO PARADOX - THE INDIVIDUAL

PURPOSE, AIMS AND OUTCOMES - RATIONAL ACTION AND BY-PRODUCTS

STRUCTURE, ETHOS AND RULES - TACIT AGREEMENT, THE SORITES AND MAGICAL THINKING

POLICY MAKING AND THE SENIOR MANAGEMENT TEAM - LEVEL SHIFTS AND SELF-BINDING

This chapter has two aims. First, I consider if my theoretical conclusions about paradox drawn in the previous chapters are borne out in practice. Second, if paradox does in fact exist in practice, the chapter begins to address the theme of this and the next chapter: that is, the ways in which paradox might be circumscribed, avoided or utilised. I use as a basis for the discussion the characteristics of organisations outlined and illustrated in Chapter 1 - People, Purpose, Structure and Management. In each case I point out where
paradox can and does occur, and where it might be ameliorated. I begin with an outline of five lessons which I used to illustrate the problems of paradox to Year 8 children in a PSE lesson. To illustrate the problems of defining a purpose for the school, I look at the difficulties attendant upon the formulation of curriculum and policy from school aims. I then consider the construction of formal and informal rules as an example of the ways in which the definition of a school structure can falter. Finally, I look at the problems senior managers can encounter. Although the domains of the paradoxes described tend to overlap, it can be seen that certain paradoxes are particularly common in certain areas. So, for instance, individuals are often bothered by the prisoner's dilemma, whereas managers tend to be plagued by level shifting and its associated problems.

Chapter 5 Living with Paradox

INCONSISTENCY

PARADOX AND CREATIVITY

DISSONANCE REDUCTION

Having examined paradox from a theoretical and practical point of view in chapters 1 to 4, I conclude that paradox is an essential and unavoidable feature of decision-making. In this chapter, therefore, I examine aspects of paradox that might be useful to us. First I consider whether inconsistency, at times, is not appropriate. I broaden this contention to a general consideration of creativity and its roots in ambiguity, inconsistency and paradox. I am able to contrast this with the process of dissonance reduction, but conclude that this, likewise, is a successful and creative response to paradox. In fact,
much of the dissertation has shown that dissonance reduction is a key feature of decision-making.

Chapter 6 The Ways of Paradox

I conclude the work with reflection upon the paradoxes I have considered, and the general place and role of paradox in decision-making and management thinking.
INTRODUCTION

Sometimes these cogitations still amaze
The troubled midnight and the noon's repose.

T.S. Eliot, *La Figlia Che Piange*

Some years ago I tried to explain Arrow's Paradox to a head teacher, and suggested to him that it showed mathematically that consensus is impossible. I made some remark to the effect that this was a great blow for management, in that it would be impossible, on some occasions, to obtain a unified staff view. "Never mind," he replied cheerfully, "we can try." I remember feeling aghast at his assumption that a logical obstacle could be disposed of by empirical action. And, rather like Fraser faced with the Priesthood of the Nemi, I began to reflect more generally on this assumption. The more I reflected, the more I felt certain that I had encountered a fundamental problem in management ethos. As I considered the issue further, I began to think of other circumstances which I had found difficult to explain or come to terms with. Broadly, it became evident to me that organisations, and the groups within them, seemed to struggle far more than they should, and that Arrow's theorem, and the head's response to it, was
Paradoxes in Decision-making

the tip of the iceberg - the harbinger of a more general organisational malaise. I will illustrate what I mean with four examples of the sorts of issues that bothered me.

1. Making and enforcing school rules

A common method of making decisions in schools, particularly on policy matters, is to form a working party. Typically, such a group might consist of an open meeting of interested members of staff. For example, such a group might meet to create a simple set of school rules for publication within the classroom, the school brochure, and so on. First, such a group will compile a long list of desiderata - perhaps fifty or sixty items - suggested from around the room. Then the list is rationalised, since the suggestions often contain duplications or items that can be included under a broader heading. These can be easily dealt with. If there are still too many items, however, it becomes difficult to know how to proceed. There doesn't seem to be an easy way of voting on items so that the most popular are included. Too often, members of the group leave the session dissatisfied, feeling the final list does not include what it was generally felt should be there.

Once the list of rules has been defined, it seems sensible to make them known, and have them accepted and understood by more than just the teaching staff. To this end the list of rules might be explained to and discussed with pupils and parents. Often all parties agree on the value and sense of the document. Nevertheless, pupils are seen every day intentionally and unintentionally breaking the rules, sometimes with the collusion of their parents and sometimes even with the collusion of staff. Further, although the rules emerged from an excessive number of items, they still do not seem to say what they ought to say. They might, for example, make comments about children walking on paths and not chewing in class, but do not directly encourage appropriate or worthwhile characteristics: to be honourable, to tell the truth, to be ambitious, and so on. However, if statements like 'always be honourable' are included then it could be argued that these
Paradoxes in Decision-making

Paradox and Decision

might be aims or outcomes, but not rules. Rules are clear cut issues that are easily enforceable because it is obvious whether they are being obeyed or not. It is equally impossible to enforce 'rules' such as 'be honourable.' But if we have created a set of enforceable rules, then why is it that they aren't obeyed? If we want pupils to be honourable, caring, and so on, how do we produce these outcomes - this ethos - if we can't make rules about them? And do we have procedures for making rules anyway, other than calling, ad hoc, a working party? If this is the case, then what are the procedures for calling working parties?

2. Assessment

It is impossible to deny that one of the primary function of schools is to assess, and to prepare pupils for public assessment in the form of examinations. The nature of assessment, however, has changed markedly in schools over the last few years. Three issues seem to have led to this. First, there has been an increase in the use of pupils' self-assessment, and their general involvement in the assessment process. OFSTED guidelines make it quite clear that children will be expected to understand the sequence and nature of the course they are undertaking, where they have reached, and what they need to do to succeed. Second, teachers are urged to assess more than the simple content of their courses. Cross-curricular skills, pupil attitude to work, and social relationships should all be commented on. Third, schools are now expected to standardise assessment practices. It is considered that children become confused by varying mark schemes and contradictory forms of feedback.

In practice, however, assessment, particularly the formulation of assessment policies, seems to cause problems. Often teachers find that the changed approach towards their subject difficult to come to terms with. The emphasis on skills seems to affect the nature of the subject itself, so that it becomes a series of exercises which may have value in themselves, but do not seem to add up to the subject as a whole. The need to ensure that
pupils know where they are can mean that children learn about the subject rather than learn the subject itself. The need to standardise assessment procedures seems to turn the subject into a vehicle for general assessment, rather than a thing of value in its own right. Fundamentally, how do teachers square the assessment ethos that they accept is desirable with their beliefs about their roles and their subjects disciplines?

3. Management

Managers generally feel that it is their job to provide leadership, to provide expertise and guidelines for activity in their particular areas of responsibility, and to provide on-the-spot support for people where required. Managers are too often frustrated, however, because the ideas and developments they perceive as necessary and common sense are often rejected by staff. This rejection might be put down to lack of understanding or lack of effort on the part of others. However it is equally evident that individual members of staff are far from recalcitrant. The vast majority are hardworking, competent professionals. So why do they appear not to be supportive? Further, these staff have several excellent ideas, but often these do not seem to reach a public forum. This is because the agendas of these forums are often dominated by inputs from the senior staff. This situation is not restricted to the dichotomy between staff and senior management. Within the senior staff itself one or two members - often only the head - has almost complete power over the calling of meetings, their membership, their agendas, and hence control over the policy making process. Meetings, in this way, can be invested with any kind of power, or lack of it, as necessary. Too often senior staff and staff generally can become frustrated by the arbitrary nature of management. Further, when decisions are reached, they are not always pursued and eventually fall by the wayside. Policy documents accumulate, but are often not followed.
4. Offices

I have observed that one thing that is common to offices in school - the school's main office, faculty bases, pastoral centres - and offices in other institutions, is the plethora of amusing notices. These vary from the simple 'You don't have to be mad to work here but it helps' to the lengthy description from Tacitus of the morale-sapping effects that continuous reorganisation had on the Roman Army. Although these notices are meant for amusement, their extent shows they strike a common chord for the members of organisations. They show that individuals in an institution feel powerless against the weight of its authority, inertia and mindlessness. They emphasise the individual's inability to influence events, and to come to terms with what is tantamount to a mystical situation. Too often it seems that an organisation, which is quite evidently a collection of individuals, ceases to act like an individual within the remit of the collective will, but takes on a life of its own and becomes a monster.

When I considered situations like these, common themes started to become evident to me. I noticed, for example, that all were characterised by the inability of the organisation to turn its theories, beliefs or needs into practical outcomes. Next, the problems were not restricted to particular aspects of the school as an organisation, but were generally applicable to all its functions. I will explain what I mean in more detail. Organisation theorists attribute several roles to organisations, but the following seem to be most common. First, organisations consist of groupings of people - it is difficult to imagine an organisation of one. These individuals not only have formal and informal relationships with the organisation, but also lead lives outside of it. Second, organisations have some kind of purpose or raison d'etre, which is usually made explicit in some kind of philosophy, a set of aims and a series of particular objectives or desired outcomes. Third, they have some kind of structure based on sets of rules which dictate the way the
organisation will organise itself. Fourth, they are usually managed, often by a discrete management group. As can be seen, the illustrations I have given pervade all these, from the feelings of the individual faced with the organisation to the problems of managers. Likewise some items such as ethos cut across the functions, being part of both structure and purpose. Most importantly, organisations construct and renew themselves by making decisions. All the entities - individual, purpose, structure, management - are the product of decision and exist to make or enable the further decisions which will allow the organisation to progress. Decisions give organisations dynamism, without which they are merely collections of people. It struck me that a common theme of all the problems I encountered was that individuals, groups and organisations seemed to falter here, in the act of making decisions. Too often the results seemed to be unsatisfactory, arbitrary or inappropriate.

The head's solution to the problems of making useful decisions - to try harder - is a commonly suggested remedy. This implies that weakness of will or incompetence in individuals is at the root of decision-making problems. But this does not seem appropriate to a hard working and dedicated profession. Other reasons for decisional difficulties have also been suggested. Eric Hoyle, for instance, in his paper *Micropolitics of Educational Organisations*, describes organisations as intractable because they are a tangled web of micropolitical activity: because people are trying to further their own interests and make decisions accordingly. The school becomes a sort of arena for their power play. This again is undoubtedly true, and we have all known examples of this. Equally, most colleagues are not closet Machiavellis. The educational press is replete with suggestions as to why schools do not make the right choices: change is too rapid; bureaucracy both within schools and without is too impersonal; teachers as a profession are becoming over stressed; educational aims, which now place schools firmly in the market place, are inappropriate. Again this is partly true, but we could equally argue that many schools do not suffer excessively from these influences, and it does not follow that any would necessarily affect the quality of decision-making.
Two other reasons for decisional difficulties are more suggestive. Mike Bottery, in *The Ethics of Educational Management*, sees the problem as one of disillusionment. We often have a normative or idealistic view of organisations, which leads us to expect them, inherently, to run smoothly. When in practice they do not, the mismatch between belief and reality makes us disheartened both with ourselves and the system. Ross Harrison, in a paper, *Ethical Consistency*, which he contributed to a collection *Rational Action* which he also edited, says that rationally consistent systems of rules are flawed or weak because they often contain inconsistencies. These inconsistencies lead inevitably to contradictions and decision-making difficulties. Both writers come very close to the position I wish to adopt. This is that *akrasia*, micropolitics, or the loss of morale give no full explanation for why organisations, and the decision-making they undertake, falters. Instead, I believe there is something in their very nature that makes them periodically and unerringly fail: that makes them imperfectible. As Ross Harrison says, systems have unavoidable, endemic inconsistency. I refer to this inconsistency, more strongly, as paradox, in terms I shall define more fully shortly. Arrow's theorem, and its constraint on decisions based on consensus, is an example of this kind of endemic weakness. An inability to appreciate this feature of rational systems is a handicap to organisations and the people who manage them. If managers believe that, in all circumstances, decisions can be rationally made and systematically pursued then, as Mike Bottery says, they will be disappointed. First, I feel that rationality is often too narrowly defined, usually as a logical exercise using principles and ideas such as dominance, utility, induction and causality. We shall see that, popular as this view of rationality is, it too often contradicts itself, becoming paradoxical. Second, and more subtly, the nature of rationality changes in situations which are paradoxical, so any form of rational action can become distinctly irrational. In these circumstances, no definition of rationality seems to suffice. Where paradox makes decision-making difficult, styles of decision-making which would normally be considered irrational are often adopted instead: a strategy in fact advocated by Harrison in the paper cited above. If these styles seem wayward or incomprehensible, it is easy for
management to assume there is a deficiency in the managed - that they cannot reason properly or cannot be bothered to make an effort - rather than a deficiency in the situation. It is these situational deficiencies - paradoxes - and the responses they provoke, that I intend to explore. My thesis is that managerial decision making is better understood and undertaken when it becomes apparent that both managers and managed can be constrained by such situations. The aim of this work is to examine this constraint: how paradoxes tend to tie us up mentally, and make it difficult for us to proceed whilst in the grip of them. In particular, I wish to consider, from both a theoretical and practical standpoint, the mechanisms which thrust paradox into our daily lives and make it more than a logical diversion or an amusing puzzle. The whole will not only indicate ways of avoiding or employing paradox, but will consider the implications of its existence and influence.

I will return to the practical problems of management, of the sort that began this chapter, in Chapter 4. For the rest of this chapter, and the next two, I wish to look at the theoretical basis of my claim that paradox is an important but largely neglected influence on the making of decisions. To do this it is necessary to start by defining and examining some terms. I will define more as it becomes appropriate. At the moment, I wish simply to consider what I mean by decision-making and paradox.

DECISION-MAKING

Reason also is choice

Milton, Paradise Lost

Decision-making can be defined both in terms of its point of origin and its nature. Decisions can, in fact, originate in a variety of places. For the purposes of my thesis I will confine myself to three of these. First, decisions can be taken by individuals. I do not propose to consider the underlying physiological or psychological processes involved
here, even though there is a school of thought that argues that these processes, indeed the whole process of awareness and being, is paradoxical in itself. For example, in *The Ghost in the Machine*, Arthur Koestler says:

Thus we have a paradoxical situation of a feedback loop where the actor's activities automatically brighten and darken the stage-lights - which in turn influence the actions of the actors.

The second location for decision-making is the group. It is often tempting to see the group as the individual writ large: Britain decided to go to war; the Sharks decided to fight the Jets. This concept is not only misleading but, as we will see shortly, is the basis of paradox. I intend to argue that group decision-making is a totally different process from decision-making by the individual. Further, a misunderstanding of this group process affects management styles: managers, after all, tend to manage groups rather than individuals. The third location for decision-making is within the management group itself. I see this as a separate location, in that management can be described neither as a group nor as an individual. Given, for instance, the autocratic government of Prussia in the nineteenth century, it is quite permissible to talk of it as an individual. "Prussia decided to invade France in 1870" could, with equal validity, be construed as "Bismarck decided to invade France in 1870." Likewise, the statement "Britain goes off the gold standard" refers to a group decision made by the British Cabinet. It is simplistic, however, to boil management decisions down to either individual decisions or group decisions as appropriate. Management decisions are different from the others, I feel, because managers not only make decisions about issues but also make decisions about decisions. Put another way, "Kaiser decides to build a Dreadnought" is not the same type of decision as "Kaiser decides to ignore Reichstag". The first decision operates within a set of rules that prescribe the Kaiser's decision-making domain, the second redefines the domain. This ability to act both within the system and outside it - on the rules that govern it - seems to be the essence of management. Furthermore, management, by
formal agreement or by usage, embodies much of the power of the system. Their decisions, therefore, deal as much with the redistribution of this power throughout the system as with matters concerning the system itself. Finally, if management is formed to manage or police the system of groups or individuals, it also has an obligation to lead the system. This raises the question of what managers can or cannot will. I hope to show that many of the desiderata of management cannot be willed: they are the by-products of activity. Unlike individual or group, management seems prone to the problems of willing what cannot be willed.

Having defined where decisions take place, it is now necessary to define their nature. At the moment, a simple definition will suffice. As it is my contention, however, that some aspects of decision-making lead to paradox, a fuller definition will emerge from, indeed be fundamental to, the dissertation. I see the process, in its broadest terms, as the procedures by which an individual, group or management body, selects its preferred action or belief from a set of actions or beliefs. The procedures for arriving at these selections can be either rational or irrational. It is important to precisely understand the meanings of the words rational and irrational, since they are open to a wide degree of interpretation. Further, a changed stance on rationality may create or dissolve a paradox. The nature of rationality and its relation to paradox is a key issue in the unfolding of my thesis, and will be discussed fully in the rest of the dissertation.

**PARADOX**

Life is a paradox. Every truth has its counterpart which contradicts it; and every philosopher supplies the logic for his own undoing

Elbert Hubbard, *The Note Book*

Having, in a basic way, defined decision-making, and aware that I have probably posed more questions than I have answered, I will now attempt to define paradox. As with
other words encountered so far, paradox obtains of a number of meanings. In its broadest sense, it simply means a senseless statement. When Augustus de Morgan produced his classic work *A Budget of Paradoxes* he meant paradox in this sense. The work is a catalogue of attempts to do the impossible: square the circle, trisect an angle, dismiss Newton (Einstein would have been paradoxical to de Morgan.) More modern definitions of the word retain this concept of senselessness or impossibility, but associate it with contradiction. A paradox is senseless or impossible in that it contradicts itself. This is close to my meaning, and is well illustrated by the "paradoxical" drawings of Maurits Escher. In these, Escher shows constructs which at first glance are realistic, but on closer inspection could not actually exist. *Waterfall* (Plate I) illustrates this contradiction between reality and impossibility exactly. We know it is impossible for water to continuously descend, but we cannot see the gap in reality that the artist has exploited to create the trompe l'oeil. If the suspicion is that the artist has taken a liberty in depicting how water falls, then *Ascending and Descending* (Plate II) is a tour de force. Here the paradoxical movement is both up and down, as the monklike figures trudge endlessly and hopelessly round. These images still do not fully capture paradox as I wish to define it, however. Although the structures are impossible and contradictory, they are essentially static. If either building could be built, then it would be a standing embodiment of the contradiction. (Another of Escher's more famous works, *Relativity* (Plate III), can, in fact, be built.) Consider, by comparison, the lithograph *Drawing Hands* (Plate IV). The contradiction here lies not in the impossibility of the structure but the impossibility of the action. Simply put, the impossibility lies in what does what. Neither hand either creates, or is created by the other, yet both are. Each actualises the other in a circular dynamism that *Waterfall and Ascending and Descending* lack. If we could have constructed either of these, as I said, we feel we could encapsulate and explain - tame almost - the contradiction. With *Drawing Hands*, its ever shifting and active character makes us realise that we fail even at the construction stage. We are left simply with the option of watching the endless round of creation: simply to observe. We can stop the contradiction only by failing to observe and turning away. While we
Plate I
Waterfall

Plate II
Ascending and Descending
Plate III
Relativity

Plate IV
Drawing Hands
observe, the contradiction swirls before our eyes. In observing it we create it, and are
unable to pin it down. Escher created a whole series of images on this theme. *Reptiles*
(Plate V), for example, captures this essential circularity and everchangingness of
paradox. The grey reptile enters the paper and dies into the pattern. The black reptile is
born from the gaps in between but is born only to become a grey reptile. And the white
reptile that never moves off the paper, can it be said to exist at all? And what is existence
when related to the other reptiles? Paradox must in my view, therefore, contain three
elements. First, it must have an element of senselessness or illogicality about it
(I hesitate at this stage to use the word "irrationality".) Second, it must be self-contradictory. Third,
it must have a dynamism about it which makes its nature or state impossible to pin down:
it must be slippery and ever-changing. In his book *Truth, Probability and Paradox*, J. L.
Mackie refers to this quality of slipperiness as *undecidability*. This is too static to
describe paradox as I perceive it. Although paradoxical situations are undecidable, this
word simply describes a gap in the system, a sort of limbo area which the system leaves
untouched or undefined. Slipperiness, conversely, is constant movement between the
borders of the undecidable. So for, example, a paradoxical statement might become true
and untrue by turns. Chris Ormell, in his *Some Varieties of Superparadox*, refers to this
type of contradiction as *dynamic contradiction*, and to the various values (there might be
more than two) that the statement flips through as its *partial meanings*. This seems a
much more apposite way of thinking about paradox - slipperiness, dynamic
contradiction, partial meaning - than Mackie's *undecidability*.

Given the extensive constraints outlined above, it seems, at first sight, most difficult to
define anything as paradoxical in these terms. Interestingly, there are a large number of
paradoxes, some of great antiquity, which amply comply. Rather than content myself
with a list of these, I have attempted, instead, to identify certain criteria, or starting
conditions, which tend to generate paradoxical states. These are: *self-reference;*
*additions to a collection which change the collection; additions to a collection
which do not change the collection*. Together they also lead to paradoxes of level
Paradoxes in Decision-making

shifting. Since these criteria have a bearing on the unfolding of this dissertation, I will describe them in some detail.

Self reference

It is quite evident that objects can be changed by manipulation. If I take a teacup and break the handle off I have changed the teacup. A more interesting consideration is the extent to which I have changed the teacup. The answer to this depends on how I viewed it in the first place. If I see it as a bowl shaped receptacle, then I have not changed it a great deal. If I see it as a piece of fire-hardened and glazed china clay, then I have probably changed it even less. If, however, I see it as a device for handling hot liquids without pain, then I have changed it quite considerably. In effect, my perception of an object defines its state. Since it can be argued that any object has a large number of properties, my perception of it varies from yours according to how I weight the importance of these properties: a brick is essentially a different object to a bricklayer than to a smash-and-grab raider. Further, comments about objects, by changing perceptions of these objects, alter the object in the eyes of the observer. A comment such as The grass is green makes me perceive the grass from the point of view of its colour, and evokes images of green fields, landscapes and the pastoral idyll. A statement such as The grass is eaten by cows makes me concentrate on its usefulness as fodder, its part in the cycle of nature, and eco-systems generally. All flesh is grass makes me think more philosophically about man's place in the world and the nature of existence. In other words, statements have the power to modify perception and, in so doing, modify the object of this perception. I will use statements to illustrate how self-reference can generate paradox.

Let us first consider the statement: This sentence is written in English. Since it is a statement about the sentence itself, it modifies it in some way. In fact, it draws my attention to its provenance. This sentence contains five words draws my attention to its
Paradoxes in Decision-making

structure, *This sentence contains a verb* to its syntax, and *This sentence is banal* to its semantic. Further, statements do not need to be true. *This sentence contains six words* is false. These points are all obvious, but I am labouring them because they form the building blocks of paradox. So far, the self-reference has been harmless. But falsity provides an element of contradiction, and statements provide an element of change and manipulation. By fusing the two, it should be possible to produce a truly paradoxical statement: one that is contradictory and ever changing. Consider this sentence:

*This sentence, which states that 2 + 2 = 5, contains two false statements.*

Obviously, two plus two does not equal five, so the sentence contains one false statement. Since the sentence claims two false statements, it has actually made a further false statement. So the sentence actually does have two false statements: the one about two plus two equalling five and the one about its containing two false statements. So it has actually made a true statement: it does contain two false statements. But if its statement about two false statements is true, then it actually contains only one false statement ('two plus two equals five') So this makes its statement false, which means it has two false statements. Effectively, the sentence is true if it is false, and false if it is true. It illustrates the three criteria of paradox that I outlined above: it is nonsensical, it is self-contradictory, and it has a circular dynamism - a slipperiness which allows us never to apprehend its meaning. The epigraph to this chapter, which is the title of a book, contains exactly the same paradox.

It could be argued that sentences of this sort are meaningless, since they make no worthwhile statement. Let us consider this issue by taking two even simpler self-referential sentences than the one discussed above:
1. *This sentence is false*

2. *This sentence is true*

Sentence 1 is analogous to the one discussed earlier, in that if it is true it contains a false statement - that it is false. So the possession of this false statement makes it true that it is false, so it is actually true. But if it is true then it contains a false statement - that it is false .... and so on. This is paradoxical by my definition. Sentence 2 is not paradoxical - when it is true it is true, when it is false it is false: there is no contradiction here. Unfortunately, it is impossible to decide which it actually is; true or false. We might describe sentence 2 as *undecidable*. Accepting the criticism made at the start of this paragraph, we might describe sentence 1 as *meaningless*, since concepts of truth or falsehood seem irrelevant in this context: there is nothing for the truth or falsehood to amount to. We might also describe sentence 2 as meaningless in the same way.

The quality of being meaningless, however, is difficult to sustain. First, the sentences are not meaningless in a grammatical sense. They are correctly constructed, and make syntactic sense, in the way that a sentence such as *this sentence is* would not. Likewise, they are not meaningless in the sense of being absurd, like *this sentence is green* would be. If I apprehended these two sentences:

*The first American President was Abraham Lincoln*

*The capital city of France is Paris*

I could point to the first sentence and say 'this sentence is false' without being accused of making a meaningless statement. If I pointed at the first sentence and said 'this sentence is' or 'this sentence is green', I would be accused of making a meaningless statement. It is easy to gainsay this argument by indicating that my statement 'this sentence is false' here
Paradoxes in Decision-making

Paradox

and Decision

refers to another sentence, and this is what gives it its sense. The meaninglessness arises when the sentence refers to itself. Unfortunately, self reference is no arbiter of meaning. The two sentences:

The next sentence is true

The previous sentence is false

do not make reference to themselves, but are still collectively equivalent to the sentence "This sentence is false". A final argument against the use of the term meaningless is that it affects the logic of the situation. When we referred to sentences as being either true or false, we had a two-state logic. The introduction of meaninglessness means that sentences can now have a third state, that of being meaningless. So as we could construct a sentence:

This sentence is meaningless

If it is meaningless, then it makes a true statement, so it becomes true. But if it is true, then its statement about being meaningless is correct, so it is, in fact, meaningless! The circularity returns. If we decide the sentence is false - not meaningless, then it can be either true or false. So, in terms of our earlier definition, it becomes undecidable. If, however, we insist on pursuing the concept of undecidability rather than meaninglessness, the three-state logic produced causes exactly the same circularity as the three-state logic which included meaninglessness. The situation is made worse in three-state logic because of the ability to make compound statements, such as:

This sentence is either false or meaningless
An examination of the possibilities shows that if it is true it is false or meaningless, and if it is false or meaningless it is true. The circularity, the paradox, seems inescapable.

As was touched on in the previous paragraph, self reference does not need to be contained in the same sentence in order for it to fulfil the criteria of paradox. Consider, for example, this pair of sentences:

Plato: What Aristotle says is false
Aristotle: Plato speaks truly

These taken together are paradoxical. What makes their paradoxical nature difficult to assess is the fact that taken, singly, they are not paradoxical. The paradox, therefore exists, within the system of sentences. This state of affairs is made even plainer if we try to pick the sentence that causes the problem. Consider these four statements made by A, B, C and D.

A. The first American President was Abraham Lincoln

B. The capital city of France is Paris

C. The Earth is the third planet in the Solar System

D. Between us we have made as many true statements as false ones

It seems at first that D's statement causes the problem, since if it is true it isn't, and if it isn't it is. A moment's reflection reveals, however, that the other three statements equally contribute to the paradox: if any of the other three were not uttered then the last would simply be false and the paradox would dissolve. The fault lies in the aggregation of the statements, not in any particular one. A practical example of these systemic paradoxes
Paradoxes in Decision-making

can be found in the *Chelsea Bun* problem. Here, two friends taking coffee in a coffee shop are faced with two cakes, an indifferent one and a highly desirable Chelsea bun. Each argues that such a desirable item must be preferred by the other, and propriety and friendship demand that the other should have it. This feeling on the part of either is not paradoxical, but taken together it is, as both try to get the other to take the bun in a bout of vicious circularity.

Finally, self referential paradoxes can be generated by the imperative. Commands such as *Don't be obedient* are obeyed if they are not obeyed, and not obeyed if they are.

Self reference, therefore, is a particularly rich source of paradox, in that it exploits the unique relationship that objects have with themselves. Objects are limited in their normal action when this action is turned upon themselves: a pencil cannot write on itself; a snake cannot eat itself. When such self-reference becomes paradoxical, these limits placed on the power of the object become manifest. One of the fundamental traits of paradox, after all, is the seeming powerlessness to escape from it. It is to be remembered, however, that not all self-reference is paradoxical. It is statements of self-reference which *change* those same statements which cause the problem. These are the sorts of situations I have described so far. Chris Ormell, in the work cited earlier, refers to statements which modify themselves as *radically self-ascriptive*. Where the resulting statement contradicts the original, which in turn contradicts the contradiction, and so on - in other words, the stuff of self-referential paradox - he uses the term *radically adverse self-ascription*. This seems to amply and succinctly describe the process of paradox by self-reference outlined so far, and can generally be taken as the meaning of *self-reference* in future.

**Additions to a collection which change the collection**

At first sight, it seems obvious that as items are collected together, the nature of the collection changes. Each item contributes its own properties to the mass, altering it bit by
bit. So the addition of blue ink to clear water gradually turns the water blue. Already, however, we enter the realms of paradox: the mixture of ink and water has neither the properties of ink nor water; it has bits of both, but is essentially a new item, neither ink nor water - you would neither want to drink it nor be able to write with it. Worse, were we to make, for example, a mixture of carbon, which is a black solid with no smell, and sulphur, which is a yellow solid with little or no smell, we would obtain the compound of carbon disulphide, which is a colourless liquid with a revolting smell. If, therefore, an aggregation of items has different properties from the original items, how can this aggregation be representative of the items? How can the aggregation speak for the items? This unpredictability of collections gives rise to several socio-political paradoxes which I will describe more fully in a moment. A second area of paradox lies in deciding at what point the collection, as items are added, changes its nature or set of properties. This area of paradox I will deal with in the next section.

Let me return to the area of paradox presently under discussion: the tendency on the part of collections to be non-representative of the properties or nature of the items that comprise them: that the properties a collection will manifest are unpredictable. Please notice that I am not referring here to chaos theory. This theory states that chaotic systems, or systems near to chaos, are very sensitive to initial conditions. If such a system is set in motion, it will reach a given state in a given amount of time. If it is set in motion again, with only a very slight change in its starting conditions, it will, over the same time as before, reach a totally different state. Since slight perturbations at the commencement of the process are too small to measure, but will nevertheless still lead to a different conclusion, such systems become unpredictable. It is important to emphasise that the unpredictability lies in the starting conditions and in the nature of the system. The systems I wish to study become unpredictable as items begin to be aggregated to form a collection. It is the act of aggregation that creates the problem, not the starting parameters. In other words these systems are not chaotic, they are paradoxical. And they are paradoxical, and hence unpredictable, in relation to the same criteria that I used
earlier: they are nonsensical; they are contradictory; they are slippery. The paradox can result both from a logical impasse and from individual actions. A fine example of a logical impasse was cited by Professor Almond in conversation. It appears that there is a Society for Useless Information. Membership of the society is obtained by the aspirant sending the secretary a piece of useless information. However, any such piece of information immediately acquires utility in that it gains membership of the society for the candidate. It is, therefore, logically impossible to join the club! A second example of a logical impasse might be that it is logically possible for an individual in a group to act in a contra-majority way, but it is not logically possible for all individuals to do so. An example of paradox resulting from individual action can be taken from a group of competing entrepreneurs. A single entrepreneur, for instance, might benefit if he reduces wages, but would not if all entrepreneurs do. This difficult situation is compounded if we make the sensible assumption that if all individuals in a group are exposed to a similar array of stimuli, it is quite likely that individuals might respond in ways similar to each other. This need to make decisions against a background of physical and logical contradiction creates situations of great complexity which rapidly become slippery, circular and paradoxical, as individuals lose the power to make beneficial choices and to control the situation generally. Any choice made from an equally available selection seems to be the wrong one. An example will explain the sort of situation I mean. Marx describes a possible and logically ordered scenario for the collapse of capitalism. Suppose that in a contracting market, an individual entrepreneur decides to cut his costs and so reduce the price of his products in order to sell more. This is a perfectly logical and rational step to take. Indeed, the choice not to undercut would lead to ruin. Unfortunately, since one capitalist sees the point of doing this, they all do. And if cutting costs includes cutting wages, then the workers, who are the consumers, can buy less, making the market contract even more. So each capitalist tries to undercut the others again, and so on. As can be seen, this scenario neatly encapsulates the simpler situations described earlier in the paragraph. It includes the illogicality of a situation where all are trying to act in contra-majority way, and shows how individual sensible actions are not
necessarily advantageous to the whole group. However, it is difficult to see how anyone can act in any other way. It is this dilemma of inescapable disadvantage which creates the paradox.

Investigation of non-zero-sum games shows how this kind of rational action on the part of individuals can, by aggregation, lead to irrational action on the part of the group. Likewise, irrational individual activity can lead to a sensible group ethic. Practical work with children and adults, the results of which I intend to describe fully, can reveal the many-faceted nature of such social paradoxes. I also intend to consider, through them, the general question of the ability of systems to regulate themselves, and the extent to which systems need to be regulated externally by management groups. Most fundamentally, I wish to look at the paradoxes created by the fact that the management group is usually part of, but separate from, the system it manages. As I said earlier, the management group, in a very self-referential and slippery way, is, at turns, part of the group it manages, being subject to the development of its attitudes and ethos, and also aloof from it.

Finally, group decision-making can be undermined by what seem to be irrelevant issues. For example, situations often arise where the order of subjects to be voted on affects the outcome. Likewise, the order in which individuals vote or make their feelings felt can change the group decision. Strength of feeling can often change an outcome - paradoxically, stronger support for an issue may, in certain circumstances, lead to its non-adoption. These situations will be discussed fully in the following chapter.

**Additions to a collection which do not change the collection**

Stalin once said "One death is a tragedy; a million deaths are a statistic." A peculiarly twentieth century statement in its use of modern terminology and grim logic, it also embodies an ancient paradox - the sorites. This is the third generic paradox I wish to
consider. I will extend Stalin's epigram in a macabre but, I hope, not distasteful way: there is a point to the activity. If we agree with Stalin then indeed a million deaths are a statistic. But at what point do they become a statistic, and stop being a tragedy? One hundred deaths? One thousand deaths? Fifty-two thousand four hundred and eighty-three deaths? Perhaps the way to answer this question is to begin at the beginning. Obviously, if we accept the statement, then one death is a tragedy. Let us add a death. Are two deaths a tragedy? I think we would agree they are. Since every additional death is itself a tragic event, it seems impossible that the addition of one more will ever turn the collection into a statistic. In this way small additions, it seems, are never big enough to change the properties the collection has already acquired. The sorites paradox, in this way, argues that collections of things can never be changed. The paradox is named after the ancient Greek word for "heap". Like the example above, suppose we wanted to create a heap of sand by the addition of grains of sand. If we don't start with a heap, then the addition of one more grain will not be sufficient to turn what we do have into a heap. Likewise, if we start with a heap, the removal of a grain will never be enough for it to lose its heapness. This may seem a long winded and esoteric way of stating the obvious, but this approach to aggregation is indeed paradoxical and leads to many powerful paradoxes which have a bearing on decision-making. The approach is certainly not esoteric, since it underpins many very basic thinking styles.

First, it is common to feel one's way mentally through a problem in small increments. If the increments are too small, the sorites paradox follows. This is best illustrated by a description of someone trying to give up a ruinous habit. Although she might think that the habit is damaging in the long run, the effect of one day of the habit is practically negligible. Therefore, doing it one more day really makes no difference. This attitude, which is the sorites encapsulated, means that the bad habit will never be given up, since an accumulation of single days are insufficient in themselves to change the status quo. A rational approach could, therefore, be construed (possibly truly as it may turn out) as an irrational weakness of will.
A second and associated process is the change in behaviour by small steps. This can be illustrated by a consideration of bird migration. We can examine, for instance, the reasons for why birds migrate over such long distances. Let us suppose, for example, that birds originally migrated over a few tens or hundreds of kilometres. Let us further suppose that continents drift apart at the rate of one centimetre a year. Obviously, having to fly one centimetre further a year later makes absolutely no difference to an individual bird's ability to migrate. Over one hundred million years the species will have stretched its migratory distance to one thousand kilometres by these insignificant centimetre steps. It seems preposterous to suggest that gross movements of this kind requiring major feats of navigation by birds could be explained as developing in this way. Equally, it is impossible to explain where the weakness lies in the argument.

A third style of thinking also leads to a sorites paradox. This involves the process by which we acquire information, and construct it into collections, taxonomies or hypotheses. Information tends to reach our brains, through our senses, in a serial manner. At first, this information is simply stored and shuffled, until meaningful patterns begin to appear. After this the pattern is pursued, and later aggregation is aimed at confirming the pattern. This whole cycle of inductive thinking is fraught. Certain paradoxes, such as Hempel's, show that this style of thinking can be meaningless, and proof of inductive constructs can be logically valid but actually worthless. Further, the construction of changeless constructs leads to the rejection of counter-examples, which, like the grains of sand, are not seen to have the power to change the nature of the collection. Eventually, such weighty aggregations may achieve the status of paradigms in the terms described by Thomas Kuhn.

A fourth application of the sorites paradox can lead to 'slippery slope' arguments and their attendant problems. These were defined by Douglas Walton, in his book *Slippery Slope Arguments*, as:
(the) kind of argument that warns you if you take the first step, you will find
yourself involved in a sticky series of consequences from which you will be
unable to extricate yourself, and eventually you will wind up speeding faster and
faster towards some disastrous outcome.

He sees the sorites paradox as a fundamental basis for this kind of argument, since it
applies in any circumstance where a word or concept begins clear but becomes vague.
He cites as an example the word 'short'. It is clear that an adult four feet in height would
be called short. Someone of six feet would be tall. Between the two heights, however, it
becomes vague as to whether a person is short, at what height one ceases to be short,
where a person becomes not short, and so on. Given, however, that a person of four feet
is clearly short, no one who is short will become tall by the addition of say one tenth of
an inch. By the continuous application of this premise, therefore, no one will ever be tall.
The argument works because it propels itself from the clear area into the vague area,
where the premises conceded in the clear area now make it impossible to get off the
argument in the vague area. This is why it is a slippery slope argument: once it has
started it is impossible to stop. If there were no vague area, but a precise cut off within
the definition, then sorites thinking, and the slippery slope, could not operate.
Unfortunately, many concepts we use on a day-to-day basis are vague. Walton proposes
as an example the concept of life. When does life begin? Certainly, it doesn't begin at
birth, since this is simply a change of environment for the child: it is already functioning
as a person within the womb. But if life doesn't begin at birth, it really can't be said to
begin the day before. In fact, a day cannot make the difference between life and non-life.
Once this premise is established, the slippery slope into vagueness begins, and life can be
held to exist right back to the time of conception. Since this seems absurd, a 'disastrous
outcome' has occurred. Walton suggests equally disastrous outcomes for vague concepts
such as abortion and euthanasia. Further, arguments either using, or warning against, the
slippery slope, are paradoxical precisely because of their contradictory and slippery
nature. Such arguments are also very common. Take, for example, this paragraph, found almost at random in a popular computer magazine.

... you may want to think about paying the extra for the better specification ..... But when you start being tempted by the next machine up the scale, where do you draw the line? You can move up another notch by paying a little more for a 33MHz rather than a 25MHz 486, and then a 50 rather than a 33, and before you know it you're looking at a 66MHz 486. If you assess your needs .... don't be afraid to stick at the low end.

All styles of thinking and action - the stepwise mental and physical progress, the too short time discount, the formation of mental constructs by piecemeal aggregation and the movement from clarity into vagueness - lead to or are exploited by the sorites paradox. I see this paradox as a combination of the paradox of self reference, and the paradox of change by aggregation. Each addition is followed by a reference back to the whole. The addition is assimilated, and the whole redefined. But the redefinition in this case retains the properties of the original. In this way, the previous state informs and defines the present. So although the process is dynamic, the dynamism serves simply to pull the previous concept forward in time, the small additions never seeming powerful enough to stop this mental juggernaut. This procedure can be contrasted with aggregation which leads to change, where the incremental additions are taken in groups (e.g. when votes are counted). The weight of aggregation here seems to be sufficient to change the mental mindset. I will argue that the paradoxes of changelessness by aggregation tend to apply to individual decision making, where information tends to be acquired piecemeal in small but continuous amounts. Paradoxes of change by aggregation, I feel, tend to apply more to the development of formal and informal group decisions, where aggregation is examined only occasionally, and so can develop a weight and authority of its own. I will finally argue that self referential paradoxes are best applied to a consideration of management decision-making, tending to handicap its efficiency.
Level shifting

Finally, it is important to emphasise that these three types of paradox: self-referential paradoxes; paradoxes of change by aggregation, paradoxes of changelessness by aggregation, are also interrelated. For example, individual decision-making, and the problems encountered there, form the basis of the group decision. In turn, individual perceptions of how other individuals or the group decide will affect the individual's decisions. Therefore, individual decisions can be made both in the light of the issues involved, or in the light of what other individuals will decide about the issues. So the decision becomes one of deciding whether to decide by issues or decisions - a meta-decision. Likewise, should a meta-meta-decision be taken about the use of the meta-decision? Infinite regression of this sort, which is based on shifting levels of decision, is a form of self-reference, which again gives a slippery, paradoxical nature to the decision-making process. In later chapters, certain activities will be adduced to illustrate this changing of level. More importantly, paradox is being encountered at several levels: within the individual, within the group, and in relations between them. An example of this type of level shifting paradox might be provided by the self referential sentence:

This sentence has one error in it.

Obviously, the only error in it is the statement it makes about itself. It is the shift from the level of content to the level of a statement about content that makes the paradox, and renders the sentence true if it is false and false if it is true. So just as two sentences may not be paradoxical until they are juxtaposed, so two levels of thought can become paradoxical in the same way.
CONCLUSION

My thesis is that managerial decision making is better appreciated and undertaken if the constraining effects of certain well known paradoxes is understood. These paradoxes manifest themselves often, and it is essential to consider the mechanics of this - to understand the environments in which paradox flourishes. First, paradox becomes evident when we take decisions. Second, it is generated in situations where items or issues are made to refer to themselves, or when items are added to an already existing ensemble. Finally, the type and nature of the paradoxes change as we change our views of rationality, since our views of what is rational circumscribe the kind of decisions we take.

It is this intimate relationship between rationality, decision and paradox, and the problems this causes, that I wish to explore in the rest of this dissertation. In particular, I want to consider various current definitions of rationality, and the varying types of paradox they produce. This will lead me to consider whether paradox can be eliminated by any approach to rationality, or, indeed, whether it should be eliminated at all, and allow me to consider the role and influence of paradox in the managerial and decision making environment generally.

The next two chapters will consider certain definitions of rationality and the application to paradox. Chapter 4 will consider the influence and outcomes of these paradoxes by reference to actual situations in schools. Chapter 5 will turn again to rationality, and consider whether paradox needs to be avoided or whether it can serve a useful purpose.
Kirk frowned. "You mean to say that you, you, embarked on a course of action without any data to back it up? On intuition?" He shook his head. "I didn't think you had it in you, Mr. Spock."

"Logic, Captain, not intuition. A lack of data is not the same as no information."

Joe Haldeman, *Planet of Judgment*

I believe that people try to act rationally, and are comfortable in the presence of rationality. This desire for the rational implies that there are certain standards by which rationality could be judged: otherwise, individuals could not be sure of either their own rationality or that of others. I intend to show that when decision-makers try to act rationally this action can lead to paradox. In this chapter I will look at a commonly applied standard for rationality, and show how this can generate particular paradoxes. In the next chapter I will attempt a more thoroughgoing definition of rationality. Even this can lead to paradox, however. Some paradoxes defy rationality entirely - become a challenge to it, in fact. In the end, we may be left to live with paradox. I consider the implications of this in chapter 5: the need for more general responses to paradoxical situations, several of which would be deemed irrational.

The first definition of rationality I wish to consider is one that often occurs in the field of paradox (certainly it is applied to several paradoxical situations). This definition I shall refer to as the Game Theoretic definition, and is implied in the epigraph at the beginning of this chapter. Game Theory began with the monumental and seminal work of John von Neumann and Oskar Morgenstem, *Theory of Games and Economic Behaviour*, and had a precursor in the development of operations research in the 1930's. Although the concept took its name from the desire on the part of Neumann and Morgenstem to
analyse games such as poker, their intent was not to produce a recreational treatise. As the name of their book suggests, they were playing for much higher stakes than that. They believed that they could found a new discipline, which attempted to put decision-making, particularly economic and fiscal policy-making, upon an objective footing. Their analysis of games represented an attempt to find a set of axioms and theorems which would rigidly prescribe and describe these activities. They believed that, in the end, all decisions could be reduced to a set of statements and procedures which would produce the optimal outcome. Although they freely acknowledged that their work did not achieve this, they also believed that they had made sufficient progress to point the way for others to complete the founding of this new science.

Game Theory is still relatively unknown, but the styles of thinking associated with it have become very common. It has led to what might be termed a game theoretic rationality. This has several characteristics. First, it is considered rational to look for the Pareto superior choice: that is, a choice which, once made, cannot be improved upon by a move to another choice. Rationality in game theory, therefore, implies self-interest: no rational person or group would choose to make themselves worse off than they should be. Second, the procedures for finding Pareto optimality are quite strictly defined. For example, in game theory, a Nash Equilibrium point indicates a Pareto optimal strategy. Broadly, game theory operates on such procedures. These are meant to be objective and universal, a strict set of algorithms which will yield the optimal choice. This style of rationality implies the conscious and deliberate use of process and processing before the decision is reached. Third, Pareto optimality needs an exhaustive list of choices from which to select. This suggests that it is rational to collect and evaluate data extensively, so that all possible choices will be discovered and included. These choices have to be thoroughly understood, and generally reduced to a numerical, or utility value so that they can be ranked. A fourth aspect of this style of decision-making is that it takes cognisance of the future only as a weighted value. These weightings are usually discounted, or reduced, at a given rate the further into the future they are extrapolated. For example, a
decision to renege on a bargain will affect one's future ability to make bargains, but the
defection will be forgotten over time. Once this value is subsumed in a utility, the effects
of time on decision-making are assumed to have been dealt with. Fifth, game theoretic
rationality demands that decisions should be consistent. That is, in a re-occurrence of
similar scenarios, the same or similar decisions should be taken. Likewise, decisional
environments should not be affected by irrelevant elements. For example, if a person
chooses a hamburger from a menu consisting of hamburger, pizza and chicken, they
should not choose chicken instead on hearing the pizza is off. To do so would be to
exhibit inconsistency. There develops, therefore, in game theory, a desire to search for
the general action or law. Sixth, game theory assumes that individuals will not only act
rationally in the manner assumed above, but that individuals will also base their actions
on the assumed rationality of other individuals. Finally, and holistically, game theoretic
rationality is pervaded by an aura of objectivity and method. The quote which opens the
chapter is in the spirit of this rationality. Decisions based on intuition, or lack of process,
are not trustworthy. There is a close attention to the exact definition of words. There is
reliance on an evidential base: no conclusions can be drawn if there is no evidence.
Rationality presents itself at heart as a logical process: indeed, rationality and logicality
are often interchangeable commodities within it.

There are several decision-making styles which are similar in intent to game theory:
decision theory, operations research, linear programming, and so on. In fact, most of
these are used in game theory and likewise use aspects of game theory for themselves.
Where game theory is of particular interest - what makes it paradigmatic of this general
style of objective or logical thinking - is its emphasis on its role as a social or political
device. Its basis in play - the interaction between a number of individuals - gives it a
broader scope than other decisional devices, which tend to relegate all aspects of the
decisional environment, both human and non-human, to the level of variables or factors.
Instead, game theory asks us to consider other 'players' rather as we would if we were
playing chess. We would not move in chess without first considering what move our
move would elicit from our opponent. A consideration of her move might lead us to consider our next move, and so on. Theoretically, after analysis of this sort, we should be able to cross from one side of the board to the other and make several moves in succession. This would especially be the case if we could assume that our opponent were as sensible and perspicacious as ourselves, and game theory demands that we make this assumption. It is this essentially dynamic nature of game theory which gives it its piquancy, and makes it particularly representative of modern objective thinking, in that it expects us to make strenuous and complex considerations of our relations with others. At the end, having quantified and prioritised these considerations, we can use formal mathematical methods to find the optimal solution. For game theory, therefore, rationality lies in self interest, complete information, formal method and social interaction, and its outcome is the maximising of utility, the pursuit of the optimal, and the search for the general and the atemporal.

Irrational decision-making, to a game theoretician, is any procedure that arrives at a decision by disregarding optimisation, consistency, the future, or other individuals. This includes decision-making by intuition, lottery or under motivations such as altruism. A typical example might be the procedure known as dissonance reduction. Here the aim is not to find an objectively optimal solution, but instead to find a solution that makes the individual comfortable. It is characterised by a tendency to focus on one or two strategies. Selection may lead to data collection, but simply to reassure the individual that she has made the right decision. It is rational, but in a post hoc, self-fulfilling way. Further examples might include decisions based on non-causal premises: that is, rather than discounting the future, decisions might be made which see the present as having some causal influence over the past (Newcombe's Paradox turns partly on such a concept.) A decision might also be made in relation to irrelevant factors, or be seen as a unique action. There will be no attempt to extract general principles or transferable skills; life will be seen as an assemblage of individual activities or events. Finally, irrational individuals do not assume rationality in others.
Game theoretic rationality and decision-making is characterised by brainstorming, logical thought, and algorithmic or mathematical procedure. The decision is a product of research and formalised process. Although these styles were not invented by game theory, they are presented in its literature as the ways to think, the only ways appropriate for the new age and new science of decision-making. A glance at the introductions to books on game theory, or books which use game theory, illustrate this belief. In *Theory of Games*, edited by A Mensch (1964), Thomas Schelling says (p 469)

"Game theory" has come to mean the mathematical exploration of the frontier of an important subject. The subject has no convenient name; it covers all the situations in which the decisions of two or more participants depend, or ought to depend, on the decisions they expect each other to make. (Italics mine)

Writing thirteen years later, John Harsanyi, in *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations* said (p3):

The purpose of this book is to present a new approach to game theory. Based on a general theory of rational behaviour in game situations, it yields a determinate solution (i.e., a solution corresponding to a unique vector payoff) for each particular game and clearly specifies the strategies by which rational players can most effectively advance their own interests against other rational players.....Thus it supplies the analytic tools needed for what may be called a bargaining-equilibrium analysis of social behaviour and social institutions.

Five years later, Andrew Colman wrote in *Game Theory and Experimental Games* (p5):

Certain important features of individual and collective rationality, cooperation and competition, trust and suspicion, threats and commitments cannot even clearly be formulated without the concepts of game theory.
Paradoxes in Decision-making

Each quote emphasises aspects of game theoretic rationality: its centrality; its all pervasiveness; its ability to deal with 'real life' situations; its mathematical objectivity; its power, particularly in analysis. Like many other institutions, schools, try as they might, have been unable to resist this approach to rationality. Writing in *Culture and Power in Educational Organisations* (ed. Adam Westoby, 1988), William B Tyler notes that there is very little literature on schools as sociological institutions in their own right, and goes on to say:

This apparent neglect may result from the schools' lack of technology that links inputs to a rationally-deducible index of performance - or even a clear and unambiguous set of procedures by which such a criterion could be defined.

This advocacy of what he later calls 'school effectiveness research' nicely presages the appearance of school management plans, performance indicators and league tables early in the next decade. I will have more to say about this reductionist, analytical style of rationality as it is applied to schools in Chapter 4, but we cannot deny that this seemingly all pervasive style - the game theoretic style - is now becoming increasingly apparent in educational situations.

I intend to argue that there are several fundamental drawbacks to this view of rationality. First, it imputes value-laden meanings to the words rational or irrational, and sees no place for irrationality. I believe that there is no absolute support for one style of decision-making over another: as we will see, there are strong views on the efficacy of both rational and irrational styles, and a major aim of later chapters will be to draw conclusions about their relative usefulness. I will discuss this issue fully there. I will spend the rest of this chapter discussing two other drawbacks. First, that the game theoretic approach seems to be based on how people ought to act rather than how they do act. Second, that game theoretic thinking can lead to paradox. I will illustrate this contention with reference to four aspects of the game theoretic style of thinking. These
are: the search for consistent and general patterns in data; the treatment of the future as a parameter, the belief in the perfectibility of information; the use of formal systems for making decisions. I will extend this last to make more general comments about this style of systematic thinking.

**THE SEARCH FOR CONSISTENT AND GENERAL PATTERNS IN DATA**

The validity of predictions may involve a lot more subjectivity than many scientists would care to admit - even when predictions arise from so-called scientific laws. In short, a hypothesis or theory is not necessarily made more certain when its predictions are confirmed.

*John L Casti, Searching for Certainty*

In a scientific age, the formation of consistent and general patterns - hypotheses and laws - has been imbued with a mystique which makes it a thoroughly acceptable - the thoroughly acceptable - way to think. Narrow rationality is often associated not only with logic, but also with induction and experimentation - what is often termed the 'scientific' method. This cosy association, however, hides problems.

To begin with, we can clearly perceive circumstances where theories of induction, though sound in a narrowly logical sense, in fact do not square with what we actually do. When we make patterns, we are often far from optimal in our techniques. For example, Karl Popper argues that hypotheses should be worded in ways that they can be negated, rather than supported. After all, no amount of support ever proves a hypothesis. One contrary piece of evidence, however, destroys it. All hypotheses are, by their nature, tentative. His proposed procedure supports this contention: hypotheses gain currency and acceptance not by their strength in logic since they have none, but by their continued inability to be contradicted. This approach is eminently logical and efficient. Unfortunately, the work of such cognitive scientists as P.C.Wason shows that the
accumulation of information to support induction, rather than to try to negate it, seems the more natural approach.

Wason and Johnson-Laird conducted a series of experiments in thinking between 1966 and 1969. Their work is summarised in *Thinking*, which they edited and published in 1977. The original formulation of the problem they set is as follows. There are four cards showing, respectively, a vowel, a consonant, an even number and an odd number. Such a set would be:

![Four cards](image)

You know that each card has a letter on one side and a number on the other. Given the rule:

*If a card has a vowel on one side then it has an even number on the other side*

what cards is it absolutely necessary to turn over in order to test if the rule is true or false? Wason and Johnson-Laird found that the overwhelming response was either 'A and 4' or 'just A'. Not only is this response incorrect, but its nature suggests that there is a deep-rooted and extensive desire among people to make proofs by confirmation. The correct answer, 'A and 7', is the only one which allows proof by negation. I will explain this in more detail. At first sight, 'just A' seems a sensible answer. If we find an even number we have supported (proved) the rule. Wason and Johnson-Laird refer to this as a level of 'no insight'. 33% of subjects operated at this level. Support of the rule was not enough, however, for 46%. Although 'just A' might disprove the rule if it turned up an odd number, a second card must be found to disprove it if A did not. The search for this second card proved impossible for them since their choice of 4 does no better. If a vowel
Paradoxes in Decision-making

appears it supports the rule, if a consonant appears it does not disprove the rule, since the rule makes no statement about consonants. This desire to disprove the rule but the inability to reformulate it so that it can be disproved is referred to as a level of 'partial insight'. 17% of subjects made choices of 3 or 4 cards, which re-emphasised either their no insight or partial insight level. Only 4% of subjects made the correct choice. 96% of subjects, therefore, were incapable of either appreciating the necessity of negation in establishing the truth or falsity of a statement, or having understood the need for negation were incapable of understanding how the statement could be reformulated in its negative form prior to negation. Both the no-insight level and the partial-insight level, therefore, reveal an inability to deal naturally and comfortably with negation.

The authors, and other researchers made later reformulations of the problem in order to remove perceived difficulties in conceptualising the problem. The authors themselves used problems based on a choice between two shapes rather than between odd/even, vowel/consonant options, options which it was considered might overly complicate the problem. This change did not substantially alter their findings. Other reformulations such as:

*If there is a B on one side of the card then there will not be a 3 on the other side*

gave a much higher correct response (B and 3). It seems likely, however, that the correct solution is simply achieved by selecting the cards mentioned in the rule rather than understanding the rule itself. In the same way, the odd/even vowel/consonant rule would lead to the erroneous choice of A and 4 simply by similar matching: partial insight may not need to be invoked for its choice.

Changes were also made (e.g. by Wason and Shapiro) to investigate whether realistic scenarios improved results. Several such scenarios did show improvement. For example, the rule and cards:
Paradoxes in Decision-making

Paradox and Rationality

when I go to Manchester I travel by train

produced a much higher correct choice (Manchester and Car). Here the linking of concrete items in the rule may well have led to the insight that the cards which match them actually are making the same statement, so their combined choice serves no more purpose than a choice of one only. The choice of another card seems, thereby, indicated. Whether success at these problems negates the original premise, that people tend to think in terms of confirmation rather than negation of hypotheses, is debatable. It is certainly arguable that since these reformulations change the appearance of the original problem, they may also change its nature. The authors claim, essentially, that artificial problems are useful in that they reveal and focus thought processes which the reformulations soften or hide. Further, these thought processes assume confirmation as a basic and natural activity for verifying hypotheses. Negation is either ignored or, where it is practised, it is little understood in its mechanics and sketchily enabled in its procedures.

I will conclude this discussion with a practical example of what I mean. In his book, The Magic Machine, A.K. Dewdeney considered, in a chapter entitled 'The Martian Dictionary', several types of word ladders. These are strings of letters which are joined to each other in a particular way. The best known ladder is that devised by Lewis Carroll, which joins one word to the next by changing one letter in it. In this way, for example, house can be joined to horse. Dewdeney proposed that all words of the same length can be joined to each other in this way. He made this proposition because all the twenty two-letter words in his dictionary could in fact be so joined. Since the chapter was taken directly from his 'Computer Recreations' column in Scientific American, he was able to suggest that readers might try to verify his proposition. Those:
with computers might attempt to test the proposition with other low values of \( n \), such as \( n = 3 \).

The addendum to the chapter contained a discussion of the resulting correspondence. With what seems genuine surprise, Dewdeney wrote:

It proved hardly necessary to use computers to explore the interconnections. Many people offered *gnu* to settle the three-letter case. For \( n = 5 \) a number of others offered *xylem*. These forays depend only on the insight that a word that cannot be connected to others settles the issue; then others cannot be connected to it.

Dewdeney had envisaged a computer based number crunching exercise to create ladders which would encompass every word. If this strategy worked, then his proposition would have achieved positive support. He had completely ignored the much simpler approach of verifying the proposition by attempting to negate it. Further, by referring to this strategy of negation as an 'insight', he sees this as an uncommon act, based on some kind of mental leap.

I noted as a preamble to these examples of game theoretic thinking, that as well as being examples of unnatural thinking styles, they were also prone to paradox. The process of induction has in fact generated several. First, the piecemeal accretions of information to an already established ensemble fall foul of the Sorites, or heap paradox, which I have already described. Here the paradox lies in the apparent inability of small aggregations to change the heap. It is revealed when the gross nature of the heap is made apparent at some future date, and the changes become suddenly evident. Several paradoxes which bring the process into disrepute have been well examined. The desire to acquire something new (for example, data in induction) is explored in the story of the farmer and the horse. Here, a farmer decides to take his horse to market. On the way, he has the
chance to change it for a cow. Later still, he changes the cow for a goat, and so on. He is pleased with every exchange, because it gets him something new. Every exchange, however, is of slightly less value. In the end, he has a tub of rotten apples. The desire for novelty has led to his ruin: the accumulation of insignificant poor exchanges is suddenly revealed to him by a passer-by. By his actions he has, effectively, improved himself to death. As I will show in Chapter 5, the work of, for example, Edward de Bono, criticises the quality of decision-making that this style of thinking produces. The desire to acquire items, which are added to what appears to be an unchanging accumulation of data, leads eventually to widely erroneous conclusions or poor quality decision-making. De Bono calls this style vertical thinking.

A second paradox which illumines the accumulation of information from a slightly different angle is Hempel's Paradox. If the paradox of the farmer and the cow shows that accumulation of information for its own sake can lead to ruin, then Hempel's Paradox shows that accumulation with some given aim in mind can also cause problems. Like Popper, Hempel attacks the act of accumulation that aims at supporting inductive constructs. The game theoretic definition of rationality would see induction by confirmation as in no way irrational. Hempel's Paradox attacks it, however, by using the very logic and rationality it purports to draw on. Suppose, he says, that we want to support the hypothesis "all ravens are black." Let us suppose, further, that we wish to support this hypothesis by observation. Then, obviously, every occurrence of a black raven that we see helps to support our inductive structure. If we apply logic to the situation we can actually make life even easier. We have an inductive statement:

1. All ravens are black

The logical equivalent of statement 1 is:

2. All non-black things are non-ravens
Please note that statement 2 does not imply that ravens are the *only* black thing; it simply makes no statement about other black things. The situation is better represented by a diagram.

![Diagram](image)

It is the area of non-black things that we are interested in.

Hempel argues that since the two statements are logically equivalent, then support for one is support for the other. That means that anything that is not black and not a raven, supports the original hypothesis that all ravens are black. I see a brown shoe. That is a non-black and non-raven, so it confirms the original hypothesis. I see a green book. That also confirms the hypothesis. I can sit in my chair at home, continually confirming my hypothesis about black ravens: I need never see a raven; I need not even step outside my door! There is obviously something wrong here. Again, the non-changing heap ("all ravens are black"), cannot be changed by the accretions that are added. More fundamentally, the paradox questions the nature of the accretions: it seems to allow a green book to be added to a heap of ravens. Worse still, the logical equivalent of

\textit{all ravens are yellow}

is

\textit{all non-yellow things are non-ravens}
So the green book not only supports the hypothesis that all ravens are black, it also supports the hypothesis that all ravens are yellow! This bizarre conclusion, it seems, undermines the very roots of induction by confirmation. Certainly, there is a school of thought which accepts the validity of proof by green book: that a green book does, in some small way, add to the belief that all ravens are black. This seems to me preposterous, and to highlight the fact that one of our very fundamental and psychologically conformable ways of making sense of the world is undermined by paradox. Hempel's paradox, in fact, underlines the reality that processes based on logic and objective procedure yield problems. It turns on concepts of logical equivalence, and logical relationships, as we have seen, are a cornerstone of game-theoretic rationality. To refuse to accept such narrow, exceptionally artificial and counter-intuitive constructs effectively dispels the paradox. This, after all, is to be expected: Godel's theorem shows the impossibility of any formal system to capture all the possible theorems within it. Likewise, any narrow rationality will not accept as rational all actual rational actions.

A second type of induction, that of backward induction is exposed by the paradox of the Unexpected Exam. Backward induction works differently to the induction we have considered thus far, where the aim is to make hypotheses: to find a general principle which relates given items of data. It is a way of tackling a series of related decisions or actions, to produce a single optimal one. It is the technique we use, for example, in finding the correct entry to use in order to successfully reach the centre of a maze. This is accomplished much more easily by starting with the goal and working back to the entrance. The backwards induction approach can be illustrated by a simple game in extended form. It is typical of the sort of game that has enjoyed extensive analysis by game theoreticians. The analysis I have given is likewise typical of a game theory analysis. The illustration itself is taken from Kreps D., *Game Theory and Economic Modelling*.
A and B are two players. Moving from left to right, A has first choice. If she chooses to move right, then she passes the choice on to B. If she chooses to move down, then the game ends with the payoff (A's score is shown first - i.e., A gets 3, B gets 1.) If she chooses to move right to B, then he has the same choice of either moving right and giving another choice to A, or moving down for the payoff - again A's score, as always by convention, is shown first. The game goes on until a payoff is reached. The question for A is, since she has to move first, which choice should she make, right or down? We will assume that it is in A's best interests to achieve her highest score. Rationally, therefore, she should behave in the following way, by backward induction. First, she should look at the end of the game. Since B has the final choice, he is likely to choose right for a payoff of 2 to A, 2 to B, rather than down which would secure only 1,1. Given this logical choice by B, A can expect that, if the game reaches its end, then she will receive a payoff of 2. This is worse than the 3 she will get if she chooses down straight off. Things get no better if she moves back one go from the end. Although she has a choice of down which will secure her 5, she only has to look a go further left to B, who will not choose to move to her go since he can secure 3 for himself by choosing down. Further, from this choice she receives only 1. The further back from the end of the game she goes, the worse it gets. So her best choice is to move down straight away.

The backward induction technique seems very sound, and has been strongly advocated as a rational decision-making technique. It has been brought into question, however, by the Paradox of the Unexpected Exam. We are asked to imagine that a class has been told by their teacher that it is to be given an exam during the following week. In order that they cannot prepare for it, the class are further informed that it will be unexpected. Over the weekend, the class argue as follows. First, if the exam is to be genuinely unexpected, then there is no possibility that the teacher can set it on Friday, because if they hadn't had
it by Thursday night then the exam would be expected. If Friday is negated on Thursday night, then what happens on Wednesday night? The exam can't be on Friday as already agreed, so it must be on Thursday and would, therefore, be expected on Wednesday night. By the same process of backward induction, the class negates all the days. They conclude that, by some error in logic, the teacher had made it impossible for herself to set the exam. Imagine their surprise, therefore, when they got the exam on Wednesday morning!

If nothing else, the paradox shows the problems attendant upon the procedure of backward induction. The stepwise nature of the induction is of particular importance in the paradox. If the logic were not based on the serial occurrence of days, then there would be no paradox. The necessity of seriality is better illustrated by the paradox of the unexpected egg. Here you are asked to picture two boxes, one of which is to contain an egg. Can the egg be placed in one of the boxes so that when you look in them the egg will be unexpected? The answer is that it depends how you are allowed to examine the boxes. If you are allowed to select one at random, then you have no way of knowing whether the egg is in that box or not. So, by this method of selection, the egg is unexpected. If you have to examine the boxes in a given order, however, you can argue that the box you will examine second cannot contain an unexpected egg because you will expect it after finding the first box empty. The first box cannot contain an unexpected egg because, by the argument above, it cannot be in the second one so must be in the first. We must conclude that it is impossible to place an unexpected egg in either box, given a definite order of inspection. The problem then is that, given that the egg is expected, where exactly is it? After all, the word expectation implies knowledge of location so it should be possible to state which box contains the egg. In fact, it could be in either box, or neither. Rather like a sub-atomic particle, the egg resides on a shadow line between existence and non-existence: it is slippery, being both everywhere and nowhere. Logically, it is both expected and unexpected. The unexpected egg paradox, therefore, is exactly analogous to that of the unexpected exam.
More fundamentally, like Hempel's ravens, the unexpected exam undermines the process of developing and supporting concepts by the serial aggregation of information. If this approach does not work, then perhaps it makes sense to collect information at random. But this is counter-intuitive, irrational and psychologically uncomfortable. It is unlikely that anyone engaged in research would ignore perfectly good data simply because some randomising device told her she ought to. Both rational and irrational approaches to decision-making here seem unacceptable. Further, it is no use arguing that the fault lies with the overly strenuous application of logic - if the proposed unexpected exam, for example, was treated as a less rigorous statement such as 'I intend to set you a test this week that you can't prepare for'. For it is exactly this lack of rigour that game theoretic styles aim to eradicate. We can't have it both ways. Game theoretical thinking as applied to induction leads to disastrous paradox. A relaxation of logic is no longer game theoretical. Either way, in a final paradoxical twist, rigorous logic destroys itself.

THE TREATMENT OF THE FUTURE AS A PARAMETER

Do I dare
Disturb the universe?
In a minute there is time
For decisions and revisions which a minute will reverse.

T.S. Eliot, The Love Song of J. Alfred Prufrock

Since the aim of game theory is to make decisions by optimisation, it is essential that all parts of a decision be manipulable and susceptible to calculation. To this end it is assumed that all parts of the decisional environment can be treated as parameters. This, as has been said earlier, can lead to proposed procedures which are unnatural and outcomes which are paradoxical.
Generally, and as a basis for calculation, the ability of individuals to define utilities is assumed by the theory. In practice, however, it seems that we neither understand how to construct them nor, more fundamentally, bother to construct them. Nicholas Rescher says in his book *Rationality* (p 109)

> People constantly make - or endeavour to make - rational decisions....Throughout, they are doing all sorts of well-advised and intelligent things - saving money, prolonging longevity, enhancing comfort, enlarging their friendships, increasing their knowledge, and the like. But to say they are throughout doing the same sort of thing - promoting 'utility' - is an eminently problematic contention.

If all decision-making transactions involved money, for example, it may not be too difficult to define utilities upon which to base optimal actions. The problem becomes very complex, however, when issues of morality or ethics are involved: can these be adequately modelled in terms of utilities? And if we are able to consider matters such as morality and ethics, are utilities and, more fundamentally, a rationality valid which assume self-interest as the only motivating factor?

If we find it difficult to accept that people construct utilities in practice, it is equally difficult to accept that people actually form stable preferences among them. There are two reasons for this inability to make preferences, both suggested by Martin Hollis in his book, *The Cunning of Reason*. First, preferences are not static and immutable. They change periodically. It is unlikely, for instance, that any individual has an optimal or permanently ordered shopping list. Tastes change, and things like shopping lists change to reflect this. Neither is this a reflection of changed circumstances - that an individual might, for example, have more money to spend this week. It is simply a reflection of the fact that people like something different from time to time. So an individual might like oranges, indeed prefer oranges to any other fruit. Any scale of utilities and preferences
might put oranges above everything else. We would not consider it irrational, however, if this week he bought apples 'for a change'. A second difficulty attendant upon forming preferences is the difficulty of level shifts. For example, a person who wants to lose weight might prefer to eat chips, but, in these circumstances, would probably prefer not to prefer them. Game theory, and other utility-maximising decisional procedures make no comment about this type of conflict over preferences. It is assumed not to exist, or if it does it is assumed that a rational individual will resolve it.

A further problem occurs when we try to treat all parts of the decision as static, manipulable or irrelevant. This applies, for example, where calculations are made of the effects that time might have on a decision. I will begin my discussion of this issue, by describing two scenarios which can be represented graphically.

The above graph shows two choices, a-a and b-b. It is obvious from the graph that a-a is a better choice, since it always does better than b-b. This outcome should be compared with the graph below. Here a-a has an advantage at present, but in the long term b-b is the better choice.
The second scenario is by no means uncommon. The choice of \( b-b \) has to be made, and adhered to, in the light of the fact that \( a-a \) is preferred at the present time. For example, we might know that eating fatty meat is not good for us, but we like it, and at the moment are faced with a particularly juicy steak. Game theory would not see this as a problem. The choice of \( b-b \) has to be made because it would be irrational to make any other: \( a-a \) is the Pareto inferior choice because it makes us worse off. Let me consider this issue in more detail by reference to the story of Ulysses and the Sirens. Ulysses wants to listen to the song of the Sirens, but knows that their seductive sound will lure him to his doom. This does not bother him now, since he cannot hear them. But his future self will. How can Ulysses ensure that his future self will act sensibly? As we know, he had himself bound to the mast of his ship, had his shipmates block their ears with wax so they could not hear the Siren's Song, and got them to agree not to release him no matter how he threatened them. In other words, his present self made a pre-commitment on behalf of his future self which his future self had to honour. Game Theory, on the other hand, would have suggested that Ulysses should not waste his time acquiring wax and rope, and convincing his men that they could disobey him without fear of reprisal, since this expenditure effectively reduces his payoff and changes his utility. His better course is to sail on in his present state of mind, and by an act of will ignore the Sirens at a later date, knowing that this is the optimal strategy. It is the easy choice to make in his present circumstances, and assumes that his future self as a separate entity does not exist, but is merely an extrapolation of his present self. I have said earlier, that
game theoretic styles of rationality tend to fail because they do not seem to relate to human nature, and because this tight style of logical thinking can lead to contradiction and paradox. Both weaknesses are illustrated by the above example. First, assumptions are made about the nature of will, the ability to will what cannot be willed, and about its constant or atemporal nature. Unlike the pre-commitment strategy, the game-theoretic approach implies perfect will: the ability to pursue a course for its worth outside of a time frame that might alter perceptions. It ignores statements such as 'it seemed like a good idea at the time' or 'I will hate myself for this', which exhibit a basic and thoroughgoing appreciation of the practical need for self-commitment. Second, the game-theoretic belief that utilities should be based on all information, including the deduced actions of a rational opponent, becomes paradoxical when one's opponent is one's future self. One's present self bases actions on the presumed actions of one's future self, which will be based on the actions of one's present self, which are based on the actions of one's future self, and so on. We are being asked, essentially, to construct a model which includes a model of itself, rather like constructing a television picture with a camera pointed at its own visual display unit. The picture would consist of a picture of the screen containing a picture of the screen containing a picture of the screen, and so on. This infinite regress is the logical outcome of the dictum to treat all items in the decisional environment as parameters, even ourselves. The paradox arises from this self reference. It can equally, therefore, be captured in a self-referential statement such as:

*I will not pre-commit*

This paradox of self-reference is obvious. If it is true, then it presumably includes itself as a pre-commitment. Therefore, it negates itself. If it is not true, then pre-commitment is acceptable, which presumably includes the original statement not to pre-commit.

The acceptance of pre-commitment as a rational activity seems, therefore, to make much more sense. By advocating the binding of future selves, it allows Ulysses to get out of
the paradox of vicious circularity. By taking the course of action he actually takes, Ulysses does not need to worry about the problem of his future self's activities because his action in the present has effectively circumscribed them. Thomas Schelling, in his book *Choice and Consequence*, takes the whole issue a stage further. He sees the two selves as locked in a complex game of conflict and co-operation. As we will see when we discuss such games, non-zero-sum games, in the next chapter, their outcome is hard to predict. Schelling's conclusions are worth quoting in full, because they underline the paucity of game-theoretic thinking (p 93-94).

(In the) two value systems that are alternate rather than subject to simultaneous and integrated scrutiny, 'rational discussion' has to be replaced with something like collective choice....So we should not expect a person's choice to display the qualities typically imputed to a rational decision, like transivity, the irrelevance of 'irrelevant' alternatives, and short run stability over time. We should expect the kinds of parliamentary strategies that go with small group voting behaviour, and the second best choices that have to be made when rights and contracts are not enforceable. Depriving oneself of certain preferred opportunities - suppressing certain states that economists call 'Pareto superior' - because the other self would abuse the opportunity, becomes an expected tactic.

I will return to the issue of self-binding, pre-commitment and the nature of will in the next chapter.
THE BELIEF IN THE PERFECTIBILITY OF INFORMATION

But I at once resolved myself to collect and correlate all the information I could discover regarding the case. If I could not employ physical force against the scholar, I could at least write a defence of the teacher, or more exactly, of the good intentions of an honest but uninfluential man. I admit that I rued this decision later, for I soon saw that its execution was bound to involve me in a strange predicament. On the one hand my influence was far from sufficient to effect a change in learned or even public opinion, while on the other the teacher was bound to notice that I was less concerned with his main object, which was to prove that the giant mole had actually been seen, than to defend his honesty, which must naturally be self-evident to him and in need of no defence.

Kafka, The Giant Mole

The typically Kafkaesque situation described above, where the narrator finds himself obliged to find out more in order to gain less, stands in contrast to the game theoretic view that more information is better, and that individuals aim to obtain perfect information which will yield an optimal decision. We have seen in the previous sections that it is debatable whether people actually use information fully, and whether they do arrive at decisions by the use of algorithm, utility, Pareto optimality, and so on. I wish now to discuss whether the aim of gaining perfect information is any more natural, and also if it is prone to paradox.

Hollis, in his book cited earlier, suggests two problems which can arise when we try to obtain information, and which are obvious. First, the information we receive changes us and our views. We can, by an act of will remember, but cannot by an act of will, forget. Consequently, the information we assimilate changes our view permanently: we cannot wish ourselves back to a time when we didn't have the information. We are stuck with it. Second, we often don't know the value of information until we acquire or experience it.
In fact, its value only becomes apparent once we acquire it. To these I would add a third difficulty, which again is obvious. This is that when we obtain information, we often acquire more than we expect or intended. Taking these three items together, I would argue that, rather than being a rational activity in game-theoretic terms, information acquisition is much more of a leap in the dark. The situation termed the Buyer's Curse or Buyer's Paradox is an example of this. Suppose I see something in a flea market which I think is valuable. I make an offer well below its value expecting to be haggled upward. To my surprise, the offer is immediately and fervently accepted by the stall-holder. I now suspect that my original assessment of the object is incorrect, and that I have probably paid more for it than I should. If I could replay the situation, given this new item of information, I would make a different offer. Unfortunately, I cannot. I only get this information from the experiencing of the situation. It is information I did not expect, and now I am stuck with it. The Buyer's Curse exactly mirrors the three problems of information acquisition described. Since we encounter these kinds of situations in whole or part often, we learn to treat information acquisition with care. For example, I recently visited Mr Straw's House, a National Trust property in Worksop, whose contents is *melange* of domesticia collected mostly between the wars. My visit was based on encouraging reviews both in the press and Trust literature. Having visited it, I shall now be far more circumspect about reviews in the press and Trust literature. If I had known how disappointing the property was to be I wouldn't have made the visit, but I couldn't know it without making the visit first. This is exactly the issue pointed by the story of Adam and Eve: that the drive to know has attendant problems, particularly where we know only by experiencing and then we are stuck with what we know. In many cases it may be preferable not to know - that ignorance is bliss. I would argue, therefore, that both as individuals and cultures humanity has developed a healthy scepticism about knowing. I intend to discuss this issue more fully in chapters 4 and 5 where I consider the issue of dissonance reduction: the procedure by which we tend to make decisions that ensure psychological comfort rather than objective optimality. The refusal to seek out information is one aspect of this strategy.
Although the acquisition of information may not always be psychologically conformable, this does not make it paradoxical. I do feel, however, that the problems attendant on information acquisition whiff of paradox. In particular, the Buyer's Curse, where the more you know the less you wish you knew has a sort of circularity. However, if we consider the perfectibility of information we do encounter a genuine paradox. Game theory argues that all individuals in the decisional milieu have perfect information. Let us consider the implications of this. Suppose a group of game theorists are going out for the day. First, we can assume they all have perfect information of the local beauty spots. Second, we can assume that they do not want to get involved in traffic jams and go to places full of other people, since this would reduce the utility of these places and the day out generally. Knowing that everyone has these preferences, and that everyone has perfect information including knowledge of each others preferences, then where do they choose to go? Obviously, it is no use picking the most favourable beauty spot, since many others will pick it, thus reducing its utility. Would it make sense to pick the least favoured, since although its actual utility is not so great, it is better by comparison with other places since it will be quieter? Unfortunately, since everyone has perfect information, they will argue in exactly the same way. So why not pick the best spot again, the one with the highest original utility? Again, because everyone else will reason similarly. In fact, as we struggle to define what we mean by 'best', it becomes evident that the final answer depends on the level at which we decide to operate. The best place could be the one which is objectively the most enticing or desirable; or the actual place where everyone else doesn't end up going; or the place which people beforehand will not decide to go to; or the places people generally do not go in this type of situation. Each definition involves a level of abstraction above the previous, and above, as a basis, the intrinsic nature of the place itself. Unfortunately, it becomes impossible to decide at which level to exit, since each is valid. Worse, as we spend more time debating the issue of where to go, we are reducing utility since we could already be on the road enjoying ourselves. So a further problem emerges: how long do we spend debating the problem? This will affect our calculations directly because everyone else will be having the same
debate, with a consequent effect upon when they actually hit the road. Game theory would dictate, therefore, that we terminate our debate at some time and find a level of decision from which to operate. Suppose we decide on the level which decides beforehand the optimal place based on the place others will decide beforehand not to go. But we now have the problem of how to predict this place since every other equally rational individual is trying to make exactly the same prediction. In the previous section, I showed how game theory becomes paradoxical when the other player in the game is one's future self. There the paradox lay in the nature of pre-commitment. In the present scenario, the assumption that other players all have perfect information leads likewise to paradox. This time the paradox lies in the nature of prediction, where the actions of the predictors affect the prediction. These predictive situations are essentially different from predictions of events which cannot be influenced. An example of both sorts might appear in a news bulletin. The weather forecast indicates that it might rain. The rational person might, consequently, take an umbrella in the car as they set out to work. The news bulletin continues with a roadwork summary, indicating a traffic build up on the main road into town. What does the game theoretician do now? If she avoids the congestion, then presumably everyone else will. This will not only congest the alternative routes, but clear the congestion on the main road. Is it more sensible, therefore, to carry on down the main road? But all other motorists have the same information and won't they presumably do the same? Action by self-reference affects the prediction and leaves the motorist wondering what to do. Paradoxes of prediction such as these make it impossible for game theoreticians to maintain the view that perfect information is essential to rational decision-making. On the contrary, the circularity of argument can only be escaped if it is assumed that some individuals do not have perfect information. If, for example, our car driver can assume that most people do not know about the traffic congestion, and that those that do will approach it in certain ways, she can begin to make more concrete deductions. Put another way, she can assume that the congestion prediction is not being influenced by everyone trying to re-predict the prediction. She can
treat it more as the weather prediction - as the extrapolation of natural events - and act accordingly.

The Paradox of Perfect Information manifests itself wherever people make predictions which include the predictions as part of themselves. If the above traffic jam example seems trivial, consider instead the stock exchange and the totalisator. Both reflect the predictions or bets of large numbers of individuals, the outcomes of which, in the movements of stock and horse race prices, affect further predictions. I will have more to say on the paradoxical nature of predictions which affect themselves when I consider Newcombe's Paradox and its implications in the next chapter.

THE USE OF FORMAL SYSTEMS FOR MAKING DECISIONS

I shot an arrow into the air
It fell to earth I know not where.
H.W. Longfellow, The Arrow and the Song

I would finally like to consider whether the whole structure of rational decision-making as described in this chapter can be self-consistent. Generally, it seems that people are anything but rational in the game theoretic sense when they make decisions. They seem to base their decisions on a whole ensemble of techniques, such as gambling, seeking obvious but not necessarily optimal agreement points, and so on. These game theory sees as irrational. In fact, my own research into this issue, described in Chapter 4, support this view that decision-making in these situations is irrational in game theoretic terms.

The paradoxical nature of narrowly defined decisional procedures or techniques are well illustrated by the work of Kenneth Arrow. His work not only shows the inherent contradictions in formal systems for making decisions, but also acts as an analogy for the more general weakness of narrow rationality. Arrow set himself the task of finding the
minimum conditions necessary to translate a set of individual selections into a group selection that fairly represented them. The conditions he felt that needed to be fulfilled were as follows:

**Freedom of choice.** There should be two or more voters, selecting from three or more choices. Any ordering of these choices by individuals should be allowed.

**Positive association between individual and societal choice.** If x is preferred to y as a societal choice, then this preference ordering should not change if individuals retain their orderings with respect to x or change their orderings in favour of x.

**Independence of irrelevant alternatives.** If w is preferred to x which is preferred to y which is preferred to z, then if, for example, y is discarded, the order of preference of the remaining items (w preferred to x preferred to z) should be retained.

**Citizen's sovereignty.** For a pair of alternatives x and y, there must be a mechanism which allows the societal statement x is preferred to y. If this were not possible, then the choice x is not preferred to y is imposed by and on society, no matter what individuals might choose.

**Non-dictatorship.** If any individual prefers x to y, then the societal view should not consist of this regardless of the preferences of others.

As I have said, I intend to use Arrow's proposals both as a direct illustration of the paradoxical nature of problem solving based on narrow rationality, and as a paradigm for the inconsistency of game theoretically rational systems generally. A consideration of the list of conditions emphasises its exemplary nature. It is a sensible, rational and minimal statement of what a voting system should provide. It allows for society to fairly reflect all possible views and changes in views. Neither society, nor any particular individual, will
be powerful enough to override the general will. The relationship between this rational system, and our general statements about game theoretic rationality are obvious. First, the system assumes that individuals are clearly able to order their preferences, and implies thereby some kind of system of individual utilities (Freedom of choice). Second, it explicitly states that the system should be consistent, particularly in its ability to produce consistent orderings under changing circumstances (Positive association between individual and societal choice; Independence of irrelevant alternatives). Third, the system is exhaustive, in that all choices must be available and represented (Freedom of Choice; Citizen's Sovereignty). Fourth, the whole list implies that a rational voting system can be produced by the application of a set of axioms such as these. In many ways, this list is highly representative of the sort of rationality game theoretic thinking produces. It is also self-defeating! Although each factor seems essential to a fair voting system, Arrow showed that all five could not be consistently applied together. In order for a system using these conditions to work, some have to be left out. I will illustrate this with reference to four examples.

Example 1

We decide to ask three people (A, B and C) to list their preferences in the following choice of foods: steak, beans, cabbage. A likes meat and needs protein. B needs protein but is a vegetarian. C prefers greens but likes meat. Their individual rankings, which are quite clear for each of them, are as follows:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steak</td>
<td>Beans</td>
<td>Cabbage</td>
<td></td>
</tr>
<tr>
<td>Beans</td>
<td>Cabbage</td>
<td>Steak</td>
<td></td>
</tr>
<tr>
<td>Cabbage</td>
<td>Steak</td>
<td>Beans</td>
<td></td>
</tr>
</tbody>
</table>
It is important to understand that, as far as each person is concerned, these preferences are transitive. That is to say, that if a first item is preferred to a second, and a second item is preferred to a third, then the first item is preferred the third. This final preference does not have to be stated: the transitive nature of the choices forces it.

Let us now generate the group preference. This can be achieved by counting how often one food is preferred to another. Steak is preferred to beans by A and C, the reverse preference being held only by B. Therefore, steak beats beans 2 to 1, and so is preferred by the group. Likewise, beans are preferred to cabbage by 2 to 1. Since steak is preferred to beans, and beans are preferred to cabbage, then, logically, steak should be preferred to cabbage. Unfortunately, once we count the preferences, we find that cabbage is preferred to steak! The aggregation of individual, transitive preferences, has produced an intransitive group preference. Although each individual can state their preferences with confidence, the group is incapable of making up its mind. Any attempts to pin the group down end up in intransitive circularity. The whole attempt to aggregate items produces a nonsensical, self-contradictory result. It is tempting to argue that this is a highly contrived, artificial example. Would this happen with more voters and more preferences? Here is an example with forty voters and four choices, generated by computer. Each voter has placed the four choices (A, B, C, D) in order of preference.

<table>
<thead>
<tr>
<th>BDAC</th>
<th>BCDA</th>
<th>CDAB</th>
<th>ABCD</th>
<th>ABCD</th>
<th>CBDA</th>
<th>CBDA</th>
<th>BDCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>DACB</td>
<td>BCAD</td>
<td>DABC</td>
<td>ABCD</td>
<td>ABCD</td>
<td>DBAC</td>
<td>BCDA</td>
<td>ADCB</td>
</tr>
<tr>
<td>DBAC</td>
<td>DCBA</td>
<td>CABB</td>
<td>ACBD</td>
<td>BCAD</td>
<td>DABC</td>
<td>DABC</td>
<td>ABCD</td>
</tr>
<tr>
<td>CABD</td>
<td>DACB</td>
<td>CABB</td>
<td>CABB</td>
<td>ABCD</td>
<td>ACDB</td>
<td>CDBA</td>
<td>ACBD</td>
</tr>
<tr>
<td>ABDC</td>
<td>CDBA</td>
<td>BDAC</td>
<td>CDAB</td>
<td>ACBD</td>
<td>DABC</td>
<td>CABD</td>
<td>DBAC</td>
</tr>
</tbody>
</table>

An examination of the preferences shows that the group prefers A to C, C to B, B to D and D to A. The preferences are, therefore, completely cyclical. The individual transitive preferences have produced an intransitive, indecisive group preference. In fact, as the
number of voters increases, the chances of such results increase. There is no method for generating a set of preferences, so citizen's sovereignty is contravened. In order for society to generate a transitive preference, freedom of choice or non-dictatorship must instead be contravened.

Example 2

A group of people wish to select between four options (A, B, C, D.) If, for example, there were 26 people, and two options were desired, they could vote using the following rules.

1. Some threshold should be decided which, if achieved, would ensure the selection of that option. Since there are 26 voters, then this threshold should be 9, since the selection of two options would leave a maximum of 8 votes uncast. This would be insufficient for a third option to reach the threshold.

2. Each voter rank-orders the preferences

3. Any option reaching the threshold is selected. Any votes over the threshold are passed on to second choices as fractional votes.

4. If no option reaches the threshold, then the bottom option drops out and its second choices receive its vote

Let us suppose that votes are cast as follows:
A in first place with 9 voters, is selected. There are no surplus votes over the threshold to be transferred. Since no second option has reached the threshold, B is eliminated. This transfers 5 votes to C (the second choice of all voters who put B first,) which gives it a total of 11 and makes it the second choice.

Let us suppose that, instead, the two voters who placed D in first place above Chad actually preferred C to D. This would give C 2 more votes and recast the table as follows:

<table>
<thead>
<tr>
<th>Number of voters</th>
<th>Order of preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>A       B       C   D</td>
</tr>
<tr>
<td>8</td>
<td>C       D       B   A</td>
</tr>
<tr>
<td>4</td>
<td>D       B       C   A</td>
</tr>
<tr>
<td>5</td>
<td>B       C       D   A</td>
</tr>
</tbody>
</table>

Again, A wins. This time, however, D has only 4 first-choice votes and is eliminated. The second choice of these voters, B, receives 4 votes, achieves the threshold and is selected. So by gaining two votes more than in the previous scenario, C fails to be selected! B, which was eliminated in the first game, is selected in the second, even though its number of votes changed in no way at all. This negates the positive association between
individual and societal choice. Again, this can only be achieved if the strictures on freedom of choice, non-dictatorship or citizen's sovereignty are relaxed.

Example 3

Five people have to select the group preference from three issues, \( x \), \( y \) and \( z \). In order to achieve this they employ the nominal grouping technique, which gives a weighting to each issue according to how each individual orders them. By this approach, also known as a Borda Count, each individual will give 3 points to their most favoured issue, 2 to the next, and 1 to the least favoured. (Obviously this procedure could be extended from \( n \) to 1 for \( n \) issues.) The five individuals order the issues as follows:

\[
\begin{array}{cccc}
X & X & Y & Y \\
Y & Y & Z & Z \\
Z & Z & X & X \\
Y & Y & X & Z \\
\end{array}
\]

Which yields a score of \( x = 10 \), \( y = 11 \), \( z = 9 \), and makes \( y \) the group choice. Unfortunately, although \( y \) wins, it can be seen from the preference list that more individuals prefer \( x \) to \( y \). Citizen's sovereignty is again contravened. Worse, suppose that the individuals are allowed to change their minds, and they do so, but only with respect to \( z \). In other words, their preferences in respect of \( x \) and \( y \) remain exactly as before. Such a new ordering might be:

\[
\begin{array}{cccc}
X & X & Z & Y \\
Z & Z & Y & X \\
Y & Y & X & Z \\
\end{array}
\]

The scores are now \( x = 12 \), \( y = 8 \) and \( z = 10 \), so \( x \) wins. However, as we specified, the orderings of \( x \) and \( y \) with respect to each other have not changed. Only \( z \)'s status has
changed, and its status would appear to be irrelevant to \(x\) and \(y\). Nevertheless, this irrelevant change has affected the final outcome. It would appear, therefore, that independence of irrelevant alternatives has been contravened.

Example 4

A committee of seven individuals has to elect another member from A, B, or C. They have the following preferences, but vote only for their most preferred one (as, for example, in a British Parliamentary election). A simple majority decides the winner.

<table>
<thead>
<tr>
<th>A</th>
<th>A</th>
<th>B</th>
<th>B</th>
<th>C</th>
<th>C</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

An inspection of the preferences shows that B is preferred to both A and C, and A is preferred to C. Please note that these preferences are not cyclical: B is the most preferred and C the least preferred. Nevertheless, C wins because she obtains 3 votes to the others' 2. This result is imposed on society by the system, so citizen's sovereignty is again contravened.

It can be argued that these are all contrived examples. They are, but only to make the issue easy to understand. There is evidence for several having actually occurred. For example, and with reference to the last of the above examples, Andrew Colman, in his book *Game Theory and Experimental Games*, cites his examination of the results of the 1966 General Election. Based on information about preferences in the 261 constituencies contested by the three major parties alone, 15 constituencies elected MPs who actually had other candidates preferred above them. Basically, the act of aggregating votes in any way is a paradoxical activity. These formal ways of finding a group view generate intransivity and contradiction. In practice, citizen's sovereignty is also difficult to
achieve by less formal ways of obtaining a group view. As we will see in the next chapter, social groups often cannot represent the desires of the individuals that compose them. Given the self-defeating nature of the list, some item must be eliminated to make the rest work. But which item? An examination of the list shows it to be minimally sufficient: to compromise any part leads to reduction in choice or power of either society or the individual. Let me emphasise that the inability of this list of items to work as a group is not a product of weakness of will on the part of the voters. Arrow's findings are mathematically determined, an outcome of the same iron logic which necessitated their inception in the first place. As teachers we are fond of forming working parties and groups, and making decisions based on the aggregation of individual or smaller group decision. Unfortunately, there is no foolproof way of ensuring that the final decision fairly reflects the balance of those of the individuals. It is the weakness of the situation, not the people involved in it, which is critical to failure. Again, an attempt at narrow rationality leads to confusion. Too often do we leave a meeting with the feeling that the decision we have reached has satisfied no one. Too often we are left to ponder the efficacy of logic as a basis for rationality.

CONCLUSION

Many decision-making difficulties stem from the desire to make rational decisions of the game theoretic type, since their mathematical and procedural basis gives them authority based on apparent rigour. An easy acceptance of aspects of this rationality seems to be a feature of modern life. But their general reliance on objectivity, logic and procedure mean that they often do not easily conform with the decisional parameters which people naturally and actually apply. Too often, game theorists operate on the basis of how we ought to think, not on how we actually think. Worse, this narrow rationality can lead to paradox. The paradox turns partly on self-reference - for example, where we need to consider our future selves as a separate contestant in the game - and partly on level shift, as we need to decide at what level of strategy to operate - at the level of utility
maximisation, or deciding how to maximise utility, or deciding how to decide how to maximise utility. If we assume that everyone is having the same debate, it becomes impossible to decide at what point we stop reasoning and start acting. Arrow's findings are the bleakest of all. Fond as humanity is of social decision-making in groups, this activity often fails at a procedural level. There is no foolproof formal way of deciding what the group actually wants.

There seems, then, a need on the part of individuals and groups to accept a broader, less rigorous, more pragmatic view of rationality. We may, therewith, pursue decision-making strategies that are satisfyingly rational, but which will also nullify paradox. I will consider this issue in the next chapter.
Chapter 3
Paradox Lost and Paradox Regained

One thing that makes Pascal’s reasoning peculiar is that it argues that it’s a good idea for you to believe something on the basis of cost-benefit analysis. It does not give you the normal sort of reason in favour of believing something - namely, reason to think it is true.

Robert Martin, There Are Two Errors In The The Title Of This Book

I began this work with a proposition that not all failure is based on weakness of will: that some situations are inherently difficult and any amount of effort will not effectively master them; they baffle or perplex in their own right. I showed in the last chapter that some of these situations stem from a perhaps overly narrow view of what constitutes a rational action or decision. In this chapter I wish to consider other views of rationality, in order to arrive at a broader definition, and hopefully dispose of the problems encountered in the previous chapter.

RATIONAL BEHAVIOUR

No one who can earn great wages or who possesses wealth will, out of zeal for philosophical doctrines, prefer to live in squalor and want

W G Sumner, Folkways

I suggested in the previous chapter that it is unfair to condemn humans as irrational, simply because they are unable to function as game-theorists would like. It is more probable that, in actual terms, most people are rational most of the time. We could, in fact, agree with Rescher, in his book cited earlier, that rational intelligence ‘is the survival
Paradoxes in Decision-making  Paradox Lost and Paradox Regained

instrument of *Homo Sapiens*. So a definition of rationality might more usefully arise from an observation of actual practice. I feel, therefore, that an attempt to relate rationality to human characteristics other than selfishness or logicality is required to make progress. I would like, therefore, to consider a definition of rationality used by Stuart Sutherland in his book *Irrationality, the Enemy Within*. As the title suggests, the substance of the book is predominantly a catalogue of what Sutherland considers to be irrational. Some of the incidents cited are taken from actual events, some from sociological or psychological experiment. The whole provides basic support for the suggestion, in the title of the book, that irrationality is not only an all pervasive aspect of the species (one wonders, after reading the book, if mankind is actually capable of rational thought or action), but that irrationality is also a bad thing - an enemy. (I have already suggested that the imputing of value to the terms rational and irrational may not be a good thing.) I will describe the main kinds of irrationality which Sutherland considers. This will lead into a more general discussion of whether Sutherland's definition of rationality captures it sufficiently well, or whether there is more to be said. I shall consider then the work of other writers on rationality. Most importantly, Sutherland's work raises the question of whether a definition of rationality needs to be rational. Can it, instead, be ambiguous, intuitive, or indeed a paradoxical definition? As we shall see, the question of the nature of rationality becomes so complex that a definition may have to consist only of a series of examples of rational and irrational actions and thoughts. To an extent, this is exactly what Sutherland provides.

I will begin with Sutherland's definition of rationality:

Rational thinking leads to the conclusion which is most likely to be correct, given the knowledge one has.....a rational action is one that, given the person's knowledge, is most likely to achieve his end.
This is a different type of definition from the game theoretic one discussed in the previous chapter. It is more succinct, possibly simplistic. Most obviously, it is outcome rather than process oriented. Since it is a broader, vaguer definition of rationality, it will necessarily include as rational many more actions than the previous one. The definition has the immediate merit of appearing more realistic and down-to-earth than the game-theoretic definition, so as the start of a search for rationality in practice, it is as good a definition as any. This move from game theory to practical considerations of rationality allows problems to be viewed from a different, perhaps more realistic angle. Hempel's Paradox and the Unexpected Exam, therefore, cease to be a problem if we stop logic chopping, and accept instead the approximate, quasi-intuitive nature of induction, a process which nevertheless has stood mankind in good stead for several thousand years. Arrow's axioms may be unattainable in practise, but they nevertheless make sense, and even if they are compromised from time to time then they are better than nothing. Finally, Sutherland devotes several pages to a discussion of the sensibleness of pre-commitment. In fact, Sutherland's book is a positive cornucopia of rationality. I feel the book's great power and value lies exactly in this extensive material. It needs no commentary, and even his definition of rationality is irrelevant to the scope of the work. Instead, we can perceive his rationality as a day-to-day and pervasive thing. Not only are we all immersed in it, but we can recognise the exact locus of our actions within the incidents described in this book. The well-springs of rationality, therefore, do not lie in some esoteric state of strategic thinking, where we find ourselves when as individuals we involvourlves in game theoretic play, or as a species consider the reasons for proliferation of nuclear weapons, or as a nation muse upon the aptness of the market ethos. To Sutherland, we are workaday rationalists; when we drive a car, smoke a cigarette, fall in love, vote, attend an interview, join a committee, meet the boss, eat in a restaurant, and so on. At almost every turn, on an individual basis, we encounter the need for rational thought. Within this melange of rationality lies a second strength of Sutherland's book: it contains references to almost every walk of life and activity.
Having considered the strength of Sutherland's work, his contention and evidence that rationality is a prosaic, day-to-day, affair, it is now necessary to turn to its weaknesses. First, although concerned to discuss rationality in general, he sees it manifest itself in practice as irrationality. Life, for him, is a veritable minefield of non-reason, a sort of physical embodiment of Murphy's Law - whatever can go wrong will. And the reason things go wrong is not explained by conspiracy theory or impersonal forces, because belief in these is itself irrational. Neither are we victims of egotism nor self-interest. Things go wrong because of human frailty and lack of knowledge. They cause us continually to deal with events in an irrational way. His definition still emphasises rightness, consistency, and the need to know. For him, as with the game theorists, rationality 'comes out' rather like a sum. And the sum is worked, in the end, by doing exactly the sorts of things game theory suggests: pull your socks up, find out more, calculate utilities, see all sides of a situation. He sees no place for, or the effect of, inconsistency, contradiction or paradox. In the end, therefore, his work promises much but delivers little. His attempt to see rationality as down-to-earth ends with his condemnation of this very earthiness. People are irrational because they think and act like people. If only they would stop thinking and acting like people then they would become rational. But, and this is what Sutherland and the game theorists ignore, they would also, by the changed nature of their thought and action, cease to be people. Sutherland makes the statement:

*People are not rational*

but ignores the question of whether such a statement can be made by a rational person. In fact, it defeats itself by self-reference, as does Sutherland. His catalogue of irrational act based on observation, from which he derives a belief in the irrationality of people, condemns himself as a person, and presumably his statements about irrationality, as irrational. It can be argued that my argument here is highly contrived and a perversion of logic, but it is exactly this type of algorithmic, logical and objective thinking which is
Paradoxes in Decision-making
Paradox Lost and Paradox Regained

proposed as rational by game-theorists. It is exactly this style of thinking I wish to abandon. Instead, I want to find a more natural rationality in the richness and diversity of human thought and action. I want to make the statement:

*People are rational*

a statement which does not fall foul of any logical inconsistency or paradox.

To begin this search for this new definition of rationality I want to return to Sutherland’s work. Early on, he elicits from his extensive catalogue of irrationality five basic reasons for irrational thought and action. Each reason sheds light on rationality generally, but I feel his obsession with seeing these items as evidence of irrationality means that he does not explore their significance fully. I shall describe each item, posing, at the end of each description, a question which will form the basis of the more general discussion of rationality and paradox which will make up the rest of this chapter.

1. People do not spend enough time untangling and thinking through any problem. In particular, Sutherland sees the exercise of rationality as both spending time thinking through a problem and knowing when any more time spent thinking rather then acting would itself be irrational.

   *Is there a distinction between rationality - thinking about an issue - and meta-rationality - thinking about thinking about an issue?*

2. People have only a sketchy and often erroneous grasp of statistics. In a numerate society, this leads to irrational thought and action. Sutherland argues that the innate human desire for structure and pattern in sensory input produces irrationality in that numerate or statistical events are often counter-intuitive in their pattern.

   *Is anything not rational simply irrational, or are there levels of rationality, perhaps based on the difficulty of being rational in a particular circumstance?*
3. People act irrationally because they want to feel comfortable, rather than solve a problem in a more realistic or sensible way. This need for comfort leads to wishful thinking and self-deception. The attainment of security, in this instance, is a sort of by-product of the action.

Is it rational to undertake an activity to achieve, not the end of the activity in itself, but some by-product of the activity?

4. People act irrationally because they can only hold a few ideas in their heads at any time, and most activities may have many parameters. Effectively, therefore, humans suffer from information overload, and take few steps, such as writing things down, to cope more easily with it.

Given the nature of Man and the world in which he finds himself, is it sufficient to try to be rational, even if the resulting action is not the best in the circumstances?

5. The structure of groups encourages selfish action on the part of individuals within the group. Interestingly, Sutherland does not see selfishness as irrational - "the selfishness of members, however immoral, is not irrational" - but believes that this collection of individual rationality leads to group irrationality. This seems to contradict his conviction that irrationality is a product of human weakness.

Can individuals be blamed if their individual rationality is warped when they form a group?

There follows an attempt to outline the requirements of a more comprehensive definition of rationality, based on the questions posed above. In order to facilitate this, I will re-iterate the questions I have posed, giving each a sub-title (italicised) which will provide the headings for the discussion which follows.
Is there a distinction between rationality - thinking about an issue - and meta-rationality - thinking about thinking about an issue? Rationality and Meta-Rationality.

Is anything not rational simply irrational, or are there levels of rationality, perhaps based on the difficulty of being rational in a particular circumstance? The Boundaries of Rational Behaviour.

Is it rational to undertake an activity to achieve, not the end of the activity in itself, but some by-product of the activity? Willing what cannot be willed.

Given the nature of Man and the world in which he finds himself, is it sufficient to try to be rational, even if the resulting action is not the best in the circumstances? Means and Ends.

Can individuals be blamed if their individual rationality is warped when they form a group? The Problems of Aggregation and group decision-making.

In other words, I think a definition of rationality should not only consider the outcomes of rationality (Sutherland's definition), or propose ideal procedures for decision-making (game theory), but should also consider the nature of rational thought, its extent, and its transferability as a concept. The above headings, I hope, capture these features in certain ways.

Since I began this chapter with a consideration of Sutherland's definition of rationality, I shall continue to use it as a basis for a broader definition of rationality.
Rational thinking leads to the conclusion which is most likely to be correct, given the knowledge one has... a rational action is one that, given the person's knowledge, is most likely to achieve his end.

In order to make this more easily fit in with the five headings I have derived, I shall restate it, using a similar one proposed by Martin Hollis, in a paper entitled *Rational Man and Social Science*. He says that S acts rationally in doing a if:

1. S wants to achieve g
2. S has a choice among alternative ways of achieving g
3. S believes that a is the best way to achieve g

A comparison of the two definitions reveals their similarity. First, the use of the active voice in the wording of Sutherland's definition implies the wanting expounded in Hollis' first axiom. The use of 'most likely' in Sutherland's definition implies the 'choice' in Hollis' second statement. 'Given the knowledge' in Sutherland equates to 'believes' in Hollis' final statement. Hollis, in his paper, went on to disagree with his definition: to use it instead as a basis of a broader definition. I mean to do the same, and will base my arguments partly on his. I intend, however, to go further. I hope to show that any attempt to make statements about rationality can, in certain circumstances, create insurmountable problems. Like Arrow's statements about voting systems, statements about rationality can tend to be self-defeating as their context changes. At the end, I will argue that all views of rationality may contain paradox, so we may have to accept it and utilise it, or avoid it.
RATIONALITY AND META-RATIONALITY

A lady once asked Dr Johnson why he had defined 'pastern' as a horse's knee. 'Ignorance, madam, pure ignorance', the sage replied

Martin Hollis, *The Cunning of Reason*

The above quotation is an excellent example of a level shift in thinking. Although Dr Johnson is questioned about a particular definition, his reply concerns his general ability to define. Decision-making often exhibits a similar change of level. At times we might make decisions about issues: shall we, for example, take an umbrella to work? At other times we make decisions about decisions: how, for example, do we generally decide whether or not we are going to take an umbrella. I have discussed level shifts already on several occasions. These can be a formidable test for theories of rationality, because their tendency towards circularity and infinite regress is particularly perplexing. The need to consider the problems of level shifts is particularly important to management groups. As I suggested in Chapter 1, management groups operate in areas where they have to both decide on issues and decide on decisions: they have to consider both the management of issues and the management of management. I shall consider the implications of this situation for managers fully in the next chapter. At the moment, I want to examine the implications of level shifts for rationality, particularly in relation to the application of skills and the taking of decisions.

Decisions about skills

Hollis believed that statement 2, S has a choice among alternative ways of achieving g, may not actually be necessary. Hollis argues that conscious deliberation is not essential to rationality. There are, he says, several rational acts that are not consciously deliberated. For example, the best way to drive a car is not to think about it, but to follow the habit which, once acquired, takes the place of deliberate thought. Several of
our basic social activities, he argues, are based on such habits. It is difficult to accept this view, particularly since the act of deciding implies some kind of choice, even where habit is involved. Although several basic physical and mental processes are automatic, where fresh problems need to be solved, a deliberated decision will be necessary in order to decide which skills to use and when. Further, the habits involved in, for example, driving a car are habits acquired when the decision is taken to learn to drive a car, and are often acquired after a great deal of deliberation about them. In other words, the driver comes to accept that the best way to drive a car is to leave it to habit. But this is not a self-evident truth, and is only arrived at during the act of learning. The learner driver, therefore, learns two lessons: or more correctly, a lesson and a meta-lesson. She learns the skills of how to control a car, and the meta-lesson that these skills are best practised until they become habitual, because this is the best way to apply them. This lesson/meta-lesson concept applies to a whole range of learning situations, especially auto-didactic ones. When a person, for example, buys a 'teach yourself golf' book, he will need to explore this issue carefully. At some point he will decide that simply imagining that you are playing every shot with a handkerchief under your right armpit does not work, and that practising until the swing becomes habit does. Nevertheless, he will have to make the decision to change his approach to the game, and organise his lifestyle differently to accommodate it. Finally, skills-based activities are constantly controlled by deliberation. Playing a musical instrument is fundamentally a skill practised until it becomes a habit, but the actualisation of the habit is continually modified by deliberate decision.

**Decisions about decisions**

I would argue, therefore, that Hollis' axiom 2 cannot be discounted: choice seems too bound up in rational activity, indeed in the human condition generally, to be ignored. Even if one has acquired a skill or habit, one still has to take the decision to apply it. Choice, therefore, might operate both on the issue itself or on the strategy to enable the issue: to make choices about the issue or make choices about choices about the issue -
meta-choices. Further, styles of decision-making might change as the level of decision changes. The potentially paradoxical nature of this shift in levels and styles was summed up by Pascal when he said that nothing is more rational than the abdication of reason. Irrationality on one level may become rationality on another. The game of Chicken provides an example of this concept. It was originally described by Bertrand Russell, and was used as an analogy for cold-war international relations. The description purported to be based on American teenage game. The two players face each other in their cars along a road, then start to drive towards each other along the central white line. The first to swerve off is the chicken. This game has been extensively analysed because it is strongly analogous to actual situations, and I will describe it in more detail below. Like all games, however, the question arises of how to win it: how to force the other player to swerve off. Russell proposes that the best way is to act irrationally: to appear at the meeting drunk; to wear dark glasses to obscure vision; even to throw the steering wheel out of the window. The opponent, assuming that such a player is mad, possibly has a death wish, and is incapable of controlling the car anyway, will be inclined to swerve off at the first opportunity. In other words, the rational way to play the game is to act irrationally. The game becomes more interesting when both players adopt this strategy. As I have said, however, I will leave a fuller analysis of the game until later.

In summary, therefore, we can choose to act irrationally within terms which circumscribe rationality and irrationality. This larger rationality, which encompasses both rationality and irrationality, we can call *meta-rationality*. This meta-rational view will itself have areas that it considers irrational. However, we may decide, as appropriate, to introduce certain of these irrational ideas under the auspices of a meta-meta-rationality. An infinite regress opens up before us, as it becomes impossible to decide at which level of rationality to operate. A simple example of this level shift occurs, I think unintentionally, in Sutherland's book. In the section mentioned earlier where he suggests reasons why people act irrationally, he says:
Second, because we can hold only a small number of ideas in our minds at any one time, in making complex decisions people do not combine all the relevant factors. One way round this problem is to use pencil and paper to set out the pros and cons and it is irrational not to do so.

In the first sentence he proposes a particular contributor to irrationality. In starting the second sentence as he does he is proposing a particular solution to the problem, but ends by advocating its general utility - that it is rational. The passage continues:

Charles Darwin claims to have used this method....

a reference back to the pencil and paper technique. The clause that refers to irrationality, therefore, stands as a non sequitur, which attempts to embed the general in the particular. In other words, it is a momentary level shift. I find the section disorientating. If the section "it is irrational not to do so" is removed, the rest becomes more homogeneous and easier to read. As I have said earlier, level shifts are paradoxical activities that seem to pervade thought, and the confusion of viewpoint they cause can prove very disconcerting.

Level problems can also be detected in the exact definition of a rational end. An end may be to achieve a particular goal. Equally, an end may be to select a sensible end. Obviously, rationality at one level may not occur at another level. For example, after citing a whole series of examples of how people irrationally assess the risks of using nuclear power, Sutherland says:

The only valid argument against nuclear energy is that its production may contain more unknown risks than does that of fossil fuels.
Not only is there a level shift here, from talking about risks to talking about the risk of risks, but there is also the statement that people are irrational about risks but rational about risks of risks. Sutherland not only ignores the effects that level shift has on rationality, but also ignores the contribution that level shift itself makes to the genesis of irrationality. Many irrational acts could be the product of rational intent which is exited at an irrational level. Likewise, as I discussed at the end of the previous chapter, irrational acts could be the product of inescapable levels of rationality based on the assumption that everyone has perfect information.

I would, accordingly, not dismiss Hollis' axiom 2, but modify it to accommodate the possibility of changing level in order to maintain rationality:

2'. S has a choice among alternative ways of achieving g, or a choice among alternative choices of ways.

THE BOUNDARIES OF RATIONAL BEHAVIOUR

....and thinking, as men do think in times of crisis, that when everything has been done there is still something that needs to be done, when everything has been said there is still something left unsaid....

Thucydides, The Peloponnesian War, VII 69

If we wish to broaden the boundaries of rationality, it still seems necessary to define these boundaries. We need to consider if there are general, objective boundaries, and if there are different degrees of rationality. The issue turns on the nature of belief, and the tolerance or fuzziness which is acceptable in a definition of rationality.
Objective and subjective criteria

Hollis says that an individual's belief in the best way of doing something (axiom 3) does not necessarily make it the best way. The question of viewpoints is, therefore, essential to a definition of rationality. In particular, an agent may internalise an act as rational which to an observer may appear irrational. Should the world view of the agent be seen as the yardstick of rationality, or are there certain objective or universal criteria that can be used? Sutherland makes no comment on this issue. Although his definition of rationality refers to people acting within the environment of their current knowledge, there is no attempt to define what is meant by this knowledge. If a person is deluding himself or distorting his view of the world, then this very distortion organises the world to the benefit of the beholder. If we are going to dismiss this 'knowledge' as incorrect, then we are applying an external or objective measure of what is the correct world view. I am not saying that there should not be some objective views on knowledge and, hence, rationality: there are, in the everyday world, quite obvious and universally accepted acts of irrationality. What I am saying is that Sutherland's definition as it stands does not deal with this issue of subjective as opposed to objective views of reality, and, therefore, he has no right to dismiss distorted views as irrational. The effect that differing views of rationality and reality have can be illustrated in the following way. Suppose I am standing in a supermarket queue. There are several checkouts, but only one is open. After a while, I may decide that it is senseless to wait in the queue, and demand loudly that a second checkout be opened. This would seem rational to me: as rational as standing around in a queue is irrational. To others in the queue, loud demands from someone at the back who has not waited as long as they may seem irrational. In fact, I would be offending social norms, an act which Sutherland considers to be irrational. After all, he argues, we move in and through society, and it is rational to make our way with as little hindrance as possible. Obeying social mores allows us to achieve our ends more easily. To a game theorist, however, social norms would be considered irrational. They are often based on prejudice, irrelevant or received wisdom, and, as such, are no basis for logical decisions.
Demanding service in this ambience would be a rational decision. I do not wish to debate the issue of social norms at present, I wish simply to illustrate the point that the rationality of an act can depend upon the standpoint and beliefs of the observer.

**Belief and evidence**

In order to mitigate a completely subjective rationality, some objective standpoint needs to be defined from which rationality can be judged, without circumscribing too closely the freedom of the individual (the game theoretic approach seems too circumscriptive, for example). Hollis suggest a possible way forward. He says that a person's belief alone cannot rationally be used as a basis for action because:

> Objective rational belief is belief justified by the balance not of evidence actually taken into account but of evidence which should be taken into account.

This seems admirably to define a standpoint more rational than simple belief. It accepts individual belief as the essence of rational action, providing that the individual has taken steps to arrive at that belief adequately. This concept is supported (in a different context) by Harrison in the paper cited above. He asks why Hitler is considered to be an evil man. Generally, Harrison states, Hitler is believed to be evil because of the beliefs he had, about the Jews, for instance. This is desperately simplistic, ignoring the fact that Hitler was evil for what he did. Nevertheless, the argument he develops about beliefs is worth pursuing. Harrison goes on to ask that, if beliefs are not an act of will - I cannot will myself to believe in God, for example - then is it fair to castigate a person for their beliefs? Is it fair to deride Hitler for something that he did not choose? Harrison says, and we would all agree with him, that it is necessary to revile Hitler because:

> ...even if we cannot choose beliefs directly, we can choose to operate processes which will result in new beliefs being acquired.
He was irrational in Hollis' terms because there was evidence available, which should have modified his beliefs and which he ignored. On the contrary, he sought out evidence which supported his beliefs. In Harrison's terms he was evil because he should have operated the process of taking into account more evidence, which would then have modified his beliefs. Neither demands a belief modification, simply an exposure to wider evidence. It is this exposure, the following of the process, the attempt to see issues from more than one viewpoint, which is rational. Hitler was irrational partly because he neither questioned his beliefs nor made an effort to change them.

**Hyper rationality and relaxed rationality**

If rationality varies with viewpoint, does it also vary in degree? Again, Sutherland's definition has nothing to say about this, even though he refers to people being partly rational. For example, in his chapter on Utility, he says:

> but in real life it seems likely that people have a mixture of overall aims - at one and the same time they want to maximise utility, ensure at all costs they do not make a disastrous loss, and also ensure that they at least come out ahead of the game

So although people are not rational enough to use utility theory, they are rational enough to want to survive. Their choice of strategies at a meta-level is, therefore, partly rational. It seems sensible to me to assume degrees of rationality, but this is ignored by Sutherland's definition. Further support for the idea of degrees of rationality can be provided by an illustration from the book *Innumeracy* by John Paulos. He asks for the reader to calculate the following:
When Caesar said "Et tu, Brute", what are the chances that you will inhale with your next breath one of the atoms that he exhaled?

The answer is that it is virtually certain, and that this can be mathematically proved. Such an answer may be statistically true, but it would be unfair to expect anyone other than a statistician to get it. It is so counter-intuitive that I would argue that it is not irrational to think otherwise. Such narrowly special examples of statistical thinking, I would suggest, would best be referred to as hyper-rational thinking. In general, I believe game theoretic rationality is a form of this hyper-rationality.

Interestingly, Sutherland sees no counterpart in people's inability to understand or manipulate language: grammar, syntax and semantic. Either he feels that people are not illiterate in the same way as they are innumerate, or that illiteracy does not lead to irrationality. I cannot say which he advocates since he ignores the issue. It strikes me as interesting, however, that lapses in literacy have opposite effects to lapses in numeracy. Lapses in numeracy tend to lead to irrational conclusions. Lapses in literacy tend to be far more tolerated, and are practical examples of how we compensate for degrees of rationality. For example, if I were to say:

*I don't like fruit but I like grapes*

This would only be seen as irrational by a pedant. More likely, a reasonable listener would fill out the statement to read:

*Generally I don't like fruit but grapes are an exception*

or

*I don't like everyday fruits (apples, oranges, bananas) but I like grapes.*
Although the statement is not fully rational, therefore, it has sufficient reason in it for me to make a sensible guess at its meaning. Even more indicative of this relaxed style of thinking is our approach to a statement such as:

*The first shall be last and the last shall be first.*

Not only does this make sense, but its inherent ambiguity gives it extra sense and meaning. This everyday, tolerant type of rationality seems much more in accord with human nature than the tighter, logicised styles advocated by game theory and Sutherland. It is this sort of rationality that eradicates many of the logical paradoxes we have encountered earlier. With this style of rationality, which entertains varying degrees of rational action and thought, we only need to be sufficiently rational to succeed. This conclusion squares exactly with the conclusion of Chapter 2. There, I suggested that the paradoxical nature of game theory lay in its very desire to be completely logical and to operate on the basis of maximisation. Here, instead, I am arguing for *satisficing* as a strategy: the need to find a solution which is good enough. Although maximisation potentially gains more, the price paid in finding the solution, and its concomitant logical and paradoxical problems, often makes satisficing a better strategy, especially where we feel, as Thucydides did in the quote which starts this section, that no solution will ever be good enough in practice. Game theorists would argue that satisficing is simply a form of maximisation. I would argue, conversely, that maximisation by satisficing is, in game theoretical terms, being rational by choosing to be irrational, which would again seem to be a case of game theoretical tenets leading to logical contradiction. By its own lights, it cannot have its cake and eat it in this way.

In order to remove the dependence of his definition on personal belief, Hollis proposes the following modification to axiom 3:

\[ 3'. \text{ A is likeliest to realise } g. \]
This also seems to relax the rigours of extreme rationality, allowing us to accept that a solution does not need to be tightly rational, but simply rational enough. Its weakness, as it stands, however, lies in its use of the word 'likeliest'. If 'belief' is too dependant on subjective viewpoints, then 'likeliest' seems based on nothing at all. To remedy this weakness, and to include as a basis for rational action the obligation to obtain a range of views, I would modify Hollis' axiom to:

3°. Given the evidence, a is likeliest to realise g

WILLING WHAT CANNOT BE WILLED

In a nutshell, Plato, we must always distinguish the incidental benefits of an activity from its inner purpose. Indeed, sometimes the benefits only ensue because we do not intend them.

Roger Scruton, Xanthippic Dialogues

Hollis next argues that the axioms do too little because of the use of the word 'wants' in axiom 1:

1. S wants to achieve g

He argues that a person's wants are really no basis for the discernment of rationality. He uses two arguments to support this view, both of which I have touched upon already. Each underlines the conflict between 'wants' and 'best interests'. First, he argues from the time discounted point of view, that a person's immediate gratification may not best serve her in the long term. Second, he considers situations where wanting as self interest is an inferior strategy if used by everyone in a group. This is not the definition of wanting - self interest - that I wish to pursue here, so I will postpone discussion of this issue until the section the problems of aggregation and group decision-making below. I would add
a third objection to the use of wanting as a basis for rationality. As I have already suggested, many desirable results may be a by-product of activity rather than a direct act of will.

**Pre-commitment and self-binding**

Any definition of rationality must acknowledge the effects that time can have on rational decisions. Not only do people's decisions vary as they are viewed in the short or long term, but an agent taking a decision in the present should presumably take into account her likely predisposition to continue her support of this decision in the future. As we have seen, however, the decision and strategies required to pre-commit our future selves are difficult to arrive at. In the discussion about Ulysses and the Sirens, we saw that committing one's future self to a course of action is not a matter of will. We need, in essence, to treat our future selves as separate entities, as other beings that we need to take into account. In dealing with our future actions, therefore, we are trying to will what cannot be willed, since we have no complete and wilful control over these different selves. I will elaborate on this statement by recalling the graph I used in the previous chapter.

![Graph showing the trade-off between time and benefit](image)

It will be remembered that this graph summarises the situation where a long term gain (b-b) is preferable to immediate gratification (a-a). As we can see, neither Sutherland's
use of ends in his definition or Hollis' wants adequately deals with this problem: we can achieve our ends, or satisfy our wants, by the inferior choice a-a, if our ends or wants are defined in the short term. Some form of pre-commitment, as outlined in Chapter 2 seems essential to achieve a more responsible outcome. In fact Schelling, in the work cited there, calls for a theory of self-management, which he calls command theory, as well as a theory of decision-making, decision theory. Central to his command theory is the need for pre-commitment. A similar need is advocated by Jon Elster in his book, Ulysses and the Sirens. In spite of their strong support of pre-commitment, it is interesting that neither see it as a complex issue in its own right. In particular, they do not see it as an issue that involves two distinct activities: the restriction of the actions of the future self (which I shall continue to refer to as pre-commitment), and the restriction of the actions of the present self (which I shall refer to as self-binding).

Pre-commitment in this narrower sense, that of the restriction of the future self, is generally made by the present self in its 'sane' moments knowing that the future self will not be so sane. For example, the present self, knowing that the future self has a great deal of difficulty in getting out of bed, will place the alarm clock across the room from the bed before retiring. The difficulty with pre-commitment is that, since it is often undertaken when the decision is not irksome, it is impossible to judge its severity. Having eaten several chocolates, I might decide to put them in the kitchen to prevent later indulgence. When I want another chocolate and go and fetch them, I might have wished that earlier I had thrown them away instead. Alternately, if I cut my hand off to prevent my future self from biting his nails, then I might conclude that this pre-commitment was too severe.

Self-binding operates differently. Here, my present self chooses to bind itself so that the future self will benefit. This is exactly what Ulysses does, and requires some personal present self-denial. Unlike pre-commitment, the present self, not the future self suffers. The difficulty with this act is that a certain wilfulness, the need to put oneself to trouble,
is required in order to make this strategy work. And for what? The future is fluid, and none of us are fortune tellers. Precautions or obligations taken now may either be useless or unnecessary later. It may make more sense to refuse to undertake a present burden and to trust in the future self to manage: to let the future look after itself.

Underpinning both pre-commitment and self-binding is the possible necessity for a third act. Both activities benefit if a third party can be introduced to support the activity. For example, Ulysses restricts his present self by getting his crew to bind him to the mast and agree to ignore his future commands. Likewise, he could have coerced his future self by lodging all his possessions, and deeds to his property, with his crew, on the understanding that it was theirs if he left the ship. The lure of the Sirens might then have been counterbalanced by his love of wealth, Ithaca and Penelope. The difficulty with introducing third parties is that they can be left with the decision of which self to support, especially if there is a need to inflict pain. If someone gives his cigarettes to a friend, and begs her to refuse to return them, what happens when the future self demands a cigarette? Which self does the friend believe? Does the person really want to give up smoking or not? In denying the demand for a cigarette, is the friend actually supporting a false self who does not really speak for the true desires of the person? And if the person needs a cigarette, is literally dying for a cigarette, has the friend the right to inflict such pain on any of the selves, be they true or false? It is interesting that the law gives little licence to individuals to help other individuals, perhaps reflecting the accepted existence of these various selves and their contradictory demands.

Although pre-commitment, self-binding and the need for the support of others all seem rational, they are beset with difficulties, and are not easy to achieve successfully. Jon Elster, in his book, *Nuts and Bolts for the Social Sciences*, sees the principal cause of this failure as weakness of will. Although we know what is best for us, and we want what is best for us, we actually refuse the take the burden, or give in to the desire for instant gratification. I would not argue with this suggestion, since this is a commonly
observed phenomenon. I have said throughout this dissertation, however, that weakness of will cannot entirely explain why things go wrong for us. I believe that, in this case, as in many others, there are difficulties inherent in the situation. I showed in a general way, in the previous chapter, that attempts at pre-commitment can be paradoxical. I shall illustrate this issue more fully now by considering, as an example of it, management groups and the problems they might have with pre-commitment. This issue is particularly applicable to management groups which, I would contend, have an essential commitment to self-binding as a strategy. I will illustrate this by the use of self-referential paradoxes, which are excellent examples of statements which bind-themselves: therein lies their paradoxical nature. The aspects of management which, it seems to me, particularly require self-binding as a strategy are:

- the devolution of power;
- the giving of orders;
- the making of rules.

I will try to frame each of these statements in the form of a self-referential paradox.

\[ I \text{ do not exist} \]

This illustrates the difficulty of negating one's power as a basis of its devolution. It is paradoxical in the following way. If the statement is true, then who made it? Truth forces the existence of a being, which is attempting to negate itself but cannot. If the statement is false, then the being exists anyway. In either case, the statement is incapable of de facto truth. The question the paradox raises is the question of one's ability to will what cannot be willed. Non-existence cannot be willed. This can be illustrated by analogy. Supposing you are lying in a hammock, gently rocking yourself. The motion is gradually sending you off to sleep. At some point you become so drowsy that you lose the ability to rock yourself. The movement, which was sending you to sleep, stops, and you wake
up. Likewise, non-existence would presumably end as the will to non-exist vanishes with the being willing it. The whole question of self-negation and self-binding is essential to a consideration of whether power can be truly devolved.

*Don't be obedient*

This statement was mentioned earlier, and is paradoxical in that if it is obeyed it isn't, if it isn't obeyed it is. The self reference lies in framing in the imperative an order to ignore the imperative. It brings into question the nature of command, particularly whether one is able to command oneself. Again the question of self-binding, and the ability to will what cannot be willed, is relevant here. Ulysses gave exactly this command to disobey when he told his men to bind him to the mast and not release him no matter what he said.

*Only rules produced from now on will be obeyed*

The paradox here lies in two areas. First, there is the question of the status of this rule. Presumably, it was produced before now, if we take production to mean the working out and formulation of rules before they are promulgated. Even if we take production as the actual writing, then presumably it existed before its sense became apparent. This also applies to its reading. In any real sense, the rule existed before the statement it embodied existed. Therefore, in its own terms, it need not be obeyed. Once its terms are negated, however, its stricture becomes apparent again, so it could be obeyed. Once obeyed, it again ceases to have force. A second area of paradox is to do with levels of rules. This rule is about rules, which is essentially different to rules about issues. The rule about rules implies an even higher level of rules. Overarching all rules is a universal rule which says "all rules must be obeyed." By definition, a rule is a statement which demands obedience. But in terms of our rule about obeying future rules, does this meta-rule have to be obeyed? It was certainly produced before now, since it is implied in all rules. So the universal rule about all rules does not need to be obeyed. So our present rule does not
need to be obeyed. Once the present rule is not obeyed, the universal rule comes into power again, forcing the obedience of our present rule. Individual rules, therefore, may not be paradoxical, but a system of rules, by self-reference, might be. Again, associations between this and previous paradoxes are obvious, in that they refer to concepts of self-binding and the nature of will. This paradoxical statement introduces the further problem of decisional levels discussed previously, and the confusion that can occur if the difference between decisions about decisions and decisions about issues is not adequately understood.

The practical and actual outturn of the problems for management outlined here I will consider fully in the next chapter.

By-products

Several ends are achieved as a by-product of action, not by the actions themselves. These ends cannot, therefore, be willed. Let me first be quite specific about what I mean by a by-product. Some by-products are unplanned, spasmodic or notional. For example, when a team loses an entertaining and well-played game, and their manager is interviewed, his response: "No matter about the score - football was the winner today" describes an unplanned by-product. After all, none of the players were playing for an abstract concept like football. They were playing for revenge, machismo, win bonuses, or a hundred and one other ends. Nevertheless the outcome is tangible and the response sensible. Other by-products can be planned for. Let me illustrate this idea, and illuminate the issue of by-products further, by a reference to Benjamin Franklin's third 'self-evident' right: the right to pursue happiness.

I would suggest that it is not possible to pursue happiness directly. Words like pursue and want imply an active move towards the direct attainment of a goal. We might want a car, so we go and buy it. Is it possible to pursue happiness in the same way? It is certainly possible to pursue things that make us happy, or things we enjoy, but this is not
pursuing happiness directly, since it does not exist as a separate entity: it is a by-product of some other more tangible activity. But neither is it unplanned or haphazard: we generally know the sort of things that make us happy. So we obtain the by-product happiness by consciously pursuing a thing we can have directly, like a car. In this active sense of wanting or pursuing, therefore, the pursuit of a by-product does become paradoxical: if we try to get it directly, we can't - if we don't try to get it we do. I don't think I am saying anything particularly profound about concepts such as happiness. I am saying simply that they are enabled through a tangible third agent. Nevertheless, their elusive nature should be borne in mind, as later I will show that the desire to consciously and directly enable such items (instead of through a third agent) bedevils educational thinking and planning.

By-products are likewise paradoxical by level shift. Suppose we are enjoying something. Is it possible to stand above or outside this enjoyment, to apprehend or formalise it instead of experiencing it? I would argue that it is not - that when we say: 'I am enjoying this' that for that time at least we are not enjoying it. We are instead making statements about enjoyment which are not enjoyment in themselves. Enjoyment or happiness is, by its nature, an informal or elusive activity, which disappears as it is defined. We operate 'within' happiness, not 'about' it. In his book *The Poetry of Architecture*, Ruskin develops the paradox, by arguing that, in some situations, the 'within-ness' only becomes evident when 'about-ness' is present. He says (paragraphs 47-8)

> unless the cricket is chirping on the lonely hearth, or the vulture soaring over the field of corpses, or the one mourner lamenting over the red ruins of the devastated village, that devastation is not felt to be complete.....And if desolation, which is the destruction of life, cannot leave its impression perfect without some interruption, much less can solitude.....one of the chief uses of the mountain cottage, paradoxical as the idea may appear, is to increase this sense of solitude.
Again, I don't think this is a particularly profound paradox, but it is nevertheless an issue that tends to be ignored during practical planning. Too often, I will argue later, do we try to formalise the essentially informal.

These kind of by-products which are realisable but uncontrollable and indefinable were called *collateral* outcomes by John Dewey, the American philosopher and educationalist. He says, for example (in *Experience and Education*), that as well as a student learning his normal lessons in school:

> The most important attitude that can be formed is that of the desire to go on learning....to extract meaning from his future experiences as they occur. (p 48-9)

Outcomes such as this can be planned for, in that we can create the kinds of third agents within the curriculum to evoke the by-product of 'the desire to go on learning.' This is not, therefore, an haphazard or unplanned outcome. Neither will the outcome appear to order. It will appear at different times for different people and for some not at all. Outcomes like this are, as I said at the beginning of this paragraph, realisable but uncontrollable.

It remains to consider whether the generation of by-products is rational, and thence to further refine our definition of rationality. Sutherland sees by-product or non-willed achievement as irrational. In the section on reasons for irrationality he says:

> Both wishful thinking and self-deception may contribute to a person's happiness, and to this extent they are rational means to an end. But I have defined irrationality as coming to conclusions that cannot be justified by one's current knowledge and in so far as someone distorts his view of the world or of himself he is thinking irrationally.
His definition, however, does not state this. Strictly, by his definition, if someone undertakes a particular activity, and as a result of this they also achieve happiness, then it is a rational end to undertake the activity entirely to achieve the by-product. Sutherland tries to get round this objection later by saying that a by-product - something which cannot be directly willed - can become a direct act of will and hence a rational end in itself, if the by-product has been achieved in the hope that it will be internalised. I will illustrate the point he makes with reference to Pascal's Wager. Pascal argues that the choice of believing or not in God can be restated in the form of a bet. If you bet on the existence of God, you are staking some of your time - a finite amount - given over to devotional activity. In return, if God exists, you gain an infinite amount. If he does not, you lose the time given to devotions. Therefore, you risk a finite loss for an infinite gain. This is an excellent bet and should be taken. We are driven to ask, however, why more people do not take it. It is not a question of ignorance, since most people could work out the nature of the bet. The answer surely lies in the nature of the stake, which itself depends upon how the chances of success at the bet are viewed. An atheist will see no point in a commitment to devotions which in his terms will not bear fruition. A believer does not need to think in terms of the wager anyway. Unlike an ordinary bet that requires an act of will, a bet on Pascal's Wager requires an act of faith which, by definition, cannot be willed: if you believe you wager anyway, if you don't you don't. This is the exact point made by Robert Martin in the quotation which opens this chapter. Pascal, himself, believed that attention to the act of worship might inculcate the belief. Sutherland supports this view, citing Aristotle's belief that if a person acts outwardly in a particular way, then he will eventually internalise the actions. It is a behaviourist argument, and may have some validity in the acquisition of skills and the eradication of bad habits. I feel, however, that the attempt to achieve belief in God by such a regime is irrational. Pascal's Wager, a game theoretic ploy, fails because it is not possible to will what cannot be willed. In the same way Sutherland fails adequately to deal with the question of willed and unwilled ends.
Conclusion

As we have seen in the first part of this section, wanting and wilfulness can cause paradoxical situations. Further, the generation of by-products, though desirable, seems not to be an act of will. In *The Cunning of Reason*, Hollis describes how Adam, in the garden of Eden, chooses to pick figs to eat. This choice is based as much on his belief that there are fig trees around as on his desire to eat them. If either the belief or desire fail, then the figs remain unpicked. Says Hollis:

Why then give motivational priority to the desire?

Further, belief implies some commitment to what is actually the case - and this exactly echoes the epigraph to this chapter and the need to base action upon some concept of objective truth. It seems inevitable, therefore, that 'wants' must be excluded as a basis of rational thought and action. In order to accommodate this weakness in axiom 1, Hollis proposes the removal of the word 'wants' and the following rewording:

1'. It is in S's overall interest to realise g.

This rewording is endorsed (not with reference to Hollis) by Rescher in *Rationality*. He refers to wanting as 'the "technical rationality" of goal-efficient action', which reduces rationality to selfishness - the sort of selfishness implied by Sutherland's ends-oriented definition. He further defines best interests as something a bystander would assume I ought to want, given my circumstances. Statement 1', thus qualified, would appear to remove the difficulties described above. First, the use of overall interests rather than wants suggests that, where necessary, long term as well as short term goals should be taken into consideration. Second, overall interests would appear to cover the by-products of actions as well as their direct outcomes. In general, a recourse to overall
interests implies that there are objective standpoints available to all, that specify an individual's best course better than his or her wants.

Before we move on to the next section, I would like to show how we have modified Hollis' and, indirectly, Sutherland's axioms so far. Hollis originally said that the following were sufficient for a rational thought or action:

1. $S$ wants to achieve $g$
2. $S$ has a choice among alternative ways of achieving $g$
3. $S$ believes that $a$ is the best way to achieve $g$

In the above discussion, we have modified these to:

1'. It is in $S$'s overall interest to realise $g$.
2'. $S$ has a choice among alternative ways of achieving $g$, or a choice among alternative choices of ways.
3''. Given the evidence, $a$ is likeliest to realise $g$.

In other words, it is rational to achieve your best interests by selecting from among a number of actions or thoughts the optimal one, based on what you know. The discussion above on levels and meta-levels has already indicated that this definition can encounter problems, especially when the level of decision is confused. However, this need not be a problem if careful attention is paid to the issue, and the level of decision accurately defined. Other than that, the definition has obvious merit. It appears to provide the broader view of rationality I advocated earlier. It allows by-products and temporal effects to become part of rational activity. It indicates that irrationality on one level may be part of a greater rationality. Finally it concedes that although there are objective criteria for rationality, these can to an extent be mitigated by subjective viewpoints – in most circumstances it is sufficient to be rational enough. In the discussion that remains, I
shall assume the definition as a basis for rational action. I hope to show that the last two items to be discussed, means and ends and the problems of aggregation and group decision-making, will not square with these rules.

MEANS AND ENDS

The difference between the reason of man and the instinct of the beast is this, that the beast does but know, but the man knows that he knows.

John Donne, Sermons

Non-rationality

One of the weaknesses of Sutherland's view of rationality is that it is ends oriented. The amendments we have made to Sutherland via Hollis in the sections above have not remedied this. I believe that an ends oriented view of rationality is impossible to support because it allows non-rational thoughts and actions to be included. I am careful here to use the word 'non-rational', not 'irrational'. We can obtain a view of irrationality by negating our definition of rationality. So we could say, for instance, that an irrational act is one that is not in our best interests. Non-rationality is, by its very nature, indefinable within the terms of any definition. It occupies the space in the present definition of rationality which is occupied by the sentence:

This sentence is true

in the two state logic system described in chapter 1. That is to say, non-rational items are undecidable in our present system. They inhabit the area to which our present definition can attribute neither rationality or irrationality. All formal systems have such gaps, but in this case the gaps are extensive and important. Let us, for example, apply our definition to animals, who seem by it to be rational. After all, they act in their best interests, choose
between alternatives, and achieve their ends with monotonous regularity. Even given man's propensity for anthropomorphism, however, it is unlikely that we would think of animals as being rational. Neither would we think of them as being irrational. More than likely, we would say that rationality as a concept is not applicable to animals, rather as loudness is not applicable to colour. They are, in this sense, non-rational. Consider, for example, Hardy's poem, *The Last Chrysanthemum* (last three stanzas)

Too late its beauty, lonely thing,
   The season's shine is spent,
Nothing remains for it but shivering
   In tempests turbulent.

Had it reason for delay,
   Dreaming in witlessness
That for a bloom delicately gay
   Winter should stay its stress?

- I talk as if the thing were born
   With sense to work its mind;
Yet it is but one mask of many worn
   By the great face behind.

Hardy divines a teleological, rational drive to the plant. Certainly it is common to talk even of plants, as Hardy does, as if they have sense. Although the concept of plant rationality is absurd, I don't know whether our definition does not, in some ways, keep it in court.
The intent to be rational

The use of the thought processes of animals as an example of non-rationality was deliberate. Comparison with animals suggests, as I proposed at the beginning of this chapter, that rationality is something essentially human, and requires for its action and apprehension, therefore, the full array of complex physiological, psychological and social attributes that man possesses. In other words, the definition as it stands is too behaviourist: it depends too much on input and output, with not enough attention to the processes that take place within the mind. This is the problem when rationality is simply judged by outcome alone: as if by an onlooker deciding our best interests. Instead we could state that the intent to pursue rational action is as valid a measure of rationality as the outcome of the action. Rationality involves, therefore, at least a consideration of the processes that go on in the mind while decisions are being made. This requirement is eloquently urged and explained by the quotation from Donne which opens this section. A commitment to investigate processes of the mind, however, includes a commitment to accept as important to the question of rationality whole swathes of stimuli which have hitherto been ignored. Simple self-interest can no longer be assumed as a basis for rational thought. Instead, we must find a place for 'trust', 'altruism', 'hate' and so on. The difficulty with the question of whether it is rational to base decisions on these stimuli is that it evokes the answer: 'it depends'. I would contend that trust, altruism, hate and self-interest, come to that, are rationally neutral. Actions based on such concepts may be rational or irrational according to the specific scenarios in which they occur. For example, at times it is rational to tell the truth, at other times it is not. The matter becomes worse when we consider the status of the stimulus 'rationality'. There is no doubt that this is a stimulus to action. People can say and mean:

Let us be rational about this

as easily as they can say and mean
Let us trust each other on this

The first statement, however, I would term a 'meta-stimulus', in that it makes a commitment to apply only certain other stimuli, such as truth-telling, which may be perceived as rational, and to exclude others, such as hate, which may not be. Such context dependent concepts should, therefore, could also be referred to as 'non-rational'. Further, and crucially, these context-dependant, non-rational stimuli become rational or irrational within the terms and environment of their actual application. It is this specific treatment that dictates their nature. Until then, they live in this limbo of undecided rationality. And the thing that gives them their distinct rational or irrational flavour is the desire on the part of the individual to be rational, and the perception of others that he is trying to be. In this context, hate, trust or altruism can become rational, and the intent to be rational the sole arbiter of rationality. If we adopt this means centred view, the ends centred definitions we have used so far become meaningless. They are simply too general and objective to be squared with the subjective, context-specific rationality associated with the intent to be rational. In other words, having gone to a great deal of trouble in the previous section to show the necessity of removing wants as a basis of rationality, we have shown equally compelling reasons for restoring them again. This conflict between 'wants' and 'best interests', therefore, seems hard to resolve. Best interests provides a more objective, less fickle basis for rationality, but cannot adequately deal with non-rational, non-human decision-making. Wants provides the necessary human dimension, but cannot deal with temporal effects and the by-products of action. The next section shows that, even worse, both can be inappropriate because, in certain circumstances, it becomes impossible for an individual to decide either what they want, or what is in their best interests.
THE PROBLEMS OF AGGREGATION AND GROUP DECISION-MAKING

It is not to be understood that the masses ever recognised their own handiwork in the Inquisition

W G Sumner, *Folkways*

I suggested in a previous chapter that an individual's actions can be constrained by membership of a group, and that this constraint is partly physical and partly logical. This constraint can eventually lead to paradox. Before describing these group decision-making paradoxes, however, let me first define what I mean by a group decision. I do not believe that there is any such thing as a group mind or group intelligence. I see groups primarily as collections of individuals. These individuals are motivated by any number of characteristics: altruism, selfishness, greed, envy, utility. They bring these bags of emotions and beliefs to the decision-making forum, and these essentially inform the group decision. The group decision is further moderated by the style of decision-making employed by individuals - whether it is rational, irrational or a combination of both. Decisions may also be about contribution and distribution. Contribution decisions delimit the nature and extent of individual contributions to the group. Distribution decisions lay out the formulas and mores which describe the division of goods which accrue to the group. These decision types are not exclusive. The type and extent of contribution affects the amount of goods produced. Likewise, the relative attractiveness of goods which the group can produce affects the extent and fervour of contribution.

This whole ensemble of decisional parameters result in a series of individual decisions. As I said earlier, and hope to show in more detail shortly, these individual decisions are themselves warped by the act of aggregating them into a group decision. Although, the group does not have a mind, therefore, it is equally impossible to equate its decisions to an ensemble of individual decisions. The variety of individual decisions and the act of aggregation produce not a group mind, but a group outcome or action which is largely
Paradoxes in Decision-making

Paradox Lost and Paradox Regained

unpredictable. It is this unpredictability which gives the group a semblance of being a living entity in its own right. An alternative way to capture the essence of aggregation, and in the spirit of this chapter, is to see it as a by-product of individual action. Although individuals know what they want, they are either powerless in the face of, or as often unaware of, the effects of these desires as they act in concert. This is the situation which is nicely summed up by the quotation at the start of this section.

Not only the act of aggregation but the mechanics of aggregation can also generate paradox. There are several opinions on how individuals interrelate in order to produce a group stance. Extreme and opposite views are presented by, for example, Robert Axelrod and Jon Elster. In his book *The Evolution of Co-operation*, Axelrod presupposes a series of individual exchanges between individual organisms (people, animals.) Each exchange produces an outcome. Then both parties to the exchange move elsewhere to relate to other individuals. With time discounting, later interactions between individuals are affected to an extent by their earlier ones. Axelrod also proposes and describes individual characteristics, or strategies. Some organisms never co-operate, some always co-operate. Some play tit-for-tat: that is, they start by co-operating, but then copy in their next meeting whatever a particular opponent did to them in their previous meeting. The aggregation of these individual exchanges gradually leads to a group ethos, based on the commoner sorts of strategies. So where individuals tend to co-operate, a climate of co-operation is generated by the group. Individuals do not think of their actions as part of a group effort: they may not think of themselves as being part of a group at all. They follow individual, evolutionary stable strategies, which work for them as individuals. Group benefits may accrue and help the individual, but the benefits only support the success of essentially selfish strategies. Elster, on the other hand, proposes a much more holistic view of the development of group mores and decisions. Like Axelrod, he sees the individual as the fundamental decision-maker. However, decisions are not the outcome of individual actions and interchanges, but are more the result of the individual's perception of how other individuals will act. Elster views the individual as
introspective rather than active. His individuals base their introspection on a view of other individuals as being homogeneous in motivation: of 'people' acting in a certain way. They can conceive of group outcomes or benefits based on these actions. Axelrod refers to very particular social interchanges: grooming or fighting. Elster refers to more abstract group activity: for instance, how a group of teachers might ensure that the school dress code is followed. Of the two, Elster produces scenarios which are paradoxical, where individuals make decisions based upon the perceived actions and beliefs of other people, knowing that they are one of the 'other people' to these other people. Axelrod, conversely, attempts to dispel paradox. I feel the actual truth about social interaction lies somewhere between the two. Individuals may develop a general view about how others will think and act, but modify this view by the experience of actual exchanges, or in relationships with particular or favoured individuals. I have tried to model some of these interactions and I will describe the results in the next chapter.

Finally, the group action or identity itself, by self reference, affects the actions of the individuals that make up the group. Group norms and decisions affect the later decisions that individuals make. These later decisions reform the group ethos, and so on. The whole activity of decision-making in groups, therefore, has this contradictory, slippery nature which is the basis of paradox.
The individual and social elements are always in interplay with each other if there are a number present. If one is trying to carry on a struggle for existence with nature, the fact that others are doing the same in the same environment is an essential condition for him.

W G Sumner, *Folkways*

The first scenario I wish to consider is the Prisoner's Dilemma. This paradox is a paradigmatic case of the group decision being at variance with, indeed undermining, the individual rationale. The resultant group ethos may present the individuals in a poor light even though, individually, they are doing their best. The dilemma, generally attributed to A.W. Tucker of the Rand Corporation, is named after two fictional prisoners. They are presently held for a minor crime, but are suspected of a major one. They are separately offered a deal, which they are not allowed to discuss with each other. If one is prepared to confess for the major crime, and the other doesn't, then the confession will help convict the other who will be severely punished. In return for his help, the first prisoner will be freed. If both confess to the major crime, then they will be punished but with some leniency because they confessed. If neither confesses, they will simply serve the sentence for the minor crime. The payoff Prisoner I receives in relation to Prisoner II's strategies can be expressed as follows:

<table>
<thead>
<tr>
<th>Prisoner I</th>
<th>Prisoner II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not confess</td>
<td>-2 Years</td>
</tr>
<tr>
<td>Confess</td>
<td>0 Years</td>
</tr>
</tbody>
</table>

This matrix shows that if I doesn't confess and II does, then I gets 5 years. If they both don't confess they get 2 years. If they both confess, they get 4 years. If I confesses and II
doesn't, I serves no sentence. The figures are given in negative amounts to represent the amount of liberty that Prisoner I loses. Conventionally, however, games in the strategic form tend to be represented in a positive way; the amount a player gains rather than loses. This can be achieved quite easily by adding a positive number to each payoff. The result for Prisoner I's payoff is:

<table>
<thead>
<tr>
<th></th>
<th>Not confess</th>
<th>Confess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Confess</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Confess</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Here 5 has been added to each payoff and the now meaningless "years" dispensed with. If liberty is of utmost importance to the prisoners, then the numbers are a weighted value of how much the Prisoner I prefers each outcome. In other words, she would prefer the scenario where she confesses almost twice as much as the one where neither confesses. This matrix can be said, therefore, to express utility values.

Further additions can be made to the matrix. First, it only shows payoffs for Prisoner I. Again by convention, Prisoner II's payoffs are shown to the right of Prisoner I's.

<table>
<thead>
<tr>
<th></th>
<th>Not confess</th>
<th>Confess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Confess</td>
<td>3,3</td>
<td>0,5</td>
</tr>
<tr>
<td>Confess</td>
<td>5,0</td>
<td>1,1</td>
</tr>
</tbody>
</table>

A second variation is to represent preferences simply in their order and not their strength. For both prisoners, the preferred order of strategy combinations, from greatest to least is:
Their place in the order is signalled by their attachment to the numbers 1 to 4; as with utilities, a higher number indicates a higher preference. These numbers are shown next to the strategy combinations.

<table>
<thead>
<tr>
<th></th>
<th>Prisoner I</th>
<th>Prisoner II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not confess</td>
<td>Confess</td>
</tr>
<tr>
<td>Not Confess</td>
<td>3,3</td>
<td>1,4</td>
</tr>
<tr>
<td>Confess</td>
<td>4,1</td>
<td>2,2</td>
</tr>
</tbody>
</table>

Both styles of representation, the utility and the ordinal, have their uses. Utility matrices are used to model actual scenarios, and can be manipulated to investigate the thresholds of behaviour. Ordinal matrices can be used to model the general preference structures, and so make comparison with other combinations of preferences possible. In particular, the use of ordinal matrices makes the generation of new models a relatively easy job. I will investigate this application in Chapter 4, where the use of such matrices in the classroom is described. The type of representation to be used, therefore, varies according to the nature of the activity under consideration.

I would like to start by analysing the structure of the Prisoner's Dilemma, and so will refer to the ordinal matrix. First, it can be easily seen why this type of game is called a non-zero-sum game. Games, commonly, are taken to mean those of the parlour variety: Bridge, Chess, Monopoly and so on. In this type of game, whatever one player wins the other loses. The total of losses and gains is, therefore, zero. These games are called zero-sum games, and are games of pure conflict. In a game such as Prisoners' Dilemma, both
players can win. For example, the strategy combination Don't confess - Don't confess allows both to benefit. The total of their gains and losses, therefore, is not zero. As we will see, the games still involve a measure of conflict, but the ability for both players to benefit enables an element of co-operation. A second aspect of the game is that it consists of two choices for each player: these combine to make four strategy combinations. These were shown above, in their order of preference. Since we are analysing the structure of the game, it may be appropriate to generalise these strategy combinations, since they reoccur continuously in social situations. I have listed the combinations again, with a name not specific to the scenario.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I confess, other doesn't</td>
<td>Temptation</td>
</tr>
<tr>
<td>Neither confesses</td>
<td>Reward</td>
</tr>
<tr>
<td>Both confess</td>
<td>Punishment</td>
</tr>
<tr>
<td>I don't confess, other does</td>
<td>Sucker</td>
</tr>
</tbody>
</table>

As I describe the nature of the dilemma, the reason for the choice of these names will become apparent. The problem for the prisoners, or anyone involved in the generalised game, is that of how to make a sensible choice. By the dominance principle, the best choice is to confess (referred to as defection in the general game.) In the ordinal game it ensures at least 2, and possibly 4 points. This is a better payoff than the equivalent 1 and 3 if non-confession (co-operation in the general game) is selected. The choice is rational in that a move to another choice would make the player worse off. Unfortunately, and here is the root of the dilemma, both players can deduce that defection is the rational and obvious choice. Since both will defect, both will score 2 points. If both were to choose to co-operate, then both would score 3 points. As was mentioned earlier, however, this is not an optimal choice. If both players choose it, they both prosper. But if one player chooses it, and the other is tempted to defect, the defector gains 4 points and the co-operator, who now looks like a sucker, gains 1. The dilemma lies in which choice to make. The choice of defection is safe, but defection by both players condemns both to
poor payoffs. Co-operation by both improves payoffs, but a co-operating player is always vulnerable to defection by the other. Finally, defection is always possible because its payoff is greatest when the other co-operates. The dilemma is inescapable because temptation to defect outweighs the reward of mutual co-operation, and the safe play of defection is punished because it fares worse than mutual co-operation. The structure of the game can be represented, therefore, in the following way:

Temptation > (is greater than) Reward > Punishment > Sucker

As will be seen later, other structures can create equally formidable dilemmas.

A final question I wish to consider is whether the dilemma is paradoxical within the terms I have defined. This can be best answered by considering the game as a repeated, or iterated game. It is important to understand that this leads to a different scenario from the original. The original game involves one turn, where both prisoners make a choice. It is unlikely that either will ever have to make that choice again. As has been shown above, however, many real world analogies of the game suggest that it is played repeatedly by individuals against other individuals or between groups of players (the multi-person game.) It is possible to argue that, in the single-turn game, defection is the most sensible strategy: co-operation could prove to be too costly. In the iterated game, however, the choice of defection leads to a group outcome which is far from satisfactory. I will give actual examples of play in the next chapter, but the accumulation of poor scores from defection becomes an obvious poor result when compared with the results of co-operation. Further, the gains from defection are high when most of the players co-operate, but are drastically reduced when everyone jumps on the band wagon of defection. The upshot is that, in the iterated game, it is not obvious what is the rational choice: what, in fact, is in the individual's best interest. The problem is compounded even before the game starts. H H Kelley and A J Stahelski, in a series of papers written in 1970, show that players bring to the game (and presumably to life generally) a
predisposition to co-operate or defect based on their view of what others will do. More particularly, co-operators will tend to be optimistic in their summary of others, assuming that some might at least co-operate. Defectors, however, seem to judge all others as defectors. This state of affairs might well be explained by the out turn of Prisoner's Dilemma-type scenarios that the players have experienced before. If a player is, by predisposition a co-operator, then they will meet both co-operators and defectors, since either style does well against co-operation. A person with a pre-disposition to defect, however, will only ever meet defection, since co-operation fails against defection and will soon be abandoned. This rather bleak conclusion would suggest, therefore, that any Prisoner's Dilemma game will begin with an element of defection built in.

As the game proceeds, however, the deleterious nature of this choice both on individual and group success becomes a cause for concern. Yet every attempt to establish a co-operative stance is ruined by individual defections which capitalise on the co-operation. The temptation, in other words, overrules the good intent, but the temptation leads to a poor payoff. Groups, and the individuals within them, cycle through the choices, never able to find one that gives a satisfactory result. The whole process reveals a very complex group dynamic, based on the relative strengths of decision-making styles, which is much too complicated to be analysed by individual players. The complexities make a consensus point impossible to find. Not only is it a rivetingly interesting process to watch, but the swirling nature of this search for the impossible is, I would argue, the very stuff of paradox as defined earlier.

The Prisoners' Dilemma has attracted an inordinate amount of literature. Some attempt to apply it to real issues. Some attempt to find a solution to it: to find a satisfactory choice. Most of the literature attempts both to relate it to the real world and to find a solution. As time has passed, attempts to find mathematical or rational solutions have declined, and been replaced by investigations of real world solutions which can be mapped back on to the game. The first authors to systematically study the Prisoner's
Dilemma were R D Luce and H Raiffa in their seminal work *Games and Decisions*. Published in 1957, it set the tone for research into game theory for the next fifteen years. Their stance was straightforward and logical. First, they believed that scenarios such as the Prisoner's Dilemma encapsulated in simple form central truths about the nature of society. Second, they could apply to these models powerful mathematical techniques for finding the optimal strategies within them. Finally, this implied that several hitherto intractable problems of social interaction could be reduced to mathematical models and solved. The outburst of research which followed in the sixties, and which found expression in several hundred papers in journals such as *the Journal of Conflict Resolution*, was aimed at what was considered to be the simple task of showing how this mathematical theory of decision-making actually squared with decision-making in practice. Even when it began to become apparent that people did not solve the Prisoner's Dilemma in the ways prescribed by theory - in fact, did not solve it all and continued not to co-operate - the belief that people actually did decide optimally was clung to. As late as 1970, writing in his book *Game Theory*, Morton Davis still finds excuses for this now obvious lack of co-operation within the game. He says that some players admitted to getting bored by the it, and started to play frivolously (boredom was an oft quoted reason for the 'incorrect' experimental result.) This was compounded by inconsequential payoffs, which gave the players no incentive to play 'properly'. He concludes with the statement:

> Our point is not that the experimenters erred, but rather that the results must be analyzed cautiously.

Although scepticism grew about the game, it was not abandoned. Instead, it has become part of a package of several games, all non-zero-sum games, which have been used for a wide variety of purposes. First, the games have been used to see if players do relate model situations to real situations. For example, in Colman's work cited in chapter 2, he provides players with a matrix game followed by an exactly analogous real world
situation, to see if they play it similarly. He found, in fact, that they did not, and I suppose there should be no surprise in this. After all, it is asking a lot of anyone to expect them to see the nuclear arms race in the selection of letters off a two-by-two matrix. In working with the Prisoner's Dilemma with children, I have always found it essential to be explicit about how it is related to the real world. Second, the use of the package of games has brought into question what we mean by rationality. Early researchers termed as 'irrational' players who did not play the games according to game theoretic standards. The trend now is to accept responses to these games as manifestations of rationality, and to change our view of rationality rather than try to modify the manifestations. This view allows this statement which would have been unthinkable twenty years ago. William Poundstone in *Prisoner's Dilemma* says:

> When one type of "rational" behaviour fails, we expect the truly rational person to step back, think things over, and come up with a new type of behaviour. Given this open-ended definition of rationality, it's hard to accept that rationality ever runs aground. In a one-shot prisoner's dilemma, the type of rationality game theory recognizes leads to mutual defection, and all attempts to come up with alternative types of rationality have failed.

This broader view of rationality is exactly my quest in this chapter.

A third use of this package of games has been to throw light on the real world. They are no longer perceived, I feel, as adequate models of actual situations. They do seem, however, to be paradigmatic of the sorts of things that go wrong, particularly of the paradoxical situations in which we can find ourselves. This role is made all the more poignant by the fact that their structure and rules are easily expressed, but that they are almost impossible to win. We can, I would suggest, examine particular real world situations and say that 'this is a Prisoner's Dilemma' or 'this is like the Chicken Game.' In this way, although we are by no means describing the actual situation, we may at least be
able to apprehend or explain possible problems within it. I shall, therefore, consider a number of non-zero-sum games throughout the rest of this chapter, and use their explanatory power to expose the sorts of situational difficulties in which we can find ourselves.

It is tempting to see the Prisoners' Dilemma as a product of its structure (Temptation > Reward > Punishment > Sucker.) But when we consider the Chicken Game, which has a different structure yet still produces a dilemma, we are forced to conclude that the move from individual rationality to group action is more widely paradoxical than is at first apparent. It will be remembered that the Chicken Game is played by two car drivers, who drive on a collision course until one swerves, the one who swerves being the chicken. This can be modelled as a matrix game in the following way:

<table>
<thead>
<tr>
<th></th>
<th>Player II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice A</td>
</tr>
<tr>
<td>Player I</td>
<td>2,2</td>
</tr>
<tr>
<td></td>
<td>3,1</td>
</tr>
</tbody>
</table>

As can be seen, this is somewhat different to the Prisoner's Dilemma and as I have said, aims to model a somewhat different social interaction. Here, the co-operative choice (A) yields the second highest score, and would probably be reflected in a mutual undertaking not to play. The payoff from defection (choice B) is best, providing the other player backs down (becomes chicken.) The worst result is mutual defection, which has very serious consequences. This scenario, different to that of the Prisoners, still produces a dilemma. Like the Prisoner's Dilemma, individual defection is more profitable than mutual co-operation. In Chicken, however, the best individual result is based, not on a hope that most players will co-operate and that this can be exploited, but on the belief that this exploitable co-operation can be forced by the threat of a calamitous mutual destruction. Individuals in the Prisoner's Dilemma react to a co-operative milieu which is
exploitable; those in Chicken pro-actively endeavour to encourage such a milieu. The
dilemma in Chicken arises from the need to balance self-interest, boldness and belief in
one's perception of others. The parallels of this style of 'negotiation' with cold war
diplomacy and industrial relations are obvious. Russell referred to the game in his work
on nuclear war, and the method is more familiarly known as brinkmanship. Following the
previous classification, Chicken could be represented as:

Temptation > Reward > Sucker > Punishment

where, in comparison with the Prisoner's Dilemma, the last two payoffs are reversed.
Other dilemmas have also been proposed, based on further reworkings of this payoff
sequence. Let us consider, for example, the relationship:

Reward > Temptation > Punishment > Sucker

This is simply the Prisoner's Dilemma with the first two strategies reversed. It produces a
matrix like this:

<table>
<thead>
<tr>
<th>Player I</th>
<th>Player II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choice A</td>
</tr>
<tr>
<td>Choice A</td>
<td>3,3</td>
</tr>
<tr>
<td>Choice B</td>
<td>2,0</td>
</tr>
</tbody>
</table>

At first it seems unlikely that this matrix would generate anything other than co-
operation, since this yields mutually the best payoff. Many game theorists have,
therefore, dismissed it as uninteresting. It does, however, produce a vicious dilemma,
since there remains a temptation to defect. This temptation lies in the desire of a player
to do better than his or her opponent: to maximise the difference between their
respective payoffs. In game theory terms this is irrational: a gain of 3 is preferable in any
situation to a gain of 2. In practice, a score of 3 each will often be seen as a gain of 0, whereas an achievement of 2 against 0 will be seen as a gain of 2. I will support this conjecture with practical examples in the next chapter. If this temptation becomes paramount, then the game becomes similar to the Prisoner's Dilemma. The failure to co-operate is even more regrettable, however, since the results of co-operation are better for everyone in absolute terms than defection. Further, the move to defect is not based on a desire to do better than co-operation, but simply on a desire to do better than one's opponent, even though the absolute gains are less. This could be seen in real negotiation as cutting of your nose to spite your face, or as beggar your neighbour. In terms of dilemmas, it has come to be called the Assurance Game. The model implied here is that of Assurance or Insurance Schemes, or schemes in general where strength is gained from a solidarity which is mutually beneficial and based on a prima facie desire to co-operate. The dark side of such schemes are that, like the other dilemmas described, defection, which is a temptation, places too considerable a strain upon the co-operators, so that the whole scheme eventually collapses. An actual example might be found in a land improvement or reclamation scheme, where a group of farmers undertake to maintain a dam or sea defence. All do better than they did before out of the scheme. Once individuals begin to defect, however, the general quality of the defence worsens, leading to a deterioration in condition of the land. But these defectors still do better than the rest, who also suffer from the deterioration, and lose further from their continued contribution to the mutual defence.

A final dilemma emerges from a further restructuring of the strategies.

\[
\text{Sucker} > \text{Reward} > \text{Punishment} > \text{Temptation}
\]

Compared with the Prisoner's Dilemma, the Sucker and Temptation strategies have now changed places. The resultant matrix is:
By the Dominance Principle, the choice of A seems assured, since this does much better than B. As with the Assurance Game, therefore, the game seems uninteresting at first sight. The difficulty with the mutual A choice, however, is that it does not do quite so well as A-B would for the player choosing A. This would result if one player decided to play in such a way that his or her opponent gained the best payoff. We can signify this as altruistic play, and its motivation can be designated as the opposite of that of the Assurance Game: here the desire is to do worse than one's opponent. Again it is an acceptable strategy until a number do it, when once more the group action resolves itself into a dilemma - the Altruist's Dilemma. Suppose, using the above matrix, one player plays altruistically, then there is no problem: one gives (chooses B) and the other takes (chooses A). If both play altruistically, however, they both end up with a poor payoff. This is analogous to many actual situations where the benefits gained by individuals acting as groups only become benefits when individuals take profit from them. If people only give and do not take, then the benefits never actualise themselves, and we are left with a situation akin to that on Lewis Carroll's mythical island where everyone barely survives by taking in everyone else's washing. In these mutually self-denigrating groups the co-operative effort will eventually collapse. If two hundred of us decide to clear the litter off a piece of grass, then if we all do it the grass will be ruined. It is better that a few make a sacrifice and clear the grass, so that the rest can enjoy it. Many social exchanges require that only a few undertake actions for the benefit of many. The dilemma lies in ensuring that the right number, not too few or too many do so. Finally, in the Altruist's Dilemma, it becomes impossible to decide whether the altruist is co-operating or defecting. His or her role seems to turn very much on the final benefits or otherwise accrued by the group. The same can be said of non-altruists.
In informal systems, therefore, where there is normally no external or coercive agency available to decide roles, responsibilities and numbers, it becomes very difficult for individuals to know what to do. In particular, 'wants' and 'best interests' are compromised as a basis of rational choice, since the very act of choice itself seems potentially disastrous. Many, instead, shuffle uncomfortably between the role of free-loader and self-sacrificer, roles which are all the more difficult to ascribe when self-sacrifice, by the Altruist's Dilemma, leads to a deleterious result for the group, or free-loading, by the Prisoner's Dilemma, leads to equally disastrous results. By turns, it seems that both too many cooks spoil the broth and many hands make light work. More purposive individual negotiation or action based on a desire for self-advancement or a desire to lead or sort out the mess seems to lead inexorably to dilemmas such as those in the Chicken or Assurance Games. Finally, I would like to consider the further exacerbation caused by the issue of self-binding and future commitment mentioned earlier with reference to Ulysses and the Sirens. This adds still more to the difficulties of dealing rationally with the dilemmas and is best illustrated by reference to a paper, *Prisoner's Dilemma and Resolute Choice*, written by Edward McClennen. Instead of arguing that the dilemma resides in the nature of the choices and their utilities, McClennen says, instead, that the dilemma is created by a player's expectations of her future actions. By skewing the payoff matrix slightly, so that only one player now has dominant strategies, McClennen shows that both will still defect. He argues as follows. Suppose that B's defection choices dominate her co-operative ones, but that A's do not. B knows, therefore, that in a mixed choice situation, A will tend to co-operate since this provides the best payoff. B, on the other hand, fares best with defection, since this is a dominant strategy for B. What is more, A knows defection is B's best strategy, and this will provide a poor payoff for A if he co-operates. B, assumes, therefore, that A must defect. But this defection rests, not on A having a dominant defection strategy, but simply on B's rational, dominance-based predisposition to defect. B, therefore, is trapped into defection by the knowledge that, whatever B wants at present, at some later time in her internal debate she will conclude that defection is the most rational course. McClennen further illustrates this point by
proposing a Prisoner's Dilemma game played sequentially. Again, no matter what a player wants to do at the start of the game, she will be pressured, when her turn comes about, into defecting. As a final twist, McClennen then argues the following. Suppose that we accept, therefore, that player B at present is debating her move in the game, using all the relevant information she can muster. From this, she concludes that herself in the future will defect. How, then, does she treat this piece of relevant information? Obviously, this future B will actually be responsible for carrying through the plans which the present B is devising. The needs of this future B must, therefore, be taken into account. This is acceptable if the present B and the future B are in accord. If they are not, however, what can the present B do? Suppose, for example, that the present B sees co-operation as the sensible action, but knows that, at the time of choice, the future B, through lack of will or wilful self-interest, will choose to defect. Should the present B take steps to bind her future self in some way? Is self-binding in her best interests, or might the future B truly have her best interests at heart? Finally, as we saw earlier, self-binding seems to be a paradoxical act anyway. For all sorts of reasons, the Prisoner's Dilemma and its associated non-zero-sum games provide a complex and many layered trial for rational thinking.

Ultimately, therefore, the work of Arrow, described in the previous chapter, and investigations into non-zero-sum games, show that rational attempts to find a group consensus which adequately reflects individual preferences is doomed to fail. The group ethos can appear perverse and irrational, even though the individuals which compose the group are anything but. Management, likewise, can view group perversity as malicious when it is not. Since rational individual action often leads to the irrational group outcome, I will argue in later chapters that irrational decisional activity on the part of individuals may be as valid as the rational variety.
HOW THE PARADOXES INTERACT

The terrible idea - terrible to some, terrible, perhaps, to all, at some time or in some form - that we are biologically predestined like other creatures, that we differ from them only in inventiveness and the capacity for reflection on our fate - treads softly behind the arrogant judgement that makes of the ant a twitching automaton.

A. S. Byatt, Angels and Insects

It is important to understand that the paradoxes also act in concert, making rational decision-making doubly difficult. I shall use as the basis of my consideration of this issue a paper, Rational Actors in Macrosociological Analysis, by James Coleman. This not only provides a practical scenario for the application of rational action, but also allows the application to the scenario of four paradoxes. I hope to show that these paradoxes survive the extended view of rationality which I have earlier proposed: they do not vanish when the narrower game theoretic definition of rationality is abandoned. The paradoxes are: Arrow's, Newcombe's, the Prisoner's Dilemma and the Sorites. I have described Arrow's, the Prisoner's Dilemma and the Sorites. It is now necessary to describe Newcombe's Paradox.

Although Newcombe's Paradox is a paradox of individual decision-making, I intend to show that it is crucial to the problems attendant upon creating a group ethos. It proposes the following scenario. There is a being (not necessarily human) who is known to be a successful predictor. You are told that at some time in the past the predictor has made a prediction, concerning your choice between two boxes, a transparent one and an opaque one. The transparent box can be seen to contain £1000. The predictor has signalled his prediction in this way:
If he predicts you will choose the opaque box only, he has placed £1000000 in it. If he thinks you will choose both boxes, he has placed nothing in the opaque box.

This scenario is seen as paradoxical because two perfectly acceptable methods of deciding which choice to make lead to contradictory decisions.

To explain this, let me first present the scenario in strategic form, showing the choices available to both predictor and player and the payoffs to the player:

<table>
<thead>
<tr>
<th>Player</th>
<th>Predictor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Put money in opaque box</td>
</tr>
<tr>
<td>Opens both</td>
<td>£1001000</td>
</tr>
<tr>
<td>Opens opaque</td>
<td>£1000000</td>
</tr>
</tbody>
</table>

This seems a satisfactory way of representing the choices, and seems to present a satisfactory solution. Obviously, by the Dominance Principle, the Player gains by opening both boxes and not one, whatever choice the predictor makes. The choosing of a strategy which gives a better payoff then any other is known as the Dominance Principle. In opposition to this, many have argued that the game in the form above does not capture all the information. It is known, for example, that the predictor is mostly correct. This supposition, presumably, is an inductive insight based on the predictor's track record. Let us suppose, for the sake of argument, that the Predictor is right 9 times out of 10. The above matrix does not take this into account: it treats all outcomes as equally likely. In fact, the outcomes should be weighted to take into account the Predictor's ability. Effectively, this means multiplying the outcomes at bottom left and top right by .9 since this is the likelihood of their actually happening. The other two outcomes, likewise, should be multiplied by .1. The result is a Utility: the expected or preferred payoff, as opposed to the ordinal or unweighted payoff which is shown in the matrix. Since the
£900000 now produced at bottom left far outweighs the £100100 now available at top left then the logical choice is to choose only the opaque box.

Unfortunately, the debate between two-boxers and one-boxers does not end there. The two-boxers have countered the utility argument by questioning the basis upon which the utilities are decided. After all, they say, they are based on the assumption that the Predictor makes good predictions. Nevertheless, it has made the prediction, placed the money and left the building. Whatever the Player now decides, the Predictor can have no control over it. If the money is in the opaque box, then it is in the opaque box. If it isn't, it isn't. So the player might as well take both boxes. To argue against this is to argue that the Predictor has some causal influence over the choice of the Player. To all intents and purposes, say the two-boxers, the one-boxers are arguing, not from a standpoint of induction, but from backward causality: that the result of an event can somehow cause it. This is irrational. The one-boxers reply from the point of a third party watching the game. If they were to bet on the game, knowing the players and their records, what outcome would they bet on? Almost certainly, they would choose the one box as the most likely outcome.

This paradox is useful because it allows certain decision-making processes - the dominance principle and maximising utility - to be described and explored. If this is as far as it goes (and in many analyses this is as far as it goes), then it is of interest to decision theorists, but may be of little interest to us. After all, the aim of this chapter is to show how a broadened rationality can dispel such paradoxes, which achieve their effect from their narrowness. It is, in fact, much broader and deeper than a simple conflict of decisional styles. For this reason it is not so easily dismissable as is, for example, the Unexpected Exam.

First, it can be used to illustrate the nature of paradox itself. In his book *Labyrinths of Reason*, William Poundstone side-steps the one-box - two-box controversy by looking at
the structure of the paradox itself. His argument put briefly says that the scenario is paradoxical because the prediction refers to itself, and in revealing itself gives information to the Player which itself will affect the prediction. It is similar to a prediction which says "I predict you will stand on your head in the next two minutes." You may have had every intention of standing on your head but now you determinedly won't. Alternatively, you may have a burning desire to prove the prediction right. Either way, knowledge of the prediction gives you power over it. In *The Ghost in the Machine*, Koestler says of predestination:

one of the conditions for fulfilling the pre-arranged pattern is that I should not believe that it is pre-arranged. Destiny can only have its way by forcing me to disbelieve in it.

Girolamo Cardan, the mathematician and fortune teller, committed suicide on the day he predicted he would die in order to make the prediction come true. In chapter 2 I described a news bulletin warning of traffic congestion where the motorists heading towards it were, by their actions, in a position to influence the prediction. In other words, a prediction about your own actions is essentially a different prediction to "There will be a small earthquake in Chile" which does not have this self-referential flavour. Following Poundstone's lead, therefore, I will reformulate the Predictor's prediction. I can appropriately accomplish this with reference to a science fiction story I once read. Here, a man invents a time machine and goes back to the time of Shakespeare. He materialises in the theatre during the first performance of Macbeth and finds he has to take on the role of the third murderer. Saying the third murderer's line, which he knows from his study of Shakespeare, he then leaves the stage. Shakespeare hears the line, and incorporates it into the play. The question is, where did the line come from? The time traveller didn't invent it: he learned it from the play. Shakespeare didn't invent it: he heard it from the time-traveller. We have a similar problem with the prediction. We can restate it as "I predict that you will believe my prediction and take just one box." More simply,
the prediction is "I predict you will follow my prediction." In that form, it is impossible to state what the prediction actually is. It is a prediction about predictions and so seems only definable tautologically, by infinite regress. Like the paradoxes I have described before, rational treatment seems to lead to contradiction.

Presented in other forms, Newcombe's Paradox can inform reflections on decision-making still further. Let us suppose we have just observed someone pick up a piece of litter in the street and place it into a bin. We ask her why she did it. The answer, a perfectly natural and common one, is: "because we should all do it." Jon Elster, in his book *The Cement of Society*, relates this reply to Kant's Categorical Imperative, which he restates as a stricture to co-operate if universal co-operation would benefit everybody. This is a perfectly sound philosophy; the notion of universal co-operation implies a general improvement in the quality of life enjoyed by the group. Indeed, if everyone picked up litter and put it in a bin then life would probably be better for all. Unfortunately, the reply "because we should all do it" is a programme for social amelioration not a reason for individual action. As a reason for individual action it is an example of what Elster calls "magical thinking." By this he means a process which supposes that an individual's actions can affect other actions and their earlier causes. In choosing to act in a certain way, a magical thinker assumes that they can affect the actions of others. In *Gone With The Wind*, Melanie exhibits this thinking style when she says that the reason she looks after destitute Union soldiers returning North is because some Northern matriarch will be looking after her Ashley in the same way. Of course, this mode of thinking is not blatantly irrational or magical in a ritualistic sense. Rather, the thought pattern is more of a wish fulfilment. It sees the individual as acting in response to particular causal stimuli. The individual goes on to assume that these causal stimuli affect others. By manifesting the action, therefore, they are enacting what others will also enact. I once asked someone why she used the local, and more expensive corner shop when she could use the more distant, but cheaper supermarket. Her reply, "because there might be a time when we need it", shows exactly this kind of thought process.
Obviously, her use of the shop would make no difference to its survival or otherwise. But her use of it as a typical part of a general desire to use it would. Her action is causal and sensible, almost, in a topsy-turvy sort of way. Further, the individual need not necessarily believe in the action: the manifestation is sufficient to assure that the causal stimuli are there. The causal stimuli are seen as ironclad, having almost a predictive quality about them. Elster refers, by illustration, to the Calvinist belief in predestination. Although an individual is elected at birth to be saved, this election itself is signalled by a willingness to behave in the manner of one elected to be saved. Action based on prediction-fulfilment is exactly the sort explored by Newcombe’s paradox. One-boxers, in many senses, are magical thinkers. I hope to show shortly that this style of thinking is common, and has extensive effects upon group decision-making.

Having outlined Newcombe's Paradox, I will now return to Coleman's scenario with which I began this section. Coleman sets out to explore the interactions which take place within organisations. He sees these as of three types. First there is the relationship which the management of the organisation has with its members. He refers to the members of the organisation as agents, in that they act for the organisation in the achievement of its goals or actions. The managers of organisations usually have contractual relations with the agents: for example, they pay them for the work they do. In return, the agents, or employees, are expected to successfully discharge their responsibilities to the organisation. The second type of relationship within the organisation is the formal one between employees: that is, the relationships between members which are dictated by the organisation. A pay and promotion structure would be an example of such a relationship. Third there are the spontaneous social exchanges which take place between different members and groups within the organisation. These might be termed the informal component of the organisation. The nature of each type of relationship obviously affects the general efficacy of the organisation. Coleman shows, by reference to published research, that there is an extensive literature on each type, but the relative importance of each in actually determining the state of the organisation has changed. In particular, the
third, or informal relationships, largely ignored in early work, have come to assume major importance. Further, and most interestingly, Coleman argues that, since these relationships are spontaneous and complex, they do not easily fit into tidy theoretical frameworks. Consequently, although theories exist on optimal formal or contractual systems, little can be said in a theoretical way about informal systems. Worse, these informal systems cannot be ignored, since they seem to generate themselves in all organisations. This is confirmed not only by observation, but by the application of theories about how people can be expected to behave in groups. Here Coleman is less precise, but I intend to argue that the very scenarios he proposes, which generate, and are symptomatic of, informal groupings and relationships, are the very ones mentioned above: Arrow's Paradox, Newcombe's Paradox, the Prisoner's Dilemma and the Sorites. I shall further argue that since these paradoxes underpin aspects of informal groupings and relationships, and that since the paradoxes by definition generate contradiction, then several aspects of group culture are contradictory and hence create obstacles to the rational development of organisations. So although the managers may be able to control the formal and contractual aspects of organisations, since they to an extent form and monitor these aspects, they can only control informal aspects with great difficulty.

Coleman proposes two scenarios to illustrate the development of informal systems. As I have suggested, however, his discussion of their impact is only tentative. Comments about their implication and application are, therefore, my own.

His first scenario relates to Arrow's Paradox, and is expounded in exactly the terms that I used above to produce an intransitive relationship between three choices A B C. I would contend that this voting procedure can be used as a representation of a formal system: a device set up by the managers of an organisation to canvass the opinions of its workforce, and to find relationships between individuals and their support for individual issues. As was observed when I described the system earlier, this can easily reach an impasse. Coleman suggests that in practical terms, however, the impasse created by the
formal procedure can be circumvented by informal action. One such action is to find the value that each voter places on an individual outcome. This could be related, for example, to the monetary gains that each voter would receive from each outcome. A practical example should make this clear.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1,9</td>
<td>3,2</td>
<td>2,3</td>
</tr>
<tr>
<td>B</td>
<td>2,1</td>
<td>1,5</td>
<td>3,2</td>
</tr>
<tr>
<td>C</td>
<td>3,0</td>
<td>2,3</td>
<td>1,5</td>
</tr>
</tbody>
</table>

In the above matrix, three voters have put 3 outcomes into an order of preference: the figures on the left of the commas indicate the outcome's position in this order. The one preferred by most will be accepted. It is easy to see, however, that there is an intransitive group preference: A is preferred to B, B to C and C to A. On the right of the commas are the monetary value to each player of the outcome's actual selection. The selection of outcome A, therefore, is worth £9 to player X, whereas the selection of B is worth only £2. Likewise, the selection of B is worth £5 to Y, and so on. Coleman argues that the intransitivity of the group selection could easily be resolved by side payments. If X were to pay £3 to Y and £2 to Z for their vote, then she would achieve the unanimous selection of A and still have £4. I said earlier that informal systems were usually complex as well, and the present scenario also illustrates this. Let us suppose that X is not content to receive £4 when the others have achieved their maximum payoff. By buying Z's vote for £2, X can still assure the selection of A by majority vote. This leaves X with £7, Z with £5 and Y, unfortunately, with £2. Y could, therefore, offer to swap to C which would gain him £3, leave Z with £5 and X with nothing. X would then need to buy Y's support at the expense of Z, who could then support Y to get back at X, and so on. In *The Theory of Games and Economic Behaviour*, John von Neumann and Oskar
Morgenstern rapidly concluded that n-person games were impossible to analyse mathematically, because the resultant formation of alliances, and, more precisely, the stable alliances which would eventually form and the side payments that would underpin them, were a result of the psycho-social mix of the players, not the result of mathematical features such as Pareto optimality. Complexity and unpredictability, therefore, are endemic to the formation of informal systems. The eventual bargain, which assures a non-transitive outcome, however, is essential to the well-being of the organisation and its members. Given our definition of rationality, it is in the best interests of the members to make a bargain. While their group choice is intransitive, they are not obeying their formal directive to make a choice. Their inability to follow the dictates of the organisation may either degrade or terminate their position in it, or so hamstring the organisation that its demise will leave their positions terminated by default. Although, consequently, they will bargain to gain their best individual payoffs, their best interests as individuals are served by reaching a bargain. Informal structures and bargains, therefore, are not only endemic in organisations, but are also the most rational way for organisations to operate. The denial of informal systems in the face of intransitivity or inconsistency in the formal system, will not lead to the best interests of either the individual or group being served.

The price to be paid by management for the growth of informal systems is, as I have said earlier, that they are complex, unpredictable and difficult to control. Further, the following scenario illustrates that informal groupings tend to fall foul of the Prisoner's Dilemma and Newcombe's Paradox. These paradoxes make it difficult if not impossible to act rationally, in that they make it impossible for the members of the group to define clearly what is in their best interests. Although, therefore, the formation of informal systems is rationally appropriate for organisations, these systems at times are prone to irrationality. Coleman proposes a second scenario which illustrates this. I use this scenario again because it is apt and insightful, but the conclusions drawn from it are my own, not Coleman's.
Coleman asks us to imagine the rapid evacuation of, for example, a cinema. This could be as a result of fire. He suggests that two responses an individual might adopt when faced with this eventuality are either to rush out or to file out. He constructs a payoff matrix, where each of the alternative strategies are related to those adopted collectively by the rest of the group trapped in the building. Simplified (a step which Coleman does not take), the result is a Prisoner's Dilemma. The payoffs are the chances of the individual's survival.

From the point of view of the individual, her ordered chances of survival in relation to the action of others is: rush if they file, file if they file, rush if they rush, file if they rush. This is the classic Prisoner's Dilemma, where the best payoff is obtained by the individual both acting selfishly but hoping the others do not. In fact, when the others do, the payoff is worse than if each had been selfless. Individual acts of selflessness, however, yield the worst payoff in the face of general selfishness. So far this is a standard scenario, which has been discussed above. Coleman adds an interesting twist, however, by adding a second scenario. To describe the impact of this second scenario more fully, I will first need to digress. I want to look in more detail at the relationships between the Prisoner's Dilemma and Newcombe's Paradox.

In one of his 'Black Widower' stories, Isaac Asimov has one of his characters talk about the belief in four leaf clovers as the bringers of luck. Although the possession of such a clover can obviously have no effect upon a person's well-being, it is conceded of people that:
The sense of security they experience is very real. The cause may be unreal, but it brings the desired result.

This is a seemingly senseless statement. How can an unreal cause bring about a result? Surely only real causes do that. Acceptance of this unreal cause flies in the face of logic and experience which says that there is no causal link between four leaf clovers and well being. To paraphrase Elster, this kind of cause might be called a magical cause. Like the Unexpected Egg, and many other objects and outcomes of paradox, it exists on a sort of shadow line between existence and non-existence, being or not being according to changing perspective. And this changing perspective is generated by the perception of the object itself: every time we seem to have established its nature this nature seems no longer to make any sense. There is, in this case, a way out of the paradox. The magical cause can be conceived of, not as acting upon its result directly, but by way of a third, or intermediate cause. This intermediate factor might, in this case, be considered as belief. If one believes in the efficacy of a four leaf clover, the possession of such feeds the belief, which in turn generates the feeling of security. The intermediate factor here is exactly the same as that I proposed earlier in the discussion of collateral outcomes and by-products: it is the mitigating factor or catalyst which generates the outcome from the action.

Many commentators have tried to represent Newcombe's Paradox in the same way. The four leaf clover scenario does not, however, capture its nuances sufficiently. Let us consider, instead, the issue of eating sweets and the occurrence of bad teeth. Suppose there was a correlation between individuals who consume large numbers of sweets and have bad teeth. Such a correlation may in fact exist, but for the purposes of the argument I shall make the reasonable assumption that it does anyway. How should such a correlation be explained? One explanation would be that the sugar in sweets rots teeth. A second, more extreme explanation might argue that the predisposition to rotten teeth might be genetic, and that this genetic phenomenon is also manifested in a compulsion to eat sweets. By this argument, eating sweets and tooth decay are not directly linked, but
are both products of a third agent. The reason that they appear to be causally linked is that the presence of the third agent insures their contiguous occurrence. Suppose now that you have heard this argument, believe it, and wish to avoid tooth decay. You also have an overwhelming desire to eat a sweet. Would you eat it? The answer to this seems to divide people in exactly the way that Newcombe’s Paradox does. A refusal to eat the sweet implies in some way a commitment to magical thinking or backward causality: that a refusal to give in to the symptom or outcome can in some way influence the other outcome through the third, causal agent. A contrary argument would state that if you have the genetic weakness anyway, your teeth will go rotten whether you eat the sweet or not, so you might as well eat it. After all, the genetic theory might be wrong or there might be other reasons why you feel compelled to eat sweets. If this seems a trivial example, the argument when it was proposed originally, by the statistician R. A. Fisher, was applied to a possible genetic correlation between developing lung cancer and the desire to smoke. In this case, the scenario proposed that if you had a desire to smoke, you may be exhibiting the signs of an early death. Again it was asked if an individual would deny or accept a desire to smoke in these circumstances. In more general terms, the assumption that contiguity implies causality is the basis of the fallacious argument *cum hoc ergo propter hoc*. Since this fallacy is well known and of great antiquity, there is certainly no general basis to a belief that contiguous events are causally linked. Beliefs that a third, hidden factor may link two other factors are equally tenable.

The connections with Newcombe’s Paradox are obvious: in fact, it becomes a special case of this general situation where two outcomes are linked by a third factor, and not by direct causality. In the case of Newcombe’s Paradox, the non-causal nature of the link is emphasised. The choice of box does not affect its contents: the contents have been determined by the predictor who has now left the building. What does link the choice, and the contents of the box, are the prediction. To deny the ability of the predictor and to take both boxes is analogous to denying the genetic relationship between tooth decay and eating sweets, and simply to argue a causal relationship. The acceptance of the
prediction is analogous to accepting the genetic relationship: to accepting that the iron hand of the prediction causes both the contents of the box to be as it is and also causes the wise selection of the single box. The only difference between the tooth decay scenario and Newcombe's Paradox is that if the power of the gene or predictor is accepted, then the tooth decay scenario has an unpleasant outcome so the action of eating sweets is denied. Conversely, in Newcombe's Paradox, the action of taking one box has a pleasant outcome.

The Prisoner's Dilemma can also be seen as another special case of an apparent causal link being actually a link by a third agent. The third agent is more difficult to give a name to, but I will call it *representability*. Just like the tooth decay gene or predictor it may not actually exist. The possibility of its existence, however, worsens the paradox, and the resultant inability to make a firm decision. This hesitation overlays the Prisoner's Dilemma and makes it a double obstacle to rational thought. Basically, representability is the individual's belief that he or she is representative of the feelings of the group. An acceptance of the principle means that the individual's choice will also be the choice of other individuals. A denial of the principle means that each individual makes a choice by their own lights which bears no relation to the choices of others. Let me illustrate this with reference to the burning cinema scenario. You believe in the principle that you are representative of the general view of the group. You have an overwhelming desire to rush from the building and save your own skin. The representability principle says that everyone is thinking the same, which will lead to the rush - rush selection from the matrix and low chances of survival. Do you deny the desire to rush?

This is the skewer that Coleman's addition to the scenario places us on. In his case, he does not argue as I have done, via Newcombe's Paradox to a third, mitigating agent called representability. More simply, he argues that you as the individual might actually have the power to lead opinion or action by your example. I do not feel that this power of leadership is accepted by sufficient people of themselves to be a factor. I do feel, as I
have suggested, that most people feel they are typical and that they think or do mostly what others think or do. If the concept of representability is too passive, however, we could replace with the idea of *expectation*. Given that a person can sum up a particular situation, she might go on to *expect* a particular outcome. Since she is not unique, it is quite acceptable for her to assume that others expect the same. This active expectation could lead to the emergence of the appropriate outcome. I will try to illustrate the idea more pointedly by reference to the voter paradox. This asks why anyone should bother to vote in an election, when it is obvious that an individual vote will count for absolutely nothing: elections are never close enough to be decided by a single vote. In fact, large numbers of voters act *en masse* and their effect is better judged probabilistically, like other natural forces. This probabilistic approach is advocated, for example, by Peter Ordeshook in his book, *A Political Theory Primer*, which is an excellent exposition of the applications of game theory to political situations. Whilst treating political processes, he says:

\[ \text{nature (i.e. the electorate *en masse*) does not condition its choices on the citizen's action.} \]

In contrast to this, however, my concept of representability or expectation suggests that an individual action speaks for, or betokens the actions of the group as a whole. In this sense, the individual action could be said, if not to condition the group choice, at least to herald or predict it. As Hollis puts it in *The Cunning of Reason*:

\[ \text{When an individual switches his vote for good reason, it is as if he were switching all the votes of those like him.} \]

As individuals in group decision-making situations - for example, when we vote - we must continually ask whether our actions have value only in themselves, or whether their
manifestations have implications that are wider or deeper. And this is exactly the dilemma that Newcombe's Paradox points.

The combination of a scenario such as the Prisoner's Dilemma, and the third agent principle exemplified by Newcombe's Paradox and transformed into a concept of representability quite powerfully describes people's inability to act, at times, purposefully and rationally. The want and need to act rationally is present, but there is an absence of a clear cut choice, and a sensible, unambiguous way of obtaining data. Instead, the members of the group circle endlessly around the dilemma. The simple choice between defection and co-operation is a problem for rationality, since both and neither seem to be in one's best interests. The problem is compounded by the meta-choice between one's commitment to or denial of one's representability. It is a meta-choice, because its solution might make the choice of co-operation or defection easier, although I suspect not. I believe that people waver between choices and levels aimlessly, and the group pays the price of indecision. Later, I will support these theoretical observations with conclusions drawn from work on the Prisoner's Dilemma in the classroom.

In summary, I have argued that although groups have formal arrangements for decision-making, they necessarily and rationally develop informal systems. At times, however, these systems are themselves prone to irrationality, not by intention, but by the choices of individuals in aggregation, and their inconsistent attitudes towards causality. It is appropriate now to consider the influence of the Sorites Paradox. In particular, I want to consider its relationship to the 'free-rider problem.' Free riders are people who, in a group which acts concertedly to provide a good, can benefit from the good without having to make a contribution towards it. So a person who avoids taxes still benefits from the public goods provided by these taxes. The free-rider can be modelled by the Prisoner's Dilemma, in that the scenario encourages defection by gain, and the definition of rationality I have proposed does not make it obvious whether free-riding is rational: if it is in a person's best interests. In a way it is, because the individual avoids their
Paradoxes in Decision-making

Paradox Lost and Paradox Regained

collection. By the concept of representability, however, it is not, because if everyone
free-rides, then no one benefits. The problem of the rationality of free-riding, however, is
better illustrated by the Sorites Paradox, because the problem is really an aspect of it.
Suppose there is a drought, and an individual wishes to water his garden. By the Sorites
Paradox, he could argue that the amount of water he removes from the pile (or reservoir)
is so slight that it makes no difference whether he waters his garden or not. So he might
as well water it. Further, if he does not water his garden, the amount saved is so slight it
makes no difference to the heap. So again he might as well water his garden. The
problem with this rationale is that if everyone follows it then everyone will suffer even
more. Like the Prisoner's Dilemma modelling of the scenario, there is no clear cut
outcome which is in the individual's best interests. Best interests are served not only by
the individual, but by the actions of other individuals. Again the solution of the paradox
for each individual depends upon their view of their representability.

CONCLUSION

The broader, non-game-theoretic definition of rationality does not seem to save the
individual and group from paradox. The Sorites, Newcombe's and the Prisoner's
Dilemma are all still ruled in, and are compounded by a spiralling self-reference of level
shift. All bedevil informal relations. In his work cited above, Rescher says an irrational
person is:

prepared to act against the welfare and best interests of himself and his
community

Rescher's refusal to acknowledge that any dichotomy might exist between personal and
communal interests shows how easy it is to ignore the lessons of these paradoxes. By
them, both 'wants' and 'best interests' are compromised as a basis for rational decision,
since choice itself is fatal. Unfortunately, if they vanish from a concept of rationality, it is
Paradoxes in Decision-making Paradox Lost and Paradox Regained

difficult to know what to replace them with. Self-preservation or at the least, a commitment to non-injury seems essential to rationality. Any action which promoted conflict ether internal or external would seem to be irrational. There seem to be two ways forward.

1. To find yet another definition of rationality, or, at least, to consider ways of living with paradox. In chapter 5, I will look at rationality based on ambiguity, psychological consonance or downright irrationality.

2. Although individuals and groups cannot avoid paradox with its concomitant rational difficulties, mostly people do not encounter these situations and so can mostly avoid irrationality. In the next chapter I intend to illustrate several paradoxical situations with practical examples. In recognising them, we may be able to avoid them.
Chapter 4
Paradox in practice

But it is worth noting that an 'inconsistent' language can be used without embarrassment by anyone who steers clear of certain questions, in much the same way that a car which would fall to pieces at ninety miles an hour can safely be driven at more modest speeds.

J.L. Mackie, Truth, Probability and Paradox

As well as the direct relationship between decision and paradox, which I have so far considered, there is also an indirect one; that of metaphor. To view decision-making as paradoxical allows the processes to be viewed from a new perspective, and allows insightful comments to be made about them. In this sense, paradox in decision-making serves a didactic purpose. It allows us to explore and illustrate the difficulties inherent in decision-making in a meaningful and interesting way. In his book Games of Life, Karl Sigismund says:

But surely the insights offered by a model are at least as important as its predictions: they help us understand things by playing with them, just like a child learns much by playing with dolls.

In this chapter I want to explore the sorts of insights which viewing decision-making as in part paradoxical provide. By looking at things in the light of paradox, I hope to draw conclusions about how individuals and groups in educational establishments manage themselves. My conclusions may not always involve paradox, but my deliberations always will. This pedagogic strain, I hope to show, will allow us to more fully understand and, thereby, control the decisional milieux in which they find themselves. Part of my thesis argues that the job of management is to steer the group away from paradoxical
situations, and this only becomes possible, at first, if they can be recognised for what
they are, accepted as existing, and differentiated from the other problems of decision-
making. In this chapter, therefore, I intend to consider the implications and ramifications
of viewing decision-making as a paradoxical process and to give actual examples of its
occurrence in education. This will not only support the contentions of the first three
chapters, but will open up the didactic strain, and clear the way for a discussion of the
amelioration or acceptance of paradox in Chapter 5.

The practical examples I intend to use are drawn from a variety of educational scenarios.
I will not only describe the situations and discuss their appropriateness as support
material, but also show how they can be used to explain and illuminate paradox and its
influence. As I said at the end of the previous chapter, if we wish to avoid the effect of
paradox, it is important that decision-makers understand its nature. Practical example, as
well as theoretical discussion, are essential in this strategy. The scenarios I will consider
are:

1. A classroom approach to paradox - the individual, the group and the prisoner's
dilemma

2. Purpose, aims and outcomes - rational action and by-products

3. Structure, ethos and rules - tacit agreement, the sorites and magical thinking.

4. Policy making and the Senior Management Team - level shifts and self-binding

It will be seen that these relate to the scenarios I introduced at the beginning of chapter 1
and the features of organisations - people, purpose, structure, management - that they
illustrated. They are not meant to be consecutive, mutually exclusive or exhaustive, but
samples of the decision-making milieu. They do not, therefore, say everything there is to
say about decision-making in schools, nor are they meant to. What I hope they do is to show the extent and depth of the decision-making space, and the wide variety of situations where paradox might be encountered: these paradoxes are added as hyphenated subtitles to the above headings. Some scenarios are manufactured and artificial, some based on everyday activity and experience, some on philosophical or generalised concepts of an educational nature. When I discussed Sutherland's work on irrationality, I said I was impressed with the wealth of example he brought to support his case. I intend in this chapter to try to support my case in a similar way.

A CLASSROOM APPROACH TO PARADOX - THE INDIVIDUAL, THE GROUP AND THE PRISONER'S DILEMMA

For civilisation to survive, the human race has to remain civilised - a small exercise in logic from the Twilight Zone.

Rod Serling

In order to more fully explore and explain people's responses to the problems posed by paradoxical situations, I devised a number of lessons based on the theme. There follows a description of the lessons, with observations on their outcomes.

LESSON 1 - THE PROBLEM OF COLLECTIVE ACTION

This is the keynote lesson, and is used to set the tone and raise the issues for later work.

The lesson centres on a film from the television series, The Twilight Zone. This consisted of a series of independent stories, of a science fiction flavour, produced in the early sixties. The particular episode under consideration is called The Shelter. It tells of a group of comfortably-off American families who are suddenly faced with a nuclear alert. Only one family, the doctor's, has a fallout shelter, which he has spent a good deal of time constructing. The family enter the shelter and close the door. As the episode
unfolds, other families arrive and ask to be let into the shelter. Unfortunately, the air filter in the shelter will only support the doctor, his wife and their young son. Eventually the mob outside the shelter become uncontrolled and they end up smashing in the door, thereby ruining the shelter for everyone. At that moment, a deus ex machina in the form of a stand-down announcement on the radio ends the action: everyone is safe. A suitable moral pronouncement about civilised people acting in a civilised manner terminates the programme. Two questions are asked of the class. The second is asked after responses have been taken to the first.

1. Did the doctor have the right to lock his own family in the shelter and refuse some others the right to survive instead?

The invariable response to this question is yes. The doctor was in the right and the rest of his neighbours were in the wrong to make demands of him. After all, the programme itself, through the doctor, makes the point that he spent a long time in building his shelter while the others were enjoying themselves. The doctor has earned the right to survive and the rest should have acted more responsibly.

2. If the doctor has the right to ensure the survival of his family, don't the neighbours also have this right?

This is a perplexing question, and classes have difficulty answering. This is not surprising, because it is the crux of the Prisoner's Dilemma. If the doctor has the right, then presumably so have the others. The doctor cannot really expect to be the only one to act selfishly while the others act altruistically. The work he has already done is irrelevant. Whether you plan for safety in a nuclear war over five years or over the last five minutes does not deny an individual's right to try to survive: death should not be the penalty for lack of foresight. If, however, we allow the right of survival to all, then the film shows clearly the consequence: the destruction of the one amenity which could save
lives. So everyone dies. Where a resource is scarce, therefore, some should perhaps forego their right of survival so that others might live. But who should survive and who should die? Certainly there are not only problems for rational choice here, but also deep moral issues to be considered.

**LESSON 2 - RULES**

This lesson aims to consider how groups develop informal systems based on rules or norms. In particular, the aim is to show the spontaneous development of these informal rules based on a willingness on the part of individuals to co-operate with each other.

The lesson consists of a series of games played in groups of three players. Two play each time and the third acts as scorer and umpire. Three games are played in a set, so each player plays every other. The games are based on choices from matrices, in the way that the Prisoner's Dilemma are represented. However a matrix representation is not used. Instead the matrices are turned into separable games. These are much simpler to play. For example, the first game uses the following matrix:

\[
\begin{array}{c|cc}
\text{Player II} & \text{Choice A} & \text{Choice B} \\
\hline
\text{Choice A} & 2,2 & 1,2 \\
\text{Choice B} & 2,1 & 1,1 \\
\end{array}
\]

This is a mild form of the Assurance Game, where the urge to defect lies not in a gain in points over what can be gained by co-operation, but by a gain in points over the other player. This game would be represented in separable form by the following rule:

*Choose a counter (Red=1, Green=2). Give the score you choose to the other player.*
An analysis of the different combinations which can be played will reveal that it generates exactly the payoffs shown in the matrix.

A second game has this rule:

*Choose a counter (Red=1, Green=2). Give the score to the other player, except if you both pick red. In this case you get 0.*

This rule generates a similar matrix to the first, except that a mutual B choice yields no score. This is a stronger form of the Assurance Game. Here there is a much greater incentive to co-operate and to continue to co-operate, because mutual defection gives a lower payoff than co-operation. The only excuse for defection is a desire to score more points than your opponent.

A third rule is:

*Choose a counter (Red=1, Green=2). If you pick red you keep the score. If you pick green you give the score to the other player.*

An inspection of the payoffs from all combinations of choices here reveals that the resultant matrix is the Prisoner's Dilemma. If the actions are reversed (give the red score and keep the green score), then we have the Altruist's Dilemma, where the player selecting green gains even more if the other selflessly picks red. Both suffer if they both act selflessly, however.

A final rule is:

*Choose a counter. If you choose green, score 1 for yourself and give 2 to the other player. If you choose red do nothing.*
This again produces a form of the Assurance Game, where the only incentive to defect lies in gaining a better score than the other player, even though a mutual choice of green gives the best score of all. A choice of red in this game is particularly negative, in that it contributes nothing in any circumstances. A mutual choice of red could well lead to a chicken run, as each tries to force the other to pick green and at least gain something.

By using separable games therefore, many games can be easily played and their outcomes analysed and discussed.

The lesson employs the first two games described. Eventually all groups learn to co-operate, but the manner by which they learn is exceptionally interesting. First, it is necessary to emphasise that although co-operation is not specifically requested, it is fully enabled by the way in which the games are presented. At no times are rival players referred to as opponents. More importantly, success at the games requires players to score higher than a target score. It is emphasised that beating the other player’s score does not count, simply beating the target score. It is possible, therefore, for all players in the room to win. Finally, it is emphasised that talking with the other player and discussing the game is entirely acceptable.

The players, although they are playing in groups of three, all broadly follow the same strategies and styles of play. Every group begins by selecting the counters at random, effectively turning the game into a game of chance. Discussion and co-operation simply involve devising sensible or effective ways of doing this. Some choose their own counters, others offer closed hands to be selected by the other player. While the target score is low (for example, to score 12 over 10 choices each), there is no incentive to change this randomising approach. Only with a rise in the target (to 17 and finally 20, the maximum that can be scored), do groups begin to think of other approaches. First one group begins to pick green every time, then others, until the idea spreads throughout the class. Most interestingly, the groups who develop this strategy are accused of cheating.
Discussion with classes reveals complex thinking about social mores. Although it is pointed out that the rules do not preclude choosing a particular colour every time, most people are adamant that this is cheating. This view seems to be based on the following arguments which individuals arrive at independently but which become, by representability, a view held by all in the group. Each individual seems to argue as follows:

1. *This is a game, and my experience of games like Snakes and Ladders is that they are a matter of luck. Therefore, to succeed at these games involves luck - picking at random.*

2. *The aim of a game is to score more than the other person.*

3. *Games are fun. To play in other ways would be boring.*

4. *I look after myself. In other words, the players do not understand that giving away points, if done mutually, is the equivalent of gaining points.*

So overlaying the formal rules of the game are a set of informal rules. Just because they are informal, they are no less valid or valued than formal rules. They are quickly formed, universal, binding and long-lasting. Anyone who contravenes them is as much a cheat as if he or she contravened the formal ones. Groups that break these rules are reviled. Interestingly, although players showed little inclination to co-operate in playing the game, the players as individuals quickly reached an informal tacit agreement over what constituted fair play and what didn't. My support of the 'cheats' reduced the resentment they suffered, but it is a moot point as to whether they actually acted rationally by defecting from the group consensus on rules. Certainly, in the scores they obtained they acted rationally to a game theoretician. In the scorn they called upon themselves, they certainly did not act in their best interests.
LESSON 3 - TACIT AGREEMENT AND INFORMAL RULES

The lesson consists of a number of activities which demonstrate that, in the vast majority of cases, groups function well because an individual's choice is reflected, in aggregation, by the group choice. In other words, representability allows individuals to be confident in their choice. Problems arise when representability predicts a group outcome which is not in the selfish interest of the individual: when the individual has to make a choice contrary to the group.

There are several activities which illustrate tacit agreement. They are based on those suggested on the seminal work on this subject, *The Strategy of Conflict* by Thomas Schelling. During all of these activities discussion is not allowed.

1. *Think of a playing card and write it down. You win if you pick the card that most others in the group pick.*

   The majority of players select the Ace of Spades. In Schelling's terms, this is an obvious choice which, without other information, the group will opt for. If the activity is repeated, even more choose the Ace of Spades, because the first game has made it even more obvious.

2. *Look at the following list of drinks. You win if you select from it the drink that most others select:*

   Tea
   Coffee
   Coca Cola
   Water
   Lemonade
   Milk
The overwhelming choice is Coca Cola. The group have little difficulty in arriving at this tacit agreement, in exactly the way they arrived at the Ace of Spades in the previous game.

3. Look at the list of drinks again. This time you win if you are the only person in the group to select that drink.

This is an activity not suggested by Schelling, because it cannot be successfully negotiated by tacit agreement on a natural choice. In fact, this scenario provides no natural choice. First, it can be argued that Coca Cola can be ignored because it is very popular. But what is the least popular. Water? Tea? Worse, if a player, picks this, then won't everyone be picking the least popular as well? So may not Coca Cola be the best choice after all? But others will think that and pick Coca Cola. For the first time members are asked to make choices which are contrary to the group choice. In other words, players are in an incipient Prisoner's Dilemma. They oscillate between the obvious choice and the unobvious choice because they are not sure whether others will make the obvious or unobvious choice. They are using their representability to define a choice which they then try to deny. But in denying it, are they not simply being representative at another level and so condemning themselves to choose what the others choose? No matter how often the game is played, a settled choice is not made: indeed it cannot be, because this very choice condemns the chooser to lose.

4. Write a number between 1 and 10. You obtain those many points providing no one in the class writes a lower.

This is the Prisoner's Dilemma in another form. Co-operation within the class would lead to a high score for everyone. Invariably, however, several players select 1 and so obtain an inferior payoff. If the rules are relaxed to allow any number to be written and to allow group discussion beforehand, stranger results obtain. Even though individuals urge the
group to write a given high number so everyone wins, players still persist in writing low numbers. Most perversely, every time I have played this game, players have actually written negative numbers, effectively losing points! So great is the urge to win, players will actually handicap themselves in order to do so. The principal problem still seems to be that players view these games as parlour games, where defeating the other players is of paramount importance, even at the expense of a personal loss.

LESSON 4 - ZERO-SUM GAMES AND NON-ZERO-SUM GAMES

The lesson consists of a multi-person game based on selections from a Prisoner's Dilemma matrix. The technique used was suggested by Douglas Hofstadter in his article on the Prisoner's Dilemma in Scientific American, later reprinted in his Metamagical Themas. Simply, using a standard prisoner's dilemma matrix, each player chooses from row and scores the payoff from the matrix multiplied by the numbers making each choice in column. The following matrix, for example, could be used:

<table>
<thead>
<tr>
<th></th>
<th>Choice A</th>
<th>Choice B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice A</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Choice B</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Suppose that 6 players pick A and 7 players pick B, then those picking A will receive $6 \times 3 + 7 \times 0 = 18$, whilst those picking B will receive $6 \times 5 + 7 \times 1 = 37$. Therefore, as with all Prisoner's Dilemma games, the highest scores are made by a few defectors free-riding on a large number of co-operators. Large scale defection leads to scores lower than can be made from large scale co-operation. But the move to large scale co-operation again paves the way for free-riding by a few defectors.

As with the individual-player games described earlier, an external target is set, the achieving of which is deemed as a win. Again, therefore, there is no value gained from
simply beating an opponent or obtaining the highest score in the group. The payoffs obviously vary with the numbers in the group, and the balance of co-operation-defection choices. For example, and using the above matrix, payoffs for a group of 15, for each round of choices would be:

| Cooperators | 15 | 14 | 13 | 12 | 11 | 10 | 9  | 8  | 7  | 6  | 5  | 4  | 3  | 2  | 1  | 0  |
|-------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Payoff      | 45 | 42 | 39 | 36 | 33 | 30 | 27 | 24 | 21 | 18 | 15 | 12 | 9  | 6  | 3  | -  |
| Defectors   | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 |
| Payoff      | -  | 71 | 67 | 63 | 59 | 55 | 51 | 47 | 43 | 39 | 35 | 31 | 27 | 23 | 19 | 15 |

The game is played over 20 choices, and the target is calculated so that up to 3 defections could consistently occur and still allow the co-operators to gain sufficient points to win. In the above game, therefore, a target of $36 \times 20 = 720$ would be set. This was an arrangement I had to introduce since all groups, even before the game started, indicated certain players who would never co-operate. These nearly all proved to be consistent defectors, but so did several others, including, in many cases, the accusers! This supports the contention outlined in Chapter 3 that people with a predisposition to defect will often see this trait in others. To win in numbers, therefore, a group had to remain cohesive and committed to co-operation. As each set of choices was made, a running total was kept on a score sheet, showing the numbers of co-operators and defectors in each round, and the scores they achieved. So although the players were not shown a full table of payoffs, they were soon able to work it out. Further, they were able to work out the total points of someone who had, for example, defected in every round. In this way, they learned about the difficulties implicit in the game as they went along, difficulties which were not evident in the simple two-by-two matrix. My role was to expedite this learning process. This meant I often seemed to act as devil's advocate, much to the chagrin of the players. Where co-operation occurred, for example, I would point out the advantage to be gained from defection. Where defection occurred, I would point out that it was scoring less well than full co-operation. In the payoff table above,
for instance, it can be seen that eight defectors fair less well than full co-operation. This point was most upsetting, since all groups began the game with full or nearly full co-operation which began to degenerate over time. I would also point out the rewards that certain consistent defectors were gaining on the backs of co-operators. This often led, paradoxically, to classes co-operating in defection, to reduce the payoff of the consistent defectors. At all times full discussion was encouraged. Periodically I would stop the game to summarise scores, and request questions or comments. Often individual players would stand up to make impassioned pleas for co-operation, or propose complex rounds of co-operation and defection in order to maximise everyone's payoff. Suggestions were invariably ignored. At this time, players were suddenly confronted with the Sorites Paradox. They become rapidly frustrated with their own inability to influence the group. This is not just an inability to influence by the weight of their particular choice, although this is certainly evident. More obviously, they were unable to influence others by word, action or debate. The passion generated in the activity was largely produced by individuals being unable to make others see what they thought was obvious and sensible.

In a series of eight games with eight different classes, only two classes produced individuals who achieved the target. This was the case even though winners were offered a Mars Bar as a prize. Invariably, groups started with good intentions, but the urge to defect on the part of some players gradually became overwhelming. This urge was partly based on the game, which provided a points incentive for defection. Also, I believe that several players just wanted to act against the crowd. One or two of these inveterate defectors in fact began to co-operate once everyone else defected. As later rounds of the game arrived and most players had lost any chance of winning, they simply aimed to spoil the chances of those that might. Generally, the session ended with a plea for me to restore order. Given anarchy, bad faith and disappointment, the result seems to be a recourse to coercion and externally enforced rules. Of the two classes that did produce winners, only one produced a total co-operative choice from first to last. This class had fared particularly poorly in the counters game, a fact I was not slow to point out. I
suspect, therefore, that they were out to beat me in this game. In other words, they treated this game as a straight zero-sum game between them and myself. And it cost me twenty four Mars Bars to realise that they were playing a different game to the one I intended. I think it is all too easy to assume that games and simulations are perceived by the recipients in the same light as we, the designers, offer them. I have tried to show in this whole section on gaming that this is often not the case. Especially with abstract games which have no context with which to interpret them, the players often bring their own context to bear. Players, consequently, play with several different agendas, and each expect different outcomes from the game.

Discussion after the game is centred on two issues. First, the distinction is drawn between the zero-sum game, where if one player wins the other loses, and the non-zero-sum games, where everyone can win or everyone can lose. A good example of a non-zero-sum 'game' is an exam with a pass mark. In most exams, if one person passes, this doesn't mean another does not. Providing they both achieve the pass mark, then they both pass. In fact, non-zero-sum games are much more reflective of the situations one meets in life than zero-sum games. Unfortunately, just as everyone can succeed, so everyone can fail, and non-zero-sum games can tend to generate Prisoner's Dilemma-type situations. The second issue considered is the problems that games of this type can cause. The lesson itself provides ample material for discussion, and parallels to real world situations can be drawn.

LESSON 5 - OTHER GAMES

An interesting feature of non-zero-sum games in the matrix form is that they can be easily manipulated to produce new games. Although the production of new games with meaning is difficult, since it is difficult to perceive the dilemmas they will produce without playing them, pupils can be invited to invent their own. A particularly interesting activity is to ask pupils to devise a game that will not cause a problem. A popular answer
is to take the Prisoner's Dilemma and swap the two left-hand numbers of the matrix. The following is the result shown in ordinal form:

<table>
<thead>
<tr>
<th></th>
<th>Choice A</th>
<th>Choice B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice A</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Choice B</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

A reference to the previous chapter reveals that this in fact is the Assurance Game. It will be remembered that here everyone benefits most from co-operation. Isolated defections do not fair well. Played as a multi-player game using this matrix, however, if more than four players defect, they actually do better than the co-operators. Nevertheless, as with all the dilemma games we are considering, defection in terms of points scored does worse than co-operation. Like the Prisoner's Dilemma, the Assurance game becomes a problem when individuals try to score more points than others, rather than being content with same score as others which co-operation achieves. The Chicken Game also obtains of modelling by the previously described matrix, and I have also used this with classes (no pupil as yet has generated it him or herself.)

<table>
<thead>
<tr>
<th></th>
<th>Choice A</th>
<th>Choice B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice A</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Choice B</td>
<td>3</td>
<td>-5</td>
</tr>
</tbody>
</table>

Here, a choice of defection has disastrous consequences. The advantage for a few defectors, however, is great. These defectors are playing chicken with the rest of the class, and basing their defection on the fact that others will not dare to defect also. The Altruist's Dilemma can be modelled differently to the way described earlier to make it work as a group game. The following matrix encapsulates the Altruist's Dilemma:
Paradoxes in Decision-making

Paradox in Practice

It can be seen that the best payoff is obtained for a defection and indeed for the whole group if everyone but one co-operates. Co-operation, however, has a meagre reward. Therefore, acts of co-operation to ensure the profit of others are, in this game, altruistic.

Again, the games are played to a specified target figure over a pre-determined series of goes. As with the Prisoner's Dilemma, great difficulty is encountered in trying to play them successfully, since the desire to do better than others, or not let others do better than oneself, seems to predominate. In discussion, pupil's continuously return to this need to outdo or not to be outdone. The idea of general improvement by co-operation, rather than on the basis of one's rugged individual strengths, seems an alien one indeed. Co-operation can be induced by reward, but this co-operation also has at the back of it overt threats from dominant members of the group who want the reward. Appealing to better nature, therefore, may be a barren strategy.

PURPOSE, AIMS AND OUTCOMES - RATIONAL ACTION AND BY-PRODUCTS

On, on, you madman, drive
Over your savage Alps, to thrill young schoolboys
And supply a fitting theme for speech-day recitations!
Juvenal, *Satire X* (Hannibal)

I wish to consider in this section some of the processes which shape the purposes and activities of schools, and how, as suggested in the epigraph, these purposes and actions,
like those of Hannibal's, can have unintended outcomes. I hope to show that, in education, we have assumed a reliance on game theoretic styles of thought at the expense of others, and in using this style of thinking, we have also enabled the paradoxes involved with it. It is not my intention to consider all the aspects of this large topic, but to illustrate my arguments by a consideration of three aspects. The first is process of turning aims into curricular objectives. The game theoretic style here seems to lead to the paradoxes of willing what cannot be willed and the achievement of by-products. The second aspect considers the role of the management team in directing policy and development. Here we will encounter problems caused by level shifts, an issue which will be explored more fully in the last section of this chapter. The third looks at the assumptions made about the use of information and the fact that, contrary to game theoretical suggestions, information is not particularly valued or prized. As I consider each aspect, I will also try to suggest alternative approaches which might reduce the effects of paradox.

From aims to curriculum

For the purposes of this discussion, I shall assume that 'curriculum' is not a simple list of subjects studied, but a series of related activities which enables successful learning to take place. I will look in particular at ways of defining a subject, at assessment, and at the 'hidden' curriculum. To open this discussion I will consider an article in Volume 27 of the Journal of Philosophy of Education, Indirect Learning and the Aims-Curricula Fallacy, by J E Adler. His contention is twofold. First he argues that aims are not easily transferred into curriculum outcomes because they are not conceived or worded in ways that make this possible. Second, and this is reflected in the title of the paper, he argues that too often the by-products of learning are ignored when we try to specify ways in which the curriculum can be enabled from its aims.
If we advocate the construction of aims which are readily converted into curriculum outcomes, it is easy to argue that these aims would emerge as simplistic, narrow or severely practical. Adler argues that this should not be the case. They should both contain a stringent and cohesive philosophy, and be expressed in a way which makes their enablement evident. We could, for example, produce an aim which says:

*Children should be encouraged to think for themselves.*

The problem with this is that it really makes no statement about how it can be achieved. The terms it uses, like *think* are much too vague in this context. An alternative proposed by Adler is:

*We need to base our model of the critical thinker not on the domain-bound individual with subject specific skills, but on the disciplined generalist.*

Whether we agree with the statement or not, at least it suggests which way we ought to go. In fact, Adler suggests that good aims make a seamless transition from "aim" to "ought". In fact, the next sentence of the above document goes on to say:

*This means that we ought to encourage the student as soon as possible to recognise that in virtually every area of our lives, cutting across categories every which way, there are multiple conflicting viewpoints and theories vying for our allegiance, the possible truth of virtually all of which calls for shifts in our global perception.*

The whole is obviously not for the faint-hearted, and is certainly not simplistic or severely practical.

The obvious move from an aim such as that above is to map out some desiderata for the actual curriculum, and then to turn these into a set of objectives and actions. I wish to
consider this procedure. I would argue that too often it is exemplified first by an emphasis on reductionism - by an attempt to define subject areas or curriculum items in terms of a series of chunks or modules - with the implication that the subject is mastered or the curriculum defined if the chunks are mastered or defined. And second, there is a belief in formalism: the desire to make all actions, activities and outcomes explicit and, if possible, measurable. Both are examples of narrow or game theoretic rationality, and as such are prone to paradox - the problems in this instance being caused by level-shifts and by-products. I shall illustrate these by considering a typical reductionist curriculum, and a document which translates a set of assessment aims into practice. I shall retain the theme of assessment and go on to discuss alternative ways in which the aims might have been enabled without recourse to the narrow rationality of formalism and reductionism.

I will begin with a specimen curriculum, taken from the School Management Competences Project.
The School Management Standards

Key Purpose
Create maintain and develop the conditions which enable pupils and teachers to achieve effective learning

Key Role A
Manage Policy
Review, develop and obtain support for school aims, policies and objectives

Key Role B
Manage Learning
Review, develop and implement strategies to support the learning of pupils

Key Role C
Manage People
Develop and deploy human resources to implement strategy and policy

Key Role D
Manage Resources
Develop, maintain, monitor and control financial and physical resources

Units

A1 Review develop and present school aims, policies and objectives
A2 Develop supportive relationships with pupils, staff, parents, governors and community
B1 Review, develop and implement means for supporting pupils' learning

Elements

A1.1 Identify opportunities and constraints on aims, policies and objectives
A1.2 Encourage discussion of school aims, policies and objectives
A1.3 Develop school aims, policies and objectives
A1.4 Seek agreement and disseminate school aims, policies and objectives
A1.5 Evaluate and review effectiveness of school aims, policies and objectives

A2.1 Identify problems and opportunities
A2.2 Develop and maintain positive relationships with interested parties
A2.3 Promote the school and its services
A2.4 Recruit pupils and operate admissions policy
A2.5 Evaluate and review relationships and promotion of school

B1.1 Identify learning needs of individuals and groups of pupils
B1.2 Review, develop and agree means of planning and supporting learning
B1.3 Implement learning program
The syllabus is given a summary statement which is called a Key Purpose. This is comprised of Key Roles, each of which is subdivided into Units, which are finally divided into Elements. A few examples only are given of units and elements: there are obviously several more of these. The hierarchical nature of the divisions is emphasised by the tree-diagram structure and the use of text decreasing in point size. There is an overt attempt to break down the subject or activity into a series of discrete chunks. The assumption here is that by achieving the chunks, in this case the elements, we have in fact achieved the whole and become successful school managers. I think there is no valid basis for this assumption. I would argue that many outcomes of action are systemic: that is they are a product of the joining together of things which were separate before. I can describe a Chippendale chair in terms of its parts - arms, legs, joints and so on. I cannot describe its quality or beauty in this way. These are a product of synthesis, of the bringing together of things. If we do not view our subject or activity holistically from time to time, then we cannot say we have fully apprehended it. In a history lesson, I might teach some particular skill, and bend the material I use towards that. In another lesson, however, I might just do some history, and pick off a whole range of disparate skills in one go. It is this that makes any subject what it is, a discipline in its own right. Subjects are not simply the vehicle for teaching a range of similar, cross-curricular skills. I am not arguing against the teaching of skills, I am simply saying that the blending of these skills in particular ways is as important as teaching the skills themselves. With a reductionist curriculum we are in danger of falling into the paradox of wanting the by-product directly. If we want something which is essentially collateral to learning, such a ethos, quality, subject-discipline, love-of-subject or love-of-learning, then we cannot obtain it by direct definition and inclusion. These items are outcomes of learning, not a direct object. I wonder if the curriculum specified above in any way captures these concepts.

We can take the issue of by-products and collateral items a stage further by considering ethos and the hidden curriculum. I shall have more to say on this area in the next section,
where I consider the formation of rules. Some discussion is relevant here, however. I propose in the next section a typical set of school rules, one of which is:

*Never threaten or use violence*

It is interesting to speculate upon the appropriateness of this rule. After all, it is a formal rule, explicitly devised and expressed, attempting to make some statement about the ethos of the school. Rules work often because they threaten. The threat is implied, because the rule does not actually state:

*Never threaten or use violence or you will be punished*

but the implied threat is there nevertheless. So the non-threatening ethos is to be achieved by threat! Nor does the dilemma end with this rule. The proposal of any formal rules implies some threat, so any rule system suggests threat somewhere within the ethos. This presumably means that a non-threatening ethos can only be achieved in institutions devoid of rules. This is a paradoxical conclusion, since an institution based on no rules is probably the most threatening of all. The paradox here seems to stem from a rule which makes a formal statement about the ethos which is essentially an informal, non-negotiated construct; a complex product of rules, standards and people. By attempting to make it explicit and formal we make it slip out of our grasp. We cannot actively and directly want an ethos. It is one of these collateral aspects of learning, like enjoyment or love-of-learning, which come out of the collecting together of the parts of the curriculum and institution. These are the products of synthesis and synergy. To try to create non-violence by formal ruling is like trying to create enjoyment by formal ruling. If we promulgated a rule:

*You will enjoy your lessons*
we would consider it absurd. I would suggest, however, that it is no more absurd than
the rule about threatening violence. In the end, enjoyment, ethos, love-of-learning, and
so on, will emerge from how the curriculum actualises itself. I am not contending that
rules can play no part in this. For instance rules which say:

*Follow the dress code*

or

*Be polite at all times*

contribute towards the ethos we require. In fact, we obtain the ethos through rules like
this. They are the third agents which generate the ethos. What we cannot do is generate
the ethos directly.

Adler, in the paper cited above, bemoans the fact that it has now become fashionable to
decry the indirect outcomes of learning, to see them as haphazard, difficult to quantify
and monitor, and impossible to plan for in any kind of systematic way. In other words,
they are mistrusted because they have a life of their own outside the narrowly rational.
Rather than accept this verdict, Adler urges us to explore and value the part the hidden
curriculum might play in enriching the whole curriculum. He makes several points about
this issue. First, he argues that the generation of by-products is unavoidable, and cannot
be inhibited anyway. Instead they should be planned for and enabled. Second, by-
products come as a free extra, a corollary of the direct and explicit curriculum. They
should be encouraged because they genuinely are something for nothing. Third, he
argues, as I have argued in the previous two paragraphs, that several concepts, beliefs,
attitudes or feelings can only be taught or developed indirectly. Fourth, he believes that
indirect learning allows progress where direct approaches, particularly the 'more of'
approach, may not. For example, failure at history will not be cured by more history if

154
the root cause is lack of motivation or reduction in self esteem, which are essentially collateral to the study of the subject. The chance to succeed at anything might be more appropriate than the chance to fail more often at history. In the light of what I have said in this section, I find his arguments particularly persuasive, especially since they point the fact that a situation which I consider paradoxical - the creation of by-products - needs to be carefully considered and managed, since it cannot be ignored. Its immanence within the curriculum is underscored by the OFSTED inspection documentation. Part 7 of the Handbook for the Inspection of Schools, the part dealing with the recording of evidence, lists the aspects of a pupils personal development which will be assessed. These include

5.1.1 Spiritual Education
5.1.2 Moral Education
5.1.3 Social Education
5.1.4 Cultural Education
5.1.5 Relationships at school

Each is to be assessed on a scale of 1 to 7, from excellent to very poor. A footnote says that 'Education' refers to the school's provision and the pupils' response. It is difficult to see how any of these issues can be dealt with by direct means. Of course, some may fall within a PSE or RE lesson, but most I feel are a product of the ethos the school generates: no school, for instance, would give a discreet timetable slot to 'moral education' or 'relationships', and I am sure OFSTED do not expect this. Instead, I feel that the inspectorate view these items as collateral outcomes or by-products of the learning process. The use of the footnoted phrase 'school's provision', however, shows equally that they should be positively sought and developed. They are not haphazard or intangible, but are as right and proper an outcome of the curriculum as the more obviously direct parts.
The role of management in policy development

Although we might consider the management of an issue, too often do we ignore management as an issue in its own right. In other words, we fail to adopt a meta-position and consider, first, our assumptions about management before we consider our assumptions about particular issues to be managed. I will illustrate what I mean with reference to a document which was presented during a course I attended on assessment, and to the question of assessment more generally.

**MANAGEMENT OF THE PROCESS**

- Where are we now? Audit
- What are our priorities?
- What are our needs? - Time
  - Support
  - Resources
  - Expertise
- What are our staff development needs?
- What do we need to do?

▼

**ACTION PLAN**

<table>
<thead>
<tr>
<th>What?</th>
<th>Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>How?</td>
<td>Strategies</td>
</tr>
<tr>
<td>Who?</td>
<td>Responsibilities</td>
</tr>
<tr>
<td>Is it working?</td>
<td>Evaluation/Review</td>
</tr>
<tr>
<td>Time Scale</td>
<td>Realistic - when in place?</td>
</tr>
</tbody>
</table>
Humberside LEA guidelines and National Curriculum policy converge on the need for the thoroughgoing assessment of pupils' work. What strategies can be employed to ensure that meaningful assessment takes place? The above document on assessment makes several suggestions. These include the need for schools to undertake an audit in order to ascertain the present status of assessment, the need to prioritise developments, to investigate staff development needs, and to form an action plan. These activities can be viewed either as rationally implied ends from the need to assess, or the rational means with which to achieve the end of assessment. In either case, there is a logical means/ends connection between assessment and the activities. The activities themselves, however, lean heavily on definitions of rationality which are associated with game theory. First, they assume the generation of a number of alternative strategies: it is impossible to 'prioritise' a single strategy. Second, they assume that all the parameters are variable, to be moulded according to the decided strategy: the reference to staff development shows that it is considered that even the behaviour of people can be varied in a desired, uniform direction. Furthermore, this manipulation and bending of people is seen as a 'practical' means of implementation - if the plan could founder on present practice or attitude then INSET will put this right. Third, the creation of action plans implies a commitment to future action without any commitment of future managers: action planning contains no pre-commitment. As we have seen, this belief in optimisation, strategic planning based on manipulation of variables, consistency, and the untrammeled extrapolation of the present into the future are all elements of the game theoretic approach. Any move to implement such a programme pre-supposes belief in a rationality of this kind. The whole process leaves nothing to chance. Instead, all parts of the plan, which is based on appropriate consultation and amendment, and the roles within it, are explicit and formal. OFSTED documentation supports this view. The technical paper on assessment presupposes a top-down management approach, beginning with reasoned policy permeating the succeeding layers of the system and eventually reaching the classroom:
25. Strong management at departmental and school level will ensure that a coherent policy is implemented that pulls together various strands......This will encourage co-operation among teachers and consistency between subjects. The quality of policy and planning will be apparent......in the extent to which there is good classroom organisation.

The ambiguity in the second sentence is particularly interesting. Does the use of 'will' imply an imperative - management and policy will do this or that - or does it imply that strong management and coherent policy will, by its very existence, lead to co-operation among teachers? If the second interpretation is the one intended, then it should be read alongside the list of evaluation criteria for management and administration contained in the Framework for the Inspection of Schools section, where both leadership and ethos figure. Does this mean that management leads by providing policy, and that the ethos which ensues is one of co-operation? I think this view of the relationship between management, leadership and policy is untenable, since the provision of policy does not imply co-operation. Neither is the provision of policy implied in leadership. To be fair to the OFSTED documentation, it nowhere states that managers should originate policy. It argues consistently that policies are a sine qua non of good management: it would seem impossible to manage entirely by adhocracy. I would argue further that it is the job of management to enact and promulgate policy. I believe, however, that it is not the sole prerogative of management to originate policy, or to interpret and dictate its particular implementation within the school. Unfortunately, I feel the tenor of the documentation I have so far considered not only assumes that management should originate policy, but also that the body politic will benefit from this imposition. Worse, there seems to be no need or perceived responsibility to justify this assumption. This omission is made more glaring by the fact that the documents are often excellent at defining issues, and exemplifying and justifying good practice. I am sure, for example, that a parent would like to receive the kind of school report described in the OFSTED technical paper. However, if we go beyond the issue to the assumptions I have outlined about how issues
should be managed, there is no justification. There is no statement which begins: 'We advocate a top-down, management inspired approach because......' This statement is never made because the efficacy of this style is assumed. I believe we cannot make this assumption. Instead I feel that as an approach it is at times useful, at times weak.

If we consider the assessment document and its associated issues of management from the point of view of a broader rationality, for example, we obtain a different view of rational ends. First, we could argue that teachers have ends: to survive, to actualise themselves, to increase pleasure or avoid pain, to impart knowledge, to improve their classes or the world. These are vague, non-game-theoretic ends, but are ends nevertheless, and could all be roughly categorised as being psychologically comfortable. Second, teachers are the people who are the actual interface between policy and practice: they will have to actually do the assessing. In that case, it may be appropriate to consider these as the centre of activity, and instead look at strategies for bringing assessment on to their personal agendas rather than strategies for implementing assessment to which they must conform. This different approach to strategic planning would require not only a consideration of the sorts of conditions in which teachers work, but also the sorts of people they are. Any strategies would be hit and miss, and the outcome in terms of assessment would vary from teacher to teacher. However, this more haphazard approach, depending upon a variety of activities and time scales, and heavily dependant upon heuristic procedures, might succeed. Third, the game theoretic approach sees INSET as a catch-all device, which will fill the gaps which occur between theory and practice. When the infeasibility of a theory is suggested, I have often heard in defence the use of the phrase 'well there is an extensive INSET implication here' - or something similar. This statement seems clearly to be indicating that INSET exists to train the teacher, to make him or her more useful or amenable to the needs of the institution: to socially engineer. This stands in contradistinction to the aims of the Teacher Appraisal Scheme, which state that teachers should use INSET to develop as people, professionals
and as participants in a vocation and career structure. This second approach to INSET seems the more humane and, dare I say it, rational.

The game theoretic approach seems not to have varied in its approach to assessment since the early eighties and seems not to have succeeded: if it had done its job it would no longer be needed. Recently, assessment has become part of the National Curriculum, and its instigation, in some senses, has now a legal basis. This legislative necessity - a needs drive - may rearrange the psychological balance of individual teachers, and the drive for psychological comfort may achieve what years of game theoretic problem solving has failed to do. I shall illustrate the appropriateness of needs driven action with an actual example from my own school. I shall go on to use it as the model of an alternative approach to relating the curriculum to aims.

To start this discussion, I would like to consider in more detail how we solve problems. It has been my contention throughout this work that we do not problem solve in as thoroughly a rational way as game theorists would like to suggest. A method I have used to demonstrate this contention both to teachers and pupils is to play the XI game. This is loosely based on a game described by Douglas Hofstadter in Gödel, Escher, Bach: an Eternal Golden Braid. The games consists of the manipulation of the letters X and I in groups. There are four rules which dictate how these groups can be changed.

1. Any group ending in X can have an I added to its right hand end. E.g. XX \(\Rightarrow\) XXI

2. Any 3 consecutive Xs can be replaced by an I. E.g. XIXXXIXX \(\Rightarrow\) XIIIXX

3. Any group can be added to the right hand end of itself. E.g. XIXI \(\Rightarrow\) XIXIXIXI

4. Any 2 consecutive I's can be deleted. E.g. XIXI \(\Rightarrow\) XX
The challenge is to start with a single $X$ (which for the purpose of the rules can be treated as a group) and, after applying any appropriate rules, end up with a single $I$.

A typical game might go as follows:

<table>
<thead>
<tr>
<th>Move 1</th>
<th>$X$</th>
<th>(Starting position)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move 2</td>
<td>$XI$</td>
<td>(Rule 3)</td>
</tr>
<tr>
<td>Move 3</td>
<td>$XIXI$</td>
<td>(Rule 1)</td>
</tr>
</tbody>
</table>

and so on.

I have played this game with hundreds of people of all ages, and their method of approach is always the same. First, they begin to manipulate the letters and rules. Sometimes individuals are confused by the rules, and need further explanation. Eventually, when everyone has grasped the rules, a period of manipulation takes place. Second, at some point, certain individuals begin to assume that the task is impossible. Third, and finally, someone in the group demonstrates that the task is impossible. They point out that the way of increasing $X$s is by doubling, but that they are replaced by $I$s in threes. Since both are prime numbers, they have no common factors. So there is not a doubling of $X$s that can be exactly remove them all by taking them away in threes. So an $I$ can never be left on its own. Often the proof is not stated as succinctly as this, but many gain an intuitive grasp of it. It is interesting to note, however, that an inspection of the rules is quite sufficient to draw the conclusion of impossibility: I have never known anyone do this, however. Instead, a period of experimentation and manipulation seems mandatory. Further, players subtly change the level of their activity without realising it. They start by trying to solve the problem of turning $X$ into $I$. At some point, however, they turn instead to the problem of whether this is possible. In other words, they move from solving the problem to solving the problem about the problem. Only when they
have reached this point, and are, in fact, manipulating the symbols for a different reason than originally, do they conclude the impossibility. Even then, I feel the search for impossibility is not based on the knowledge of impossibility, but simply on the belief that the amount of manipulation already undertaken would have produced a solution if there had been one. In other words, the problem is solved by a combination of experimentation, level shifting, shrewdness, and a growing boredom with the original problem. Indeed, as Karl Sigismund said in the quotation used in the first paragraph of this chapter, we learn by playing with things. I would argue that this is not an uncommon approach to problem solving. Certainly there is no evidence of solution by deduction even though, as I have pointed out, all the information required for a solution without any need to play or manipulate is contained within the rules. In reflecting upon this game, as well as in discussion with players, I have drawn several conclusions. First, the way we attack a problem is anything but optimal, in that we work with only a sketchy grasp of the parameters and with an equally weak grasp of our route to the solution. I feel that this unorganised or ad hoc approach to problem solving, is closer to people's actual strategy than that of the game theoreticians, who would have us believe that problem-solvers embark on their quest having ensured beforehand that they have as sound a grasp as possible of all aspects of the problem. Second, people succeed where they have the power to find the answer. This empowerment includes a clear statement of the problem with no deliberate concealment of information, the provision of resources and support, and a clear reason for involvement - a need to succeed. The need to succeed may be generated by the individuals themselves, but may equally be based on a legal or institutional requirement to complete a task. Either way makes success a psychologically desirable thing. I believe that this style of problem solving, combining ad hoc approaches, empowerment, and the need to reduce psychological dissonance, can often succeed.

I will try to illustrate this contention from an actual management scenario in which I was recently involved. Although it deals with IT solutions to particular problems, I have no
intention of advocating these solutions or suggesting that these are general problems. It is rather my intention to exemplify this non-formal approach to decision making, and to strengthen my case that management should try more regularly to support and encourage this as a procedure.

The changes in reporting procedures which began with Circular 1494 have made it very difficult for schools to be flexible in their reporting times. The need to report a pupil's progress at least yearly, coupled with the need to report on the ends of key stages at their ends, forced us to place nearly all our reporting at the end of the academic year. Further, we were obliged to comment on progress in some detail within each subject, which made short or single line comments impossible. Faced with this situation which the staff did not relish, I obtained a copy of a profiling package for the computer network which made possible the writing of reports by computer from a comment bank. The package became available about a month before the first reports were due to be started. In the six weeks following, we had to complete four years of reports (years 7 to 10), each year group consisting of approximately two hundred and forty pupils, taking between eight and twelve subjects. Faced with such a daunting task, several staff, either as individuals or comprising faculty or subject groups, asked if they could attempt the computer solution. Even though none of us knew the package, there was no management structure to systematise the task, and no dedicated hardware, we decided, nevertheless, to try. During the activity, which was essentially grass-roots led, and which had no formal management sanction other than my personal support as system manager, I observed the development of several interesting phenomena. First, ideas spread rapidly by personal contact, or 'buckshotting' through the group. Spare machines were commandeered as they became available, so that the package could be experimented with. This playing created experts in certain aspects of the package who became known and sought after. Second, tacit agreements in the form of 'time-sharing' arrangements became common, as particular groups were allowed access to the machinery at certain times. More broadly, we tacitly agreed on the worthiness of what we were doing. We shared a belief in the IT solution to
the problem, and commitment to self-help and mutual support. Third, both management of the issue and management of its management occurred simultaneously but discretely. That is to say, the exploration and development of the package went hand in hand with the emergence of systems to support and enable the information flow. In this way, the task of writing reports was not only accomplished, but also a system for managing the process in the future was developed. Fourth, networking as a means of proceeding was particularly in evidence. *Ad hoc* meetings occurred often, and members of the informal group of report writers would drift to the room where the computers were to see what was happening or if anything new or interesting had been discovered. Gradually, these disparate activities and discoveries formed into a single structure or system which allowed the task to be accomplished. My role was of importance here, since I acted as a clearing house for information, taking questions and suggesting contacts as well as proposing ways forward. I had the overview, and in this sense only did I 'manage' the network; both the network of computers and the network of staff which the computer network supported, enabled and empowered. This process of networking also allowed individuals to support each other, an activity I came to call *peer* or *mutual* empowerment. Last, and most interesting, we achieved far more than we set out to, in areas which we would not have predicted. Not only did our personal expertise increase, but so did the database with which we worked. In particular, we made some valuable discoveries about the manipulation of comment banks to create correctly parsed but varied sentences. Although we wanted to complete our reports, the management systems we developed to enable this were an unforeseen bonus. Of most value was the work we did in linking the output of the package to a word processor rather than simply to a printer. In this way, we were able to greatly improve the presentation of the product, an outcome again which we had not intended to explore when we began.

Eventually, about 20% of all reports were written using the package, which it must be remembered was non-existent in school up to a month before the reporting cycle began. The concomitant management systems and databases, combined with general enthusiasm
for the quality of reports and their ease of production, make the use of the package virtually *de riguer* in future. Given a clear need and a commitment to IT solutions - both of which empowered individuals to complete the task, and a philosophy of self help based on networking or peer-empowerment, the group which pioneered the changes needed no management. Interestingly, the group has now disintegrated. Once the task was accomplished, and even with a general desire amongst us to meet and explore more aspects of the computer system, no further work has been attempted. The need to find new approaches and radical solutions, based, in this instance, on the prospect of writing freehand what might, in some cases, be more than a dozen sets of reports, has vanished with the successful completion of the task. The new approaches and radical solutions remain, however, for future exploitation and development when other needs arise.

When we consider the work of Edward de Bono and other writers on creativity in the next chapter, we will examine their emphasis on the need for restructuring, for putting ideas together in new ways. The exploratory, non-teleological way that grass-roots or bottom-up networking proceeds exemplifies this process. The whole provides ownership and uniqueness. It is kept on course by the need to succeed and the overarching philosophy and clear aims. Other than that, its outcome is unpredictable, a matter of forming larger structures and systems by a procedure akin to natural selection. And these structures and systems are the intangibles, the by-products, whose apprehension gives the whole quality and meaning. It is these that make the whole greater than the sum of its parts. I have suggested above that networking, ownership and dissonance reduction have a part to play in ensuring that the curriculum actualises itself positively. I do not wish to argue for an entire curriculum, or management structure, based on networking or needs led, grass roots development, just its appropriate use. First, such a procedure does not always guarantee that particular obligations are fulfilled. Second, networking takes time and commitment and several activities might be worthy of neither. I shall illustrate these contentions with reference to the statutory requirement to register pupils.
In this instance, we wanted to make afternoon registration arrangements more efficient. In particular, we wanted to abolish afternoon registration in tutor groups, since this involved two movements of the entire school population within the space of five minutes. Again it was decided to seek an IT solution, using an Optical Mark Reader and a computer database which contained details of the afternoon teaching groups. By using this package, we could register the entire school in these: the entire school would go to its lessons after lunch and not need to move again until the end of these lessons. When I set up this system, I consulted very few staff. This was partly because the software dictated certain aspects of the system, so that there was no flexibility there to consult about. Where there was latitude in the system, no one was really interested in discussion. Registration to the staff was a necessary evil, and provided it did not involve them in much extra work, they simply wanted to be told what to do. There was no need to explore the procedures, because the whole thing was just something that had to be done. This should be compared with the IT solution to report writing, which obviously saved a great deal of work and produced an improved result. Further, staff cared about their reports because they made important statements about their pupils and would be read by parents. Registration, on the other hand, was purely an administrative necessity, implying no values and commitment. It was to be dispensed with as quickly as possible so that the business of teaching could proceed. Although I proposed a discussion group, no one took me up on it. I felt quite happy to devise and impose my system. In fact, I feel the staff would have felt that I was derelicting my duty as a manager if had done anything else.

The problems of management outlined in this section seem to reside in the inability of managers to consider first the nature of management before beginning to manage. Rather than applying a particular style to all issues or problems, it would seem sensible first to consider a range of styles and then apply them appropriately to particular issues. There seems to be a need for managers, from time to time to adopt a level shift, and take a meta-management stance. Sometimes the outcome might be to do nothing, but doing
nothing about an issue is still managing it at a meta-management level. At other times the mere oversight, direction and empowerment of a ground swell of activity might be appropriate. Proscriptive, top down action might be necessary, but not always. I have tried to show, however, that this last style of management, a control model which deals with issues, is often seen as the whole end of management. I would argue, conversely, that this is but one style. The heart of management, I believe, lies at the meta-level of management, where the most appropriate of several approaches is selected and then pursued.

**Information**

Much play is made of information transfer, storage and interpretation in schools. At a recent lecture on Special Educational Needs provision, this battery of questions was fired at the managers present:

- how is information from primary schools used in decisions about SEN provision for individual pupils?
- how is information about SEN children communicated to staff?
- how is information about SEN developments communicated to staff?
- how do faculties pass information about SEN children to each other?
- how does the SEN co-ordinator gain information about pupils from faculties?

Each question could be answered in some detail except the first. It seems that although there were some contacts with the primaries, their information seemed not to influence faculty plans at all. Worse, information from the primaries on all the new intake children seemed not to influence planning. I decided that a useful indicator of the influence information from the primaries had on a given secondary school lay in the secondary school’s setting policy generally and its policy in year 7 in particular. If the school did not operate setting at all, or did only at Key Stage 4, then it was of no use to me. The interest lay in schools which setted throughout their age range. I felt it was reasonable to argue that if any such school began setting at the beginning of year 7, then it must be
basing its sets on information received from the primaries. Likewise, I could assume that setting later in the year or at the beginning of year 8 was leaning more heavily on home grown data. Interestingly, no school I polled began setting straight away. This seems a bizarre situation. First, the schools are not philosophically against setting: after all, they set in later years. Second, they cannot complain that there is no information available: the primaries have been assessing their pupils for seven years. This assessment has been both statutory and informal. Pupils exit Key Stage 2 with a set of National Curriculum grades, as well as a whole set of work, presumably marked, in a whole range of subjects. Why, then, is this valid and extensive collection largely ignored?

*Is it not trusted?*

I think not, although this is a difficult issue. I certainly know of one pyramid of schools where strenuous efforts at standardisation have been made between partner primaries, and the secondary they feed. Certainly, all of these schools know, understand and agree on exactly what, for example, a level 3, AT 1 in English means. Still the secondary school does not stream until year 8. More interestingly, every secondary school uses pastoral information from the primaries extensively to arrange its pastoral and tutor groups. It seems that this information is perceived to have far more relevance and use than academic information.

*Is the information not transmitted?*

I know of no primary school which does not send copious amounts of information out with its leavers.

I think that the reason why this information is not acted upon is more subtle. The tacit assumption behind the accumulation and communication of information is that information is good for us, and that the more we have the better we will be. When I
discussed this issue in a previous chapter, however, I suggested that this may not be the case. People treat information with respect, because it can be a dangerous commodity. As we saw, the more we get the less we might want it, and the less we want it the more we might get it. And often the more we get the more confused we get. I believe that teachers see the divide between Key Stage 2 and Key Stage 3 as a hiatus, as a time of fresh beginnings and the putting away of childish things. Children, likewise, think of themselves as going to 'big' school and expect it to be different. The transfer of information includes within it, as more subtle information, the mores and attitudes of another life. This view may not be true, but the hiatus effect seems to be. Very little information seems to permeate the divide. If we can assume this disinterest in information, can we not also assume a disinterest in that style of thinking - the game theoretic - which is information based?

Interestingly, when the setting process finally begins, I know of no school which adopts a top-down approach to it. Invariably, sets are formed on the basis of assessments made by individual teachers, not on the basis of some objective criteria which are applied across the board in an scientific way. The argument against this approach is that the individual teachers know the children best. Here, therefore, the grass roots, not reasoned policy, seems to apply.
STRUCTURE, ETHOS AND RULES - TACIT AGREEMENT, THE SORITES AND MAGICAL THINKING

The most unreasonable things in the world become more reasonable, because of the unruliness of men. What is less reasonable than to choose the eldest son of a queen to rule a state......For whom (alternately) will men choose, as the most virtuous and able? We at once come to blows as each claims to be the most virtuous and able. Let us then attach this quality to something undisputable. This is the king's eldest son. That is clear, and there is no dispute. Reason can do no better, for civil war is the greatest of evils.

Blaise Pascal, Pensées

The work of Schelling mentioned in the first section of this chapter shows the extensive use that we make of tacit agreement in our day-to-day social intercourse. Let me define again Schelling's contention. He maintains, contrary to game theory, that all points of decision are not equally plausible or likely - some are more obvious than others. Where agreement is necessary, these obvious points will be selected even though they may be Pareto inferior. The fact is that any agreement, even though it may not be a particularly efficient, is often better than no agreement, or an agreement that has to be exhaustively calculated for optimality from a whole range of potential agreements. I illustrated this contention in the first section with reference to playing cards, where even out of fifty-two choices it is very easy to find an agreement on one or two without discussion or negotiation. I now wish to discuss this issue in relation to the formulation of rules, both formal and informal.

Formal Rules

No one would deny that rules are essential to the smooth running of organisations, and that it is the job of both management and the managed to ensure the formulation and
enforcement of rules. This is particularly important in schools, where rules are seen as defining the milieux within which successful learning can take place. As I have indicated above, these rules can be both formal and informal. Further, I would argue that their efficacy depends to a large extent on how closely they correspond to tacit agreement points as defined by Schelling. If not, they become prone to the Sorites slippery slope argument. Let me give an example of what I mean by this with reference to a typical set of school rules.

1. *Always be polite*

2. *Be on time for school and lessons*

3. *Remove coats at start of lesson*

4. *Always walk round school*

5. *Never threaten or use violence*

6. *Care for people and property*

7. *Never eat or chew in class*

8. *Bring books and equipment to lessons*

9. *Keep the school clean and tidy*

10. *Always walk on the footpaths*

11. *Keep to the dress code*

12. *Be calm and sensible at all times*

13. *Always do your best*

These rules are designed to appear in classrooms, in information booklets for pupils and parents, staff handbooks and so on. As such, they are typical. It is interesting that they tend to look for obvious points of agreement based on all or nothing at all. NO eating in class. NO violence. Try your best ALWAYS. A rule which said, for example,

*Six children can wear their coats in class*
would by comparison be fraught with danger, because it does not conform to a natural agreement point. The problems with a rule like this are twofold. First, it would be necessary to choose the six children. This would obviously cause difficulties, since it would be vital to negotiate a set of criteria by which the six might be selected. This in turn might fall back on obvious agreement points: the six who have worked hardest, the first six on the register followed by the next six, and so on. The fact is, that apart from encouraging some group decision-making, this rule simply adds complications for no gain.

A second more subtle problem with this rule is that it is unstable. It is particularly prone to the slippery slope Sorites argument described at length in chapter 1. It is to be recalled that this type of argument begins with a premise which although acceptable in a clear area, becomes impossible to deny, and disastrous in its consequences when conditions become vague. In this case, the rule clearly states that six children are to wear coats. It is easy to argue that what applies to six can equally apply to seven - one child more wearing a coat will not make an obvious change in the appearance of the class. Once this is conceded, however, the next coat and the next coat mean that eventually all the class will be wearing coats, even though the addition of one coat, it is agreed, would make no difference at all. Once the slippery slope is entered, the consequences are disastrous. The banning of all coats prevents this impasse. If no coats are to be worn, then the wearing of one coat is not an insignificant act, but a fundamental change in the dress habits of the class. Therefore, the concession that one coat makes no difference need never be made and the slippery slope never entered.

This case is fairly obvious, and such a rule as the one discussed above would never sensibly be devised. The slippery slope, however, is subtle, and is likely to emerge anywhere where there are concepts such as life and short, which are clear in themselves, but are vague in their edges, termination points, or interfaces with other concepts. Let us reconsider our original set of thirteen rules and inspect them for vagueness.
example there is an obvious one - what constitutes a lesson? We would all agree that a
lesson is clearly a place where formal, managed learning takes place. The problem lies in
deciding when a lesson actually begins, so that coats can then be removed. No lesson can
start the moment the lesson signal goes or its appointed time arrives because transfer
between lessons is not instantaneous. Further, some time entering the lesson and
preparing for work must precede the lesson proper. At some point, the preparation time
ends and the lesson begins. At what point does this actually happen? There are several
points at which a lesson could be said to have begun for an individual child: when she
enters the building or environs of the lesson; when she lines up outside the classroom;
when she enters the room; when the whole class is in the room; when the teacher has
taken the register; when the books are handed out; when the teacher starts to describe
the lesson content; when work begins. The difficulty is that there is not necessarily any
formal agreement about this within a school or among a particular set of teachers, and
there is certainly not an obvious tacit agreement point. I have discussed this issue with
many teachers, and find that each has his or her own view, and the range of views about
commencement of lessons covers all those mentioned above. In this case, a clear concept
is actually vague at its edges, and is therefore open to slippery slope abuse. The problem
is not helped even if individual teachers make their point of view on lesson
commencement obvious. First, their view will probably differ from other individual
teachers, so that although an individual teacher might know exactly what he means by
the beginning of a lesson, he will be aware that the views of his colleagues vary from his,
and will probably respect their right to vary. Second, individual children themselves have
a view of when a lesson starts. So although the teacher might have one view, it will
almost certainly be at variance with most of the children in the class. This problem is
compounded by the fact that, as a child's day unfolds, she will not see a consistent
attitude from her teachers. So although she knows she must remove her coat at the
beginning of a lesson, she will find that the time into the lesson when she is actually
expected to remove her coat will vary. Although, therefore, a lesson does begin, there is
obviously no hard and fast rule as to when it should. It would be absurd to say, for
example, that all lessons must begin three minutes after the previous one finishes. The relation to the slippery slope problem is obvious. If we concede that no matter when we define the beginning of a lesson, a few extra seconds taken in preparation will make no difference to the lesson itself, and further that we can often expect to add seconds because of unpredictable circumstances, then lessons can, by the Sorites slippery slope, start later and later by small increments. Likewise, they can start earlier. This is impossible when they are backed against a previous lesson, but is quite possible where they start the day or join a break or lunch time. We can all remember, for example, the concern amongst public as well as teachers when the first appearance of computers in schools a decade or so ago led to children arriving up to an hour-and-a-half early. This was generally not planned: no matter how early the computer teacher arrived the children were already there, and the earlier and earlier opening seems to have emerged by increments.

I am not here arguing that lesson preparation time should be prescribed, or that lesson preparation will always lengthen. What I am saying is that the use of words such as 'lesson' introduces vagueness into rules which can be prone to Sorites. To say

Remvoe coats at start of lesson

can create as many problems as

Six children can wear their coats in lessons

Teachers will not deliberately ignore the rule or try not to enforce it, but the Sorites paradox has this stealthy quality which means that, over time, states of affairs can change. Even though all teachers might begin a year starting lessons at the same time, the starting times begin to drift apart as the year progresses. It becomes very easy to condemn teachers for not trying hard enough when, in fact, they are not aware that they
are doing anything different. This, after all, is the insidious nature of the Sorites: you
don't realise things have changed because the change is gradual. Worse, the children
receive mixed messages and begin to develop their own interpretations of when a lesson
starts. I would propose, therefore, that rules should be carefully scanned for vague
concepts, and these removed as far as possible. In my own school, for instance, we have
the rule that coats should be removed at the beginning of lessons. The sort of creep in
timing I have described above does take place, and children become very confused about
what is required of them. When the children go to assembly, however, they are told to
take their coats off before they enter the hall. Here no children wear coats. The rule is
tied not to a vague and hence moveable concept such as 'lesson', but is instead associated
with a more concrete concept. In this case, they do not wear a coat once they have
walked through the door. This admits of much less vagueness than the term 'lesson', and
so is more consistently enforced and accepted. Instead of removing the vagueness, a
second approach might be to remove the clearness. Suppose that we abandoned the term
'lesson' altogether, and simply left it to the discretion of the individual teacher to ensure
that coats were removed at an appropriate time. The slippery slope would no longer
apply since there is no clear area to begin with. The time when coats are removed is the
time when coats are removed. The Sorites slippery slope cannot be entered since there is
no fixed or clear time to which seconds can be added and the move into vagueness
begun. As we will see below, however, the Prisoner's Dilemma and the free-loader
problem beckons with this approach.

I will conclude this section on formal rules and the sorites paradox anecdotally. On one
occasion I disciplined a boy for throwing a piece of paper out of the classroom window.
His justification of his act: 'It was only one piece of paper' was a clever answer and also
the quintessential sorites. In disciplinary situations we are often faced with exactly this
problem of escalation: when does an act of disobedience become unacceptable. The boy,
effectively was saying that in his opinion, one piece of paper is too insignificant to mark
the border between acceptable and unacceptable behaviour. My response, to punish, was
a tacit statement that I had no intention of slipping down the sorites slope. Too often, however, we are faced with the sorites paradox in these situations and often, usually with the best of intentions, ignore it. In this way discipline, by small increments, can ebb away.

**Informal Rules**

Having considered the problems with formal rules, I would now like to consider informal rules - ethos, norms or standards. I have discussed these elsewhere. In fact I have argued strongly that the development of informal rules is not only endemic to groups but is also a rational process. Problems arise with informal rules for a number of reasons. First, teachers may develop a different ethos, or set of expectations, from their children. Second, norms tend to be based on tacit agreements which, by their nature are, not discussed or aired, and are consequently much more difficult to influence, direct and apprehend. Third, the ethos of a group changes as the group changes, and so is difficult to relate to consistently. The outcome of these problems is often conflict, made worse because the parties to the conflict firmly believe they are in the right. There are several groups within a school - groups of children, teachers, managers, ancillary workers, and so on. Each develops their rules or norms. Then they use these rules or norms as a basis for influencing the rules or norms of other groups. Unfortunately, the membership of groups is intermingled. Although a teacher has certain teaching staff norms, she also becomes a member of the groups of children she teaches. So her norms not only modify, but are modified by, these other group norms. These, in turn, modify the norms formed with other teachers. Management groups have even more complex relationships, as they endeavour both to relate to teaching norms and class norms. This ever-shifting milieu becomes easily too complex to be predicted. I will try to illustrate what I mean by analogy, with reference to a computer model.

It can be argued that it is relatively easy to model simple social interactions on computers. Chess programs, for example, do this. It might be objected that these merely
model conflict as a social interaction, but this need not be so with other models. For example, Axelrod in his book cited above shows how the Prisoner’s Dilemma matrix can be used as the basis of a computer model, since it represents both the results of conflict and co-operation. In fact, all non-zero-sum games represent a particular blend of this dichotomy. It is relatively easy for a computer to represent the mechanics of a matrix game in that it can easily determine the scores and apportion them accordingly among the players. What is more difficult is to cause different strategies to be modelled. Axelrod used human competitors to do this, since players were invited to submit a program which would describe their strategy and determine the choices made during each game turn. In this way, programs could be pitted against programs, to see which would accumulate the most points. After each program had played every other, the overall winner was determined, and conclusions drawn therefrom concerning the appropriateness of the strategy. Following Axelrod’s lead, I set out to model a game between a number of players. On this occasion, however, each followed a similar strategy. Further, I tried to find a strategy which played the multi-person Prisoner’s Dilemma described in the first section of this chapter in such a way as to produce a seemingly random output of collective choices, as players circle aimlessly around, trying to decide whether it is better to co-operate or defect. This unpredictable output is surprisingly difficult to achieve. Suppose, for instance, we assume that players will only change their choice if they see it doing worse than expected. We could build in a ‘tolerance’ factor, an amount they are prepared to lose before they change. Even though these tolerances might vary from person to person, the game settles to a set of stable choices - each player repeating its choice continuously - within a few moves. If we try to vary the tolerance over time by some function of a player’s previous scores, it still rapidly settles to a stable choice-set. Only when the variation in tolerances is related to other players’ scores does the outcome become unpredictable. In other words, players begin to struggle when their expectations are influenced by the actions and achievements of others. The resulting sets of choices show a satisfying, almost chaotic oscillation. For example, in a choice between co-operation and defection in a game of twenty players, the following numbers chose (if we
can as refer to a player modelled in a computer as having a choice) to co-operate, turn by turn:

<table>
<thead>
<tr>
<th>Turn Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number co-operating</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>9</td>
</tr>
</tbody>
</table>

This is not a truly chaotic outcome, since it is not entirely affected by initial conditions.

As well as considering the numbers making each choice on each turn, it is also possible to follow an individual's choices over a series of turns. An inspection of these choices show that some players' initial tolerances never change, and yield a choice continuously the same. Others players change occasionally, but have a dominant choice. Others switch between choices, using one for a while then another for a while. Some, however, switch every few moves. These, as would be expected, are sensitive to initial conditions. If a game is replayed, and one of these sensitive players has their starting tolerances slightly varied, then, as would be expected, they make a different set of choices. What is not so obvious, however, is that they occasionally skew the whole outcome. I will print the above table of turns, with a second and third game where one of these sensitive players now commences with two slightly changed tolerances

| Turn Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|-------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|
| Game 1 co-operators | 8 | 7 | 7 | 7 | 7 | 10 | 5 | 9 | 6 | 7 | 7 | 7 | 7 | 7 | 8 | 5 | 5 | 2 | 9 | 4 |
| Game 2 co-operators | 8 | 8 | 7 | 8 | 7 | 8 | 10 | 6 | 7 | 6 | 6 | 7 | 8 | 7 | 8 | 4 | 11 | 5 | 9 | 4 |
| Game 3 co-operators | 8 | 8 | 7 | 8 | 7 | 8 | 10 | 5 | 8 | 6 | 6 | 7 | 6 | 7 | 7 | 4 | 5 | 3 | 7 | 4 |

Although the difference in numbers co-operating is usually only one, reflecting the changed tolerance of the sensitive player, occasionally it is quite high. For example, Game 2 move 17 shows a change of six players. Here, the particular circumstance at this point in the game has led a cascade of other players to follow the changed choice of the
sensitive player. The sensitive player has been far more influential than its individual choice would seem to warrant. In her book *Mathematics in Daily Life*, Joanne Growney describes a real-life situation analogous to this one. She asks us to imagine a group of people, each of which will involve themselves in a particular concerted action provided so many others will. These particular 'thresholds' of numbers vary from individual to individual. If no individual, for example, has a threshold of none ("I will act even if no one else does"), then the action will never start. Even if an individual has such a threshold, the action may involve only a minority, since the numbers involved might never cross most people's thresholds. Alternately, all thresholds might be crossed in a chain reaction of involvement. The computer program reflects this scenario, and shows how individuals can indeed make a telling change to outcomes. More interestingly, the individual does not have to be particularly persuasive or popular. Often his addition to numbers is sufficient to cause others to cross the threshold and begin the chain reaction.

This whole discussion of the computer model, I believe, is central to a grasp of my suggestion that the formation of informal rules and norms occurs in an unpredictable, self-referential, paradoxical way. Individuals come to groups with expectations of success - particularly, they have their own agenda which they would like to satisfy as much as possible. Their views on how this can be achieved are influenced as much by others in the group as themselves, so that, over time, their expectations alter. In time, certain typical styles of thought and action will evolve. In the computer example, some always co-operate, some co-operate most of the time, and so on. We can think of these as discreet groups or, as flavours within a larger group. If we think of the computer as modelling a number of groups, then the sensitive players are the ones, like teachers or managers, who move between groups, affecting and being affected. Not only are they influenced by the groups they are in, but, from time to time, they can also influence, by cascading, the mores of the groups into which they move.
The swirling, paradoxical and unpredictable nature of this situation is obvious. Again, people who cannot adequately manage it may be accused of laziness or lack of will, but could just as easily be suffering from a weak situation.

The most obvious and extensive of these weak situations are free-riding and the Prisoner's Dilemma, and I will now discuss these in relation to the making and enforcement of rules. As we observed in Chapter 3, free-riders - people who defect but still benefit from the co-operation of others - move within the orbit both of the Prisoner's Dilemma and the Sorites. We might first consider free-riding from the point of view of a pupil. Since the adherence to rules will tend to enhance the learning environment and the ethos of the school generally, all children will benefit from a co-operative ethos. In addition, their future career prospects will be enhanced if they are known to go to a 'good' school where children are well-behaved and motivated. Any child can argue, however, that sticking to the rules is a burden. If he were to break the rules, then it wouldn't make a difference since every one else would still be following them. This is Sorites thinking. Further, he has the incentive to defect since the rewards are greater than co-operation. This strategy, however, skewers itself on the horns of the Prisoner's Dilemma as each person reasons in the same way and mass defection leads to the worsening of the situation for everyone.

In order to examine free-riding from the point of view of the teacher, we will return to the coats-in-lessons problem. We saw earlier that rules like this one, if not thoughtfully enough worded, could lead to the Sorites slippery slope. I suggested that one remedy would be to remove the change from clarity to vagueness, which is essential to the slippery slope, by dispensing with clarity. We might suggest that children remove their coats at the discretion of the teacher. The incentive for an individual teacher to freeload is immediately apparent. Whilst it is in teachers' best interests to present a united and consistent front with the children, so that they get used to obeying rules, the trouble this causes for an individual teacher could be side-stepped if he chose not to enforce rules,
whilst letting others do so. Again, the defection strategy of the Prisoner's Dilemma applies.

For all the possibilities of freeloadng, I still feel there is an advantage to giving discretion in rule enforcement to teachers. Primarily, I believe the Sorites slippery slope is an insidious problem and more difficult for an organisation to cope with than freeloadng. The slippery slope, by its nature, is entered slowly and proceeds slowly. It is change by attrition rather than rapid movement. People on the slippery slope can be excused if they do not notice the situation they are falling into: that, after all, is the nature of the Sorites dilemma. Freeloading, especially in small groups like groups of teachers in schools, can be mitigated against. An obvious example of this was manifested in my school which decided, by a landslide majority, to do without end of lesson signals. There was an obvious incentive to defect, and release one's pupils a few minutes early. By and large this did not happen. Discussions with staff revealed a whole set of complex and intertwined calculations:

- If I defect my benefit comes not at the expense of a nebulous thing called group ethos or co-operative spirit, but at the expense of the next teacher, whom I know, who may be a friend, and who gets my children early.

- If I defect I get rid of my children early, but I could get someone else's early. If I am a magical thinker, I won't defect because in that way I might 'influence' the member of staff who sends their children to me. Put another way, and copying the statement in the previous chapter - 'what would happen if we all did it?'

- If I defect I will be seen by others to defect - it is easy for other staff to ask whose these children are who are wandering around without a teacher. I would be seen to be letting the side down and would lose respect.
- if I defect the lesson signals, which we all agree we don't want, will be reinstated - I could be the teacher which causes the management to cascade against the arrangement.

In other words, the system had a reasonable chance of policing itself, because the results of defection were public and dishonourable - they would affect friends and colleagues individually.

This might be contrasted with a similar suggestion, which were much less successful, that staff should encourage pupils to move quickly between lessons, not drop litter, not walk in out-of-bounds places, and so on. The public nature of the effort was as great, but in this case, no other particular teacher was affected if an individual teacher defected. The same applies to absence where staff have to cover for absentees. No one who was off knew in advance who would be affected - it was just somebody. In other words, the need to support one's friends and colleagues, people we know and respect, is of paramount importance. The need to preserve dignitas in the eyes of the general is somewhere second, according to the individual. The need to co-operate for the good of the body-politic, especially in situations where there is no public display, does not seem to be a consideration at all. This was particularly evident to me from my work as an appraisal trainer. I travelled to different schools and took teachers through the appraisal procedure, explaining carefully their roles and responsibilities both as appraiser and appraisee. When it became obvious that the process contained virtually no external monitoring, many course members asked: 'Why should I take it seriously? Why should I bother? I' ve enough work on already. What's to stop me making it easy on myself and just going through the motions?' The only answers I could give were first that we were being treated like professionals who are capable of doing a job properly without being checked. Second, I suggested that if appraisal did not work - if it was treated as a soft option which people could ignore then, if no apparent development took place in the profession as a result of it, it could easily be replaced with something more rigorous.

Effectively we are in a Prisoner's Dilemma, with the results of defection being a worse
situation for all of us. The comments I received to elicit these responses show that, generally, people do not think in terms of freeloading, defections and dilemmas at all. If they had some perception of this thing called group ethos or co-operative effort, the reasons to act co-operatively over appraisal, given its lack of rigour, would have been self-evident.

POLICY MAKING AND THE SENIOR MANAGEMENT TEAM - LEVEL SHIFTS AND SELF-BINDING

The emphasis on the unpredictability of organisations is a significant counter to the view that problems can be solved by a rational process.

Tony Bush, *Theories of Educational Management*

In the book cited in the epigraph to this section, Tony Bush provides a useful summary of what he calls the *ambiguity models* of organisations. The situation described might be extreme, but I hope I have shown that schools are sometimes unpredictably ambiguous, and often because of paradox. Ambiguity causes problems not only for the institution, but also for the leadership of the institution. In particular, managers feel that their status and the extent of their power is ill defined, and that their success is difficult to assess or learn from because feedback is often non-existent or ambiguous. Cohen and March, two *principal* exponents of the ambiguity model, propose that managers should cope with this by using a variety of tactics, including the packing of meetings and the overloading of agendas. In this way, at least some of the desiderata of the management might become policy. I tend not to support this cynical manipulation of the system. If managers try to engineer systems, they can create even more ambiguity by falling foul of paradox. In particular, the paradoxes associated with self-binding and with level shifting become apparent. I intend to argue, therefore, that rather than adopting a devious or stealthy approach, they should endeavour to be much more rigorous in their conduct. I will explain in more detail what I mean with reference to both self-binding and level shifting.

183
Self-binding

It will be remembered that self-binding refers to the way that an individual either restrains her present self so that a future self might benefit, or restrains a future self so that a present sensible action will not be upset. If we do not set up these restraints, then we can get into paradoxical situations of self-reference, whereby we are trying to make decisions on behalf of our present selves which affect future selves whose free choice will affect the choices we should make at present, and so on. If we restrict either the present or future choice, then the paradox dissolves. Too often, managers ignore this need for self restraint, but hope that their future selves will still operate within the same remit. When the future selves do not act as hoped, ambiguity or contradiction in decision-making occurs. I would argue, therefore, that managers should build pre-commitment into their management strategy.

I will start my consideration of this with an example. It is easy to see from any analysis of public examination results that boys achieve less well than girls. It would seem appropriate to consider strategies and then implement some kind of plan to improve the work and results of boys. This activity might become the responsibility of a particular senior member of staff, or involve some kind of group or collegiate approach. I have known several instances of schools deciding to address the problem, and I presume that most schools feel that they should do something to try to deal with the problem: that it is an issue certainly worth addressing. Nothing, however, seems to be done, and boys continue to underachieve. I would suggest that any group or individual that seriously wishes to undertake such a task, should also consider the ways they will ensure that their future selves will not renege on it. Please note that I do not imply that one's future self is incompetent or a backslider. What I am suggesting is that, with the multiplicity of jobs which beset teachers, it is often easy to prioritise away certain issues. Given this, how do
we ensure that any particular issue of importance does not get prioritised away? How, in other words, do we bind our future selves to action?

Both Schelling and Elster, in their works cited earlier, make suggestions about this. Some are appropriate, some are not. All, however, advocate that we give our future selves a price to pay if the work is not tackled. Schelling suggests an obvious method of achieving this. He proposes an individual lodge a sum of money with a friend, which will be forfeit if the task is not completed. In other words, a person makes a one-way bet that they will complete the job. This is obviously not practical in schools, but it does lead to an approach that both Elster and Schelling advocate: that is, that we wager our standing or esteem. We make some kind of public announcement that we intend to complete the job; we literally stake our reputation on it. In the case of a policy to improve the results of boys, this might mean a letter to parents and governors stating the school's intention. *In extremis*, an article for the local newspaper might be written. The whole point is that commitment must be obvious and extensive: it must be public.

Public commitment need not rest with these kind of large scale exhibitions. All teachers have job descriptions which are, or should be, public. These again, are a statement of pre-commitment which can be examined by others. In particular, I feel that senior managers should make themselves open to this kind of peer pressure. All management groups should commit time in their meetings to a discussion of each person's actions. Managers here should make commitments to future actions, and report back on success achieved in previous commitments. This making public not only provides the necessary self-binding, but also ensures that individual targets are made sensible and coherent by discussion with other managers.

A final form of public pre-commitment might be termed the formal. Schools or groups within them are often given grants by external bodies to achieve particular ends: to install a computer suite, to evaluate a reading scheme, and so on. It is also customary for
schools to provide grants internally, from the capitation budget, to support certain schemes. Any form of grant or disbursement is a form of pre-commitment, in that it publicly and financially ties the recipients to the achievement of their targets. This style of finance should, therefore, be encouraged.

Although pre-commitment by public statement makes sense, it can suffer from the law of diminishing returns, in that too many public commitments begin to pall. A more thoroughgoing way for management to bind itself is to develop an ethos of management which encourages a ground swell of opinion. The sort of approach I am talking about is the one I described in the section on Purpose, Aims and Outcomes, which described the development of a computerised reporting system. Such 'popular' movements bind managers in that they are carried along by them. By developing such a ground swell, a manager is forced to go along with it even if it becomes hard work or tedious at a later date.

**Level Shifting**

In the previous section I referred to managers adopting a management strategy which included self-binding. As I have said before, in the section above on Purpose, Aims and Outcomes, the need for managers to decide on appropriate styles of management before managing particular issues is essential. The adoption of a particular strategy such as self-binding is, therefore, a meta-management issue. The decision to adopt appropriate styles of management is a meta-meta-management decision. I have also indicated earlier that these shifts of level can become paradoxical, since what is rational and sensible at one level of decision becomes irrational at another. I wish to pursue this issue more fully here. Not only will we see the need to handle the concept of levels clearly and carefully, but also further illustrate the problems attendant upon self-binding. In particular, I wish to explore the paradox of using one's power to negate one's power.
The DFE publication, *School Governors: A Guide to the Law*, gives a detailed account of the roles and responsibilities of school governors. The information is precise, exhaustive, and grounded in law. The section on meetings, for instance, contains details on the size of the quorum, how meetings should proceed and be minuted, where withdrawal from meetings should operate, detail on the formation of committees, and a paragraph on the power of the chairman. There are other sections on relations with parents, budgetary management, special educational needs, and so on. Speaking as a governor, I know that meetings proceed in ordered ways, and expert advice on procedures and competence is provided by the clerk, an employee of the LEA, who also minutes the meetings. All the formal aspects of a governors work are, therefore, prescribed and restrained. There is no need for governors to control themselves or their power in any way, since the government has done this for them. This style of management is in sharp contrast to schools. There are no such things as articles of government for schools: no constitution which inhibits or guides managers in any way at all. Heads especially have almost total control of the school invested in them. In order for this power not to become tyranny, most heads would accept, I feel, that they need to restrain or share their power in some way, not least because it is physically impossible for them to manage all aspects of a complicated organisation themselves. Heads, therefore, delegate tasks and arrange for the production of policies. I wonder if this is enough, however, to adequately ensure that the organisation runs without ambiguity. School governors know the limits and extent of their powers, because they have a legal constitution to adhere to. Where no coercive power exists, such as in schools, is it possible for heads to create constitutions? Creation of rules of government means that there have to be sets of procedures which are followed by all. Can the head use her power to create such a system and place herself under it? Effectively, can she bind herself in a complete way with her own power? I represented this problem in Chapter 3 with a discussion of the paradoxical nature of the statement:

*I will negate myself*
and showed there it is a complex statement: a statement leading to conclusions about the absolute power of head teachers which I feel most schools do not tackle. And if some power is devolved to other managers, then they run into exactly the same self-binding impasse. I know of no school which has an instrument of government, to which all are subject, which clearly states, like the governors documentation, exactly how the institution will be run and the role of people within it. This is a meta-management task which is too often ignored. Instead, managers more often content themselves with managing issues. The lack of coherent framework in which to do this creates confusion and ambiguity. In particular it allows managers to operate within and outside the system by turns. I will explain in more detail what I mean, with reference again to the problem of dealing with the underachievement of boys.

Suppose a school decided to deal with this problem. A common approach is to set up a working party of interested staff to investigate the matter. This often involves an input of suggestions and ideas followed by a discussion. As we have already seen, this procedure can be fraught because Arrow's Theorem shows the difficulty of achieving a consensus from a disparate group of individual suggestions. Nevertheless, such a group will usually produce a number of recommendations. These will be forwarded to the head teacher or the senior member of staff responsible for the activity. This person is then in an interesting position. What does he do with the information? Does he treat it as a policy, which is now to be followed by the school? Does he treat it as a recommendation, to be treated as advice towards a policy which he now determines? Does he ignore it if he doesn't agree with it, or does he have some obligation to include it somewhere? Interestingly, and this is an issue that managers often fail to appreciate, the internal debate the manager is having is not about the issue of boy's achievement at all, but is actually about how to manage the issue. In other words, the manager is now in a meta-management position: his power is unrestrained because there is no policy which dictates the nature of decision-making. He can use the findings of the working party in any way he likes, if at all. He can use it to bolster his own view, or ignore it and propose his own
view anyway. If he has the job of finally producing a policy, then this policy can be his and totally arbitrary, bearing no relation to any other thinking than his own. Because the working party, the manager, and their relationship is not embedded in some higher constitutional format which overarches them, the manager can become a tyrant. Effectively the manager has the freedom to say 'I take the decisions'. But this does not mean only deciding on issues. From the meta-management level-shift he is allowed to make he can also decide who makes the decisions. This ability to make decisions about decisions puts managers into a very powerful position. By turns they can use suggestions from committees and working parties to empower or enable their views if the suggestions conform with them, or can treat them merely as suggestions to be disregarded if they do not. However, a manager faced with a forthright committee might find himself manipulated: the committee, or influential members of it, can use the confusion over meta-positions to gain the advantage. Arbitrarily, therefore, committees become makers of policy or not, and members of them, as well as the people who manage them, often become confused and disillusioned. This would not be an issue if working parties or committees were not popular, but in schools they seem to be almost the universal decision-making forum. Please note that I am not suggesting that managers or the managed consciously manipulate the situation; quite the reverse. The thrust of my dissertation is that paradox manifests itself stealthily: most managers are probably not aware of their shifts in level. Further, they are mostly as confused and disillusioned by the outcomes of their activities as are the people they manage. Tony Bush, in his work cited above, quotes on this issue from Noble and Pym, Collegial authority and the receding locus of power, *British Journal of Sociology*

In the lower levels officials or committees argue that they, of course, can only make recommendations. Departments must seek the approval of .... the general management committee. It is there we are told that decisions must be made .... In the general management committee however, though votes are taken and decisions formally reached, there was a widespread feeling, not infrequently
expressed even by some senior members, of powerlessness, a feeling that
decisions were taken elsewhere .... as a committee they could only assent to
decisions which had been put up to them from one of the lower tier committees

I am arguing strongly that this state of affairs, this receding locus of power, occurs
because of the ease with which level shifting can take place: the easy way in which
decisions about both an issue and its management are resolved simultaneously and often
with consequent confusion. Worse, this resolution of the issue is usually specific to the
issue. The next issue involves a whole new round of level-shifting and confusion. The
root of this difficulty lies, I believe, in the lack of an overall constitution or set of rules
which defines the roles of manager and committees, and which is subscribed to and
followed by all.

Is the creation of such a document possible within schools? I would argue that the
paradox of self-binding would mitigate against it. The head teacher would need to bind
herself to the constitution - to become subject to it - by using the very power she wishes
to negate. She cannot create a constitution to which she is beholden since this
constitution, like all policy documents within schools, has validity only because it has the
support and authority of the head. In other words, there is no absolute authority other
than the head's power to make this constitution work. So she would have to use her
power to negate her power, and the paradoxical nature of her position is complete. This
situation should be compared with the power of governors, which is defined and
circumscribed by law. Here a higher authority does exist, so the paradox of self-binding
is avoided. The only form of restraint on a head's power can, therefore, be voluntary
restraint. As we have seen, however, level-shifting is a stealthy phenomenon, and
voluntary restraint, unlike compulsory restraint, is exactly the binding which is prone to
it. Further, it is no use investing authority in someone else, such as the chairman of
governors. Although this may remove problems for the head, it simply puts the chairman
of governors into the self-binding paradox. As I have said before, the problem here
resides not in the weakness of individuals, but in the weakness of the situation. Autonomous organisations with autonomous leaders will always be vulnerable to this situational weakness, the paradox of self-binding.

CONCLUSION

In this section I have tried to look at the practical manifestations of paradox: to look in greater depths and the scenarios I described at the beginning of chapter 1. I have tried to show how people in organisations can find themselves in difficulties by describing my work with children on the Prisoner's Dilemma. I have showed how the purposes of an organisation can be confused by the inability to distinguish between outcomes which can be achieved directly and ones which are only by-products. Too often we try to obtain the by-products directly, particularly by the application of narrow or game-theoretic views of management and rationality. I have shown how the structures of organisations become problematic, again because rules are as important in the by-products or informal structures they develop as in their formal or intended outcomes. Also the efficacy of rules and structures can be affected by slippery slope problems and the problems of the Prisoner's Dilemma. Finally, I have shown that management can suffer both from the problems of self-binding and level-shifting, and that these problems seem difficult to resolve.

My aim in the next chapter is to counteract these somewhat pessimistic conclusions about paradox by considering its positive side. I wish to explore the uses of paradox, to see if the deficiencies of paradox can be outweighed by its benefits. I shall review the role of paradox in creativity as a way of dealing with problems, and its spur to dissonance reduction which itself provides a different but nevertheless equally useful way of dealing with problems. In Chapter 6 I will consider some of the positive aspects of this present chapter, particularly empowerment and networking.
Chapter 5
Living with paradox

'Allow me,' said Mr. Gall. 'I distinguish the picturesque and the beautiful, and I add to them, in the laying out of grounds, a third and distinct character, which I call *unexpectedness*.'

'Pray, sir,' said Mr. Milestone, 'by what name do you distinguish this character, when a person walks round the grounds for the second time?'

Thomas Love Peacock, *Headlong Hall*

Paradox may be genuine and unavoidable under any definition of rationality. In this case, it may be necessary once more to consider the nature of rationality. Unfortunately, these more broadly rational or irrational procedures seem to lack the rigour required for decisions of consequence, and so tend to be frowned upon. The result, in most decision-making scenarios, therefore, is at best, an attempt to make decisions under known uncertainty. At worst, like the head teacher mentioned at the start of this dissertation, we feel we must try a bit harder. The title of Sutherland’s book suggests that irrationality, his view of it at least, is bad for us and should be avoided where possible. Certainly, irrationality - inconsistency, paradox - is not seen as a bad thing by everyone. Edward de Bono, for example, in work described in detail later, encourages irrational thinking as a means of creativity. More simply, and central to my thesis, there are situations which create an irrationality from which it is impossible to escape. In his poem, *He Wonders Whether To Praise Or Blame Her*, Rupert Brooke points this dilemma - this paradox - of inescapable irrationality.
Paradoxes in Decision-making

I have peace to weigh your worth now all is over,
But if to praise or blame you cannot say.

For, who decries the loved, decries the lover;
Yet what man lauds the thing he's thrown away?

Be you, in truth, this dull, slight, cloudy naught,
The more fool I, so great a fool to adore;
But if you're that high goddess once I thought,
The more your godhead is, I lose the more.

Dear fool, pity the fool who thought you clever;
Dear wisdom do not mock the fool who missed you!
Most fair, - the blind has lost your face for ever!
Most foul, - how could I see you when I kissed you?

So....the poor love of fools and blind I've proved you
For, foul or lovely, 'twas a fool that loved you

The poem shows much more eloquently than I that irrationality - the act of being a fool - cannot be avoided: or, at least, can only be avoided with hindsight. Further, the predicament does not only apply to falling in love. The stuff of Greek and Shakespearean Tragedy is this inability to escape the irrational act. When, in King Lear, Edmund says "reason be thou my goddess" we instinctively flinch, knowing no good will come of such an invocation. If irrationality is, therefore in certain circumstances, inescapable, and by Sutherland's evidence, as a species, we have a natural predilection for it, may it not serve some useful purpose? May there not be some good attached to it? In the down to earth way advocated by Sutherland, may we not tolerate inconsistency and fuzzy logic? I shall argue first that inconsistency is pervasive in practical terms and cannot be avoided. I shall argue next that paradox is actually beneficial in situations that demand creative or
original thought, and here a style of rationality is indicated which tolerates ambiguity and inconsistency. I hope finally to show that a commitment to this style implies that supposedly irrational styles of decision-making: lottery, dissonance reduction, and so on, are, in fact, acceptable.

INCONSISTENCY

We are still caught up in the backwash of the ancient Greek conviction that only those disciplines whose rational systemization proceeds by way of mathematico-linear development are rigorous and solid, and that those disciplines that proceed by way of dialectically cyclic argumentation are somehow inferior and unsatisfactory.

Nicholas Rescher, *Rationality*

At first sight, consistency appears to be a *sine qua non* of rationality. After all, given similar situations, we would expect rational people to act in similar ways. There are, however, compelling reasons for, if not abandoning consistency, at least relaxing its constraints. If we take an inductive view of the world, which groups actions into larger structures - motivations, causes, laws and so on, then a case might be made for the immanence of consistency. If, however, we see the world as a series of individual, unique, contiguous actions and events, then no two situations will ever be exactly alike and consistency becomes not only impossible, but largely irrelevant: there is no larger structural background to be consistent against. Consistency, in effect, is not achievable in practice. Much of my thesis tends to lead to this conclusion, and I shall argue it further with reference to the work of Rescher cited in the epigraph. Rescher sees inconsistency as inherent in all logical systems, because our knowledge and our axiom systems can never be perfect. Most logical systems contain inconsistencies, and our application of them depends upon the state of our knowledge. Given the full extent of what could be known, Newton's was as good a stab as possible at an explanation of the universe. Only
the increased knowledge available to Einstein made his theories possible. Both, however, were a product of rigorous logical deduction. They differ only in their inputs and premises. Rescher sees the extent and validity of this starting knowledge base as a function of the society upon which it is based. Some societies are more advanced than others, so the application of logical deduction to their knowledge produces an answer closer to the absolute truth than less advanced societies. Given the state of scientific knowledge in the eighteenth century, for example, the phlogiston theory was a perfectly logical result. It is not the case, therefore, that weak or inconsistent reasoning is a product of weakness of will or frailty of intellect on the part of individuals. It is simply a result of the extent and soundness of knowledge available at the time. He argues that we accept this case in fact by having two sets of beliefs. These are first beliefs of which we are absolutely certain, and would stake anything upon. Second, and more ubiquitously, there are beliefs which we consider plausible, given the evidence. Although we operate on a day to basis as if these second types of belief are true, we accept a certain ambiguity and inconsistency in their outturn simply because we know that they are tentative. Rescher says:

For once our approval and belief is recognised as being in some degree tentative and fallibilistic, it makes perfectly good sense to contemplate the acceptance of incompatibilities - to take inconsistencies in our stride.

He sees this lack of consistency as due to the impingement of the real world. He says, for example,

one must regard perfect rationality as an idealization and acknowledge that we humans are 'rational animals' because of our capacity for reason, and certainly not because of our achievement of perfect rationality.

and
If rationality were only possible in the light of complete information it would perforce become totally irrelevant to us. It lies in the inevitable nature of things that we must exercise our rationality amidst conditions of imperfect information.

He proposes the use of the dialectic as a natural extension of this type of ambiguous yet psychologically acceptable mode of thought. He sees several disciplines, particularly those that are evidential or anecdotally based, as proceeding - as having to proceed by the imperfect nature of their information - by the dialectic: by the conflict of inconsistencies. The resolution of these inconsistencies leads to a general advancement towards the truth, with a concomitant move towards further inconsistency. The changing perspective continually generated by this effort to resolve problems leads us to reconsider already resolved issues from a new angle, and so to redefine our previous conclusions.

One 'tightens the net', so to speak, through continuous construction of an increasingly adequate case.

This swirling and cyclical activity he describes is similar to the swirling nature of paradox although, of course, no paradox is implied by the dialectic approach as such. I shall argue further than Rescher below by showing that this reasoning from ambiguities is not just the prerogative of certain disciplines. In many ways, it underpins the whole cycle of creativity. In effect, there is not one type of rationality which can be applied to all information in all disciplines. The logico-deductive method is appropriate to subjects such as mathematics and decision-theory, but is not as useful, in a subject like history, as a disputive or dialectic approach.

Just as Rescher argues for this dialectic style of thinking because it is conformable with inconsistency, he argues equally against linearly deductive styles because they are not conformable with it. His general aim is to show that the quest for consistency is not only
chimeral but a drawback to genuine advancement in certain disciplines. The quote from his work which opens this section shows the burden under which the 'soft' or non-experimental sciences, in particular, labour, because the discursive and dialectic style, although suited to the nature of the data from which they work, lacks the rigour demanded of a truly rational approach. His view concurs with mine in that he sees that there is something essentially imperfectable about this rigour - this reason and logic - so that inconsistency and supposed irrationality become unavoidable. We disagree, however, on the reasons for this. For Rescher, rationality breeds inconsistency because the raw material, the information on which it acts, is imperfect. Given perfect knowledge we could, presumably, achieve a perfectly consistent action. Although I accept that information is naturally imperfect, I have tried to show that no amount or quality of knowledge will improve certain situations. It is difficult to see, for example, how any increase in knowledge would actually prevent the problems which occur in the Prisoner's Dilemma. It is the situation which is imperfect here, not the information. I am arguing, therefore, for a much more robust rationality which tolerates, accepts and uses ambiguity and paradox, since these are so often unavoidable.

PARADOX AND CREATIVITY

it is on such occasions that my ideas flow best and most abundantly. Whence and how they come I know not; I cannot force them.

Wolfgang Amadeus Mozart, Correspondance

Ambiguity and inconsistency can be the basis of a broader rationality, which achieves ends in a more oblique way. In particular, it can be seen as the basis of much creative activity, and in this sense does have an overarching rationality, a rationality based on inconsistency. I hope to show that the seeds of creativity can be found in the elements of paradox: juxtaposition of opposites; circularity; level shifts. Each of these have been advocated either singly or as a group by a wide variety of authors. I shall explore the
Paradoxes in Decision-making

Living with Paradox

issue of creativity and its possible relation to paradox centrally by considering the work of Edward de Bono.

The medieval philosopher and savant Ramon Lulls felt he had discovered a foolproof method for converting the Muslims. He constructed a device consisting of two discs, pinned together at their centres so that they could revolve independently, the one slightly smaller than the other. In segments around each disc, short mottoes and tags based on biblical texts were written. As the discs were revolved at random, texts on the outside wheel were brought into juxtaposition with texts on the inside wheel. The combination of texts, and the ideas they would generate, would be so startlingly original and profound, Lulls felt, that they would be bound to convert the heathen. He further believed that a complex enough set of wheels would be capable of generating all the ideas that could possibly ever be thought. Lulls, in fact, had discovered one of the fundamental techniques of creative thought, the forced associations, a technique proposed explicitly by Charles S Whiting in the 1950s and also advocated by de Bono. In fact, much of de Bono's work is based on earlier (and mostly unacknowledged) commentary. The earliest what may be termed systematic description of the creative process is contained in a fine piece by the great French mathematician and philosopher Henri Poincaré. In his Science and Method, he produces a sustained and brilliant passage on mathematical discovery. The centrepiece of his thesis, like Lulls and de Bono, is the act of combination: herein lies the key to creativity:

To create consists precisely in not making useless combinations and in making those which are useful and which are only a small minority. Invention is discernment, choice.

Like de Bono, and unlike Lulls, Poincaré sees creation as an act of combination mitigated by choice. Lulls, one feels, is more influenced by the Cabalist tradition which sees the point of combination as exhaustive: to find out everything about everything. De
Bono sees choice within combination as a necessary part of the creative act and, as we will see, suggests a number of strategies which may be employed in order to ensure useful choices. Poincaré, although he concedes a central role for choice, follows a much less rigorous line about its nature. First, he argues that the act of combination is largely subconscious. Second he argues that the act of selection is likewise subconscious. The aim of conscious action is to prove, rationalise and exemplify the subconscious or intuitive combination and selection. Poincaré, like Mozart in the extract quoted at the beginning of this section sees creation as a thing that cannot be willed. It is essentially a by-product, and does not occur when it is concentrated upon. He cites an extensive example of the generation of one of his own mathematical proofs, which came into his mind when he was out walking and later when he was boarding a bus. The only wilful part of creativity is the preparation for it. Poincaré urges the inventor to steep himself in his subject: and then effectively to go on a long walk and forget the issue. Poincaré himself uses the word paradoxical to describe this state of affairs. Certainly, wilfulness and its by-products can lead to paradox, as we have seen throughout this dissertation. I would also argue that Poincaré's description of the process of creation is paradoxical by self-reference, although he does not say this. I will consider this self-referential nature by first considering one of his memorable images, which I shall extend by considering a similar image contained in the work of Douglas Hofstadter.

Poincaré likens the mind to a vessel containing atoms, which are stuck to its inside wall. The preliminary and conscious work sets certain of the atoms in motion, those associated with the particular topic involved. The more extensive the work, the more atoms are dislodged. The atoms are not just dislodged, however. The act of conscious work also shakes up the vessel, so that the atoms fly in all directions. The atoms collide and combine. The subconscious mind selects the combinations which apply to the topic, and in particular the fresh combinations which have not appeared before. The subconscious mind further selects from these fresh combinations the ones which are particularly viable. The whole point of Poincaré's hypothesis, and the place where he differs fundamentally
from de Bono, is his belief that the conscious mind does not have the ability to make this selection. In taking this stance he feels that he lessens the nature of the paradox of creativity where the most result seems to emerge from the least activity. He argues, in fact, that although the conscious mind is at rest, the subconscious mind is working overtime; it is this work that produces, quite naturally and not paradoxically the creative outburst. Unfortunately, he is not prepared to propose a mechanism by which the subconscious mind does the selection. Like many of his contemporaries (he was writing in 1908), he seems to feel that an evocation of the subconscious, along with all its supposed miraculous and incomprehensible powers, is sufficient. This deficiency was remedied, some eighty years later by Douglas Hofstadter, in a description of the mind as powerful, memorable, and somewhat more witty than Poincaré's.

In his *Metamagical Themas* Hofstadter includes a discussion on the nature of consciousness under the title 'Who Pushes Whom Around Inside the Careenium.' The Careenium is exactly analogous to Poincaré's mind (his cranium), where a large number of marbles are flying around and careening off each other. These marbles are magnetic - Small Yellow Magnetic Marbles (SYMMs) - and like Poincaré's atoms they stick together. These groupings Hofstadter refers to as SYMMballs. Because they are large, they deflect and channel the movement of marbles so that these tend to follow similar paths over time. This is analogous to our thoughts being directed by the symbols and ideas which we carry around in our minds. However, these SYMMballs can themselves be changed by accretion or breakage from the surrounding flow of marbles. So a slow speed view would see marbles being deflected by apparently immovable, unchanging SYMMballs. If the view of the scene was speeded up, however, so that the movement of individual molecules becomes a blur, the SYMMballs would be seen to be moving and changing. In this way, our ideas and concepts not only channel (shove around) our thought processes, but are themselves moulded and restructured by the same thought processes. It is this essentially recursive, level shifting, paradoxical type of activity that I feel accounts for the pattern forming that takes place in the subconscious mind. I would
argue that the intense mental activity which predates creativity - which starts the atoms flying to use Poincaré's analogy - also forces the breakdown of pre-conceived ideas and concepts. The new structures - new SYMMballs - might possibly contain a way forward which would not have been perceived if the marbles had not been allowed to careen off each other and it is these SYMMballs that become articulated by the conscious mind at some later date. As we will see when we discuss the work of Edward de Bono, creativity always seems to involve the breaking down of preconceived ideas. Hofstadter's concept of the careenium seems to be a good description of this.

If we concede, therefore, that creativity is based upon some kind of self-referential and hence paradoxical mental activity, it would seem that the existence of paradox is, therefore, a benefit. I will now argue that paradox is of benefit to creativity in a practical as well as a philosophical way, and to do this I will return again to the work of de Bono. In particular I wish to examine his contention that creativity can be an act of will, a stance which directly contradicts much of the earlier work on creativity, as exemplified by the work of Poincaré. The mechanics of this act of will are centred on the acts of generation and association. In the several methods he proposes for these I will argue that paradox has a useful role to play, particularly in its natural tendency to produce opposites and absurdity. In fact, the role of paradox in the act of creation is strongly hinted at in this quotation from C. R. Rogers, in a paper *Towards a theory of Creativity*, written in 1954. He says a basic trait of creativity is 'openness to experience' which he defines:

> It means a lack of rigidity and permeability of boundaries in concepts, beliefs, perceptions, and hypotheses. It means the tolerance for ambiguity where ambiguity exists. It means the ability to receive much conflicting information without forcing closure on the situation.
The reference to ambiguity is interesting in that the first sentence of the passage is itself ambiguous. Although I feel it is unintentionally ambiguous, it nicely illustrates the point Rogers is trying to make. Does the sentence in fact mean that openness to experience is both a lack of rigidity and a lack of permeability? If this is the case, then in what state is openness to experience, if it is neither rigid or permeable? The hunt for a third state which is neither of these is the beginning of an act of creation. Edward de Bono suggests that the sort of thing that has happened here by accident, and for Poincaré happens in the subconscious, in fact happens often by design. Creativity for him is partly a matter of will, and can be helped by tricks and techniques which generate conflict and ambiguity. The resolution of this conflict and ambiguity leads to the creative act. And paradox, I have argued throughout this dissertation, is an excellent generator of conflict and ambiguity.

I will first describe the techniques that de Bono proposes to support creativity, and show where these can be related to paradox. He proposes a two stage approach to creativity. First is the divergent stage, which equates roughly to Poincaré's first two stages: the finding of a wide range of ideas which are combined in novel ways. Second comes the convergent stage, where the fruitful or useful ideas are sorted from the useless ones. This roughly equates to Poincaré's third and fourth stage, where choice is exercised and the intuitive leaps consolidated into stepwise proofs. It is in the ideational stage where ambiguity and conflict are at their most prominent. A number of ideational techniques illustrate this point. These are taken largely from his book, *Lateral Thinking*.

In order to find a way of approaching de Bono's work, I would like first to consider briefly the development of views on creativity. A useful summary of this can be found in the introductory chapter of Doris Shallcross' *Teaching Creative Behaviour*. The pioneers of the study of creativity combined a variety of programmes and objectives. The *Aptitudes Research Project* led by J P Guilford succeeded in devising tests to assess creativity which isolated certain character traits associated with it. Guilford believed that
creativity was common to all humanity, and could be located within any individual. D McKinnon and F Barron, however, sought the well-springs of creativity in a study of the work and thought-processes of accepted creative thinkers. E P Torrance supported this work, as well of that of Guilford, by assessing, identifying and investigating the work of creative schoolchildren. The sorts of conclusions Torrance reached are of particular interest. By 1963, his book *Education and the Creative Potential* illustrates the extent to which the promotion of creativity as a discrete activity had developed. First, he shows how creativity can be fostered in children: in other words, he is no longer content to simply describe or identify the traits associated with it. In this he leans on the work of A F Osborn, whose book *Applied Imagination* was probably the first creativity primer, containing activities that might be used to promote or train creativity. Second, he accepts without discussion certain traits that are associated with creativity. They are not vindicated or urged: the traits by the 1960's had become accepted as essential to creativity. I hope to show that in his work, de Bono likewise accepts these characteristics as essential (thereby accepting the pioneering work of the American Universities in the 50's), and then uses their promotion as a basis for the development of creativity (thus aping the work of Osborn and Torrance). The traits that were isolated in the fifties, and I believe de Bono pursues although he does not name them as such, are as follows. First, creative people manifest the characteristic of fluency: that is, they are able to produce a large number of ideas on a particular theme. These ideas may be of no great worth, and may later be rejected, but their plethora is the basis of the choice which we have seen lies at the heart of the creative effort. The second characteristic is flexibility - the ability to find a number of different categories of ideas. The third trait is elaboration, the ability to conceive ideas in detail and specifics, rather than just in vague or general terms. Finally, creative people are original thinkers, devising ideas and concepts that have not been perceived by others. I shall use these headings to analyse de Bono's work. Much as he likes to trumpet the originality of his thoughts, I believe his approach relies heavily upon the research that went before him, particularly in the isolation of the traits just mentioned. The merit in de Bono's work lies in the extensive number of techniques he
adduces to help develop these traits in individuals, many of which I will argue can themselves be enabled by the use of paradox.

The techniques which de Bono proposes are a series of mental exercises. He does not propose a holistic approach to creativity, but instead tries to break the process down into a series of steps. The activities are often artificial, bearing little relationship to 'real' problems. De Bono accepts this, explaining that we learn maths, for example, in the same way, by dealing with artificial problems to acquire skills, which can then be transferred to real life situations. Further, he accepts that the style of thinking he propounds is not a natural style. The mind is much more at home with 'vertical' styles. The ways that these creativity skills, when acquired, will actually be transferred to the arena of real life is never adequately dealt with. De Bono insists, therefore, that he wishes to illustrate the processes involved in creative thinking, and to allow agents to explore these processes for themselves.

The making of patterns is central to his thesis. He sees this as the creative act. In fact, he seems to argue that all pattern making is creative, in that the mind

picks out messages from the environment even though no-one has deliberately put them there. (p 27)

Therefore, the mind takes some unformed stream of data and forces it into a pattern. De Bono never claims to be a Berkleian, but he certainly seems to suggest that the universe is a personal and internalised construct. This act of creation, however, is not seen as true creativity. De Bono is quite specific about what creativity is: it is the creation of novel or original patterns, not simply the creation of any pattern. Further, this originality is achieved by what he calls lateral thinking. This is contrasted with 'vertical' thinking, and both are described in some detail. De Bono is quick to place values upon these types of thinking: unlike vertical thinking, lateral thinking is good and should be encouraged. He
Paradoxes in Decision-making

Living with Paradox

is fond of (and good at) producing pithy statements and clever phrases. His use of language emphasises his commitment to the worth of lateral thinking:

vertical thinking is analytical, lateral thinking is provocative;

with vertical thinking one has to be correct at every step, with lateral thinking one does not have to be;

with vertical thinking one concentrates and excludes what is irrelevant, with lateral thinking one welcomes chance intrusions.

The use of terms like exclusion, analysis and correctness presents vertical thinking as lugubrious, boring and negative. Lateral thinking, on the other hand is brave, welcoming, risk-taking and open. It operates with verve and élan. Lateral thinkers are dynamic and forthright, whereas vertical thinkers are conservative fuddy duddies. The fundamental disadvantage they have is threefold (p 53). First, they neither look for new ideas or information, nor have successful techniques for handling the information they already have. This is also seen as a root cause of irrational thinking by Sutherland. Second, they make no attempt to restructure the information they already have. Essentially, a vertical thinker takes information in a serial form, making patterns as he goes. Any further information is subsumed into the previously formed pattern. In terms I have used, vertical thinkers may be called 'Sorites' thinkers. Third, vertical thinkers accept adequate solutions which work rather than trying to find different solutions which work better. To use a reference to work described below, they might be called 'dissonance reduction' thinkers. In order to become a lateral thinker, therefore, one has to attack this mental obstruction and lethargy. Broadly, de Bono's approach to the problem is by Poincaré out of American University.
Paradoxes in Decision-making

As I said earlier, Poincaré described a series of states through which the conscious and unconscious mind cycles in order to create. De Bono tries to formalise this activity into an entirely conscious regime, using a whole range of thinking techniques. He says at the end of his book (p 260):

Lateral thinking is directly concerned with insight and with creativity. But whereas both these processes are usually only recognised after they have happened lateral thinking is a deliberate way of using information in order to bring them about.

Using the headings outlined earlier, I will now describe the techniques in more detail.

Fluency

This is the ability and need to create a large number of ideas, and is the cornerstone of creativity. All writers on creativity argue that the act of creation lies in choice: the choice of the apt or original insight or solution. Choice can only operate in a milieu of alternatives. The more alternatives, the greater the choice. Whether the choice is made consciously or subconsciously, it has to operate on something. Poincaré saw the choice field as a by-product of intensive work on a subject: de Bono sees the choice field as a direct product of techniques. The basic technique is the use of his word 'po'. This means 'I accept the idea but will not discuss it at the moment.' S J Parnes refers to this technique less sensationally as the 'deferred-judgement principle.' Both writers (and many others) are saying that ideas or material should be accumulated without assessment: that any contribution, no matter how bizarre or outrageous, should be acknowledged and noted; that creative thinkers are not satisfied with their first idea. Assessment should be applied at some later date, unlike the more natural procedure of think and then evaluate, think and then evaluate, and so on. The handling of large amounts of information is worth the resultant mental stress and psychological dissonance for the opportunities thus created.
for improved insight. The two techniques proposed by de Bono to facilitate fluency are the use of alternative views and the use of quotas. Alternative views encourages the thinker to see an object or idea from the point of view of its different qualities. So a piece of paper might be seen variously as having a shape, texture, colour, use, composition, and so on. His well-known application of this technique is to the problem of thinking of as many uses as possible for a house brick. This is, in fact, the technique of Attribute Listing, attributed by Shallcross to R. P. Crawford. The quota can be used with the alternative views technique, or by itself. It simply consists of placing a lower limit on the number of ideas to be generated. So a person might be asked to think of ten uses for a house brick. Hybrids are also proposed. The 'why' technique is designed to generate information by forcing answers to question 'why?'. The '5 whys' technique (not proposed by de Bono), generates answers sufficient to reach a quota.

Elaboration

Elaboration is the act of looking at an object or idea in increasing detail. It is seen by de Bono mostly as a technique to develop fluency: that is as a means to generate alternatives. I believe, however, that it belongs under the heading of elaboration because he uses the problem or object itself as a starting point for the solution of the problem or to make insightful statements about the object. In this way, I believe, de Bono, without realising it, is presenting creativity as a paradoxical activity: he is tacitly assuming that a problem, for example, contains within it the seeds of its own solution. Appropriately, self-reference is the device we use to apprehend these seeds. I will give examples of these self-referential techniques to illustrate my point.

The first technique I will consider is that of fractionalisation. It is highly contrived. De Bono argues that this does not matter, since the aim is to generate ideas, not necessarily sensible ideas. For example, a geometric figure can be broken down into smaller geometric figures to see what results. Phrases and descriptions can be decomposed in
any way desired. For example, the design of an apple picking machine could begin with a fractionalisation of the words of the problem. So 'apple' could be fractionalised to its being delicate, and separate from other apples. Likewise, 'picking' could be fractionalised to removing and carrying. The fractionalisation then continues at the next level. 'Removing' is fractionalised to grabbing and jerking, and so on. This concentration on minute detail not only more closely defines the problem, but conversely broadens and generalises it. So an apple, a tangible item is associated with a concept like removing. Further, the result seems to get more out than it puts in. Starting with three words, we can go on to generate as many as we want: to reach a quota or fill a time slot devoted to generating ideas. This is only to be expected of the technique, since self referential systems are capable of generating infinite amounts of data: a consideration of the computer-generated graphic of the Mandelbrot Set, an iterative procedure, amply and dramatically illustrates this.

A second powerful technique is reversal. This technique is close to paradox, in both its performance and outcome. The aim is to generate changing views of an object or problem by reference to the object or problem itself. More precisely, the reference is to the negative or reverse of the target. The result is paradoxical, in that one considers both the object and its opposite: the tension between the two leads to insight. Again, this is not an original concept, bearing obvious resemblance to the dialectic of Hegel and Marx. Ever the pragmatist, however, de Bono applies the technique not to a consideration of the cycles of history but to the problem of directing traffic. In this example, de Bono urges us not to consider the problem of a policeman directing traffic, but to see it instead as traffic directing the policeman. The two images, considered separately and together, may allow a unique and viable solution to be elicited. Again, the technique has generated more information than we started with.

A third illustration of the process of elaboration by self-generation is de Bono's advocacy of the use of analogy. Always a fruitful and powerful device for looking at the world in
new ways, he refines it into an important and deliberate thinking tool. He proposes that
the problem or object be analogised, and the analogy then pursued. As time goes on, the
relating of this development to the original object will lead to insight and changed views.
For example, the problem of finding ones way in the fog could be analogised to the
problem a visually impaired person has generally in finding their way. Alternatively, it can
be related to a similar mental exercise such as solving a crossword puzzle. The analogy
does not need to be exact - in fact, it does not need to fit the target at all. For example,
'life is a bowl of spaghetti' is a quite acceptable starting point for the analogy technique.
All that is required, as with all de Bono's techniques, is the ability for it to generate large
amounts of extra information.

When I discussed self-reference in Chapter 1, I commented that self-reference could be
found in systems or collections of objects, each individual of which was not in itself self­
referential. De Bono proposes a technique which, interestingly, is analogous to this
systemic reflexion. He refers to common, real-life situations where we are only given
partial information. It is incumbent upon the thinker to make sense of this by filling in the
gaps where no information exists. In an artificial way, data can be accumulated by
generating information with deliberately created gaps in it. So, for example, a picture can
be produced with parts omitted. The aim here is to complete the picture. Likewise, a
story could be produced with events omitted. (Here de Bono refers in particular to the
detective story, which omits information by starting the narrative at a point where a fair
amount of the story - the planning of the crime - has already taken place). As we shall
see below, this process of filling the gaps is the basis of the creative leap made when
items are associated. But de Bono also advocates it as a means, in its own right, of
garnering more information. Since the combination of images or words generates further
information by the omissions in the set, I would argue that the technique is a self­
referential one. It is systemic self-reference, however, in that each piece of information in
the set (each word or picture), is not in itself capable of self-reference.
The outcomes of these efforts at elaboration are seen at their best where de Bono describes work conducted with children. Here they are continually encouraged to put more detail in their work. Examples of drawings show a plethora of arrows and writing, indicating all manner of detail.

**Flexibility**

This is similar to fluency, but aiming instead to generate categories of ideas rather than ideas themselves. So ten uses for a cup which all involve the containment of some kind of fluid would be fluent but not flexible. A. J. Cropley, and C. R. Rogers, both cited above, see creative people as tending to assimilate new data items into unusual combinations: they do not code them, give them meaning, in stereotypical ways. They see sensory inputs as individual objects (this tree as opposed to a tree), and tend to relate these items to their own individual experience. It is this essential self-centredness that leads to novel categorisation and increased flexibility. Although de Bono acknowledges the need for flexibility - in fact, he sees the challenging of pre-conceived categories or patterns as essential to the act of creation - his techniques for developing it are not particularly numerous or inspired. He refers to categories severally as dominant ideas and crucial factors (where he sees them as inhibiting) and as labels (more neutrally). A technique he proposes for generating alternative labels is to try to do without the accepted (or dominant) one. So he suggests one should describe war without using the label 'fighting' or school without using 'teaching'. This forces the would-be creative thinker to generate other categories such as 'movement' or 'learning'. A much more thorough and practical study of flexibility techniques is described by Osborn in *Applied Imagination*. Here he proposes a checklist of actions which will increase flexibility. For example, *Put to other uses* might request suggestions for what to do with a hundred roller skate wheels other than put them on roller skates. *Adapt* might ask for a list of things that are like a bathtub. *Combine* is a technique I have found particularly fruitful. It requests a classification of an outrageous grouping, and so breaks down the stereotypical
groupings into which the items would normally go. So, for instance, children might be asked to devise an advertisement for an invention consisting of a disparate group of objects: perhaps a combined vacuum cleaner, toaster and transistor radio.

**Originality**

Much of the work on creativity emphasises the fact that creative thinkers are original: they come up with ideas that no-one else has thought of. In fact, the word original is often synonymous with the word creative. It is surprising that, since de Bono sets such great store by finding new and original patterns, he makes few suggestions on how to be original. I think this is because originality is his overriding aim: the whole point of all the activities described is to facilitate the original, the creative thought. In this sense, therefore, all his techniques are a means of cultivating originality.

As well as proposing the above sets of techniques, de Bono also proposes milieux in which they might operate. Two are of particular interest. The first is the use of random stimulation. Instead of using the problem or object under consideration as a starting point for the techniques, a random object might be used instead. I have tried this with a number of classes, and found it possible to apply several techniques to a useful end. For example, I set several classes the following problem:

*Think of as many ways as possible of finding your way somewhere in the fog.*

I then provided a number of objects picked at random for them to consider: a felt tipped pen; a pen holder; a clock wheel; a book; a drinks mat; a rubber. They were to use these objects as a basis for solving the problem. Generally, the random nature of the objects forced the use of the technique of analogy. The use of a number of objects rather than a single one tended to force flexibility: it was difficult to persist in seeing such a disparate collection as suggesting the same sorts of solution. The objects taken as individuals also
served as a starting point for several techniques and thereafter solutions. Consider, for example, the clock wheel. By applying fractionalisation to it, 'clock' suggested time and 'wheel' suggested vehicle. From this emerged the suggestion that one might wait until the fog had cleared and then proceed upon one's way, or that one might catch a bus to the desired location and let the driver find it. The application of alternative views revealed that the clock wheel had mass: it could be thrown forward into the fog to see if it collided with anything. It also had lustre (it was made of brass), so that it could be used to reflect a light into the fog. It had teeth, so could it be used to cut the fog away - to disperse it? It is important to understand that no individual pupil found all of these solutions, and none applied the techniques in a systematic way. But all found something, and taken as a class several interesting and original solutions were suggested.

The reference to an outcome produced by a class of children illustrates another of de Bono's environments for creativity: the brainstorm. Like so much of what he presents, this is not an idea original to him. According to Shallcross, it was actually devised by Osborn. De Bono uses it unacknowledged, but advocates it with such verve that it has come to be a practice associated with his name. The activity is well known, so I do not propose to dwell on it. I wish to reiterate, however, that it is not, in my opinion, a technique for creativity. It is, instead, a forum or arena, a means of creating an environment, wherein creative techniques such as fractionalisation, reversal and quota can operate. A failure to appreciate this, I believe, leads to the activity mostly being wasted. It can only be successful where the techniques which encourage creativity have already been isolated, accepted, practised and mastered. Too often have I been in 'brainstorming' sessions where no understanding or perception of thinking techniques have been displayed, and where the activity has been simply discussion, argument or gossip under the banner of an acceptable and status-conferring name.
The 'selecting mechanism'

Having generated a whole set of material, it is now necessary to combine it into useful patterns or hypotheses. De Bono has an essentially Darwinesque view of this activity. He says (p 28)

> It does not matter whether they are right or wrong so long as they are definite......Once the patterns have been formed, the selecting mechanism of usefulness (fear, hunger, thirst, sex, etc.) will sort out the patterns and keep those which are useful for survival.

So a successful pattern is based on utility, and de Bono implies throughout the book that whenever a useful pattern is generated it will be recognised as such. He does not develop techniques for recognising utility: he seems to feel that this can safely be left to the thinker, and his or her fundamental drives (fear, hunger, etc.) Whether this is appropriate and sensible, or whether he is advocating exactly the procedures that Poincaré suggests but is simply relinquishing control to the sub-conscious at a later date I cannot say. Certainly, the accepted view of creativity is that a genuinely useful pattern or solution can be recognised. This recognition is often belated, a fact which would explain the surprise with which discovery has been attended, from Archimedes' 'Eureka' to Martin Gardner's book on paradoxes, *Aah! Gotcha*. However, humanity's ability to spot utility even in the apparently useless is extensive. For example, iron was traditionally used to make weapons, small amounts of agricultural paraphernalia and domestic ware. With the increased production of iron in the eighteenth century, its availability far outstripped these traditional needs. John 'iron mad' Wilkinson, the Shropshire iron master, found a plethora of uses for iron: barges, water pipes, a chapel, his own coffin. Likewise, the development of the computer in the forties as an electronic device to crack enemy codes by the rapid handling of information was appropriated to the creation of a revolutionary, unprecedented and wealth-generating style of games and game-playing. I think,
therefore, that we must accept that the mind can recognise utility even in the most irrelevant or unpromising areas. This does not mean, of course, that each individual mind can recognise a useful idea when it apprehends it and, as I have said, de Bono debates this issue not at all. In the light of the general human instinct to recognise utility, however, I do not think it is unfair of him to expect most individuals to appreciate usefulness, at some conscious or unconscious level. Further, we can make the same assumption about the mind's ability to make patterns: ley lines and the laws of physics are particular examples of this general and fundamental human activity. Some work I did with schoolchildren emphasises the innate nature of this trait. I used a computer program which generated, obviously at random, 'misfit' characters. These were based on that style of children's book which creates comic pictures by dividing the page horizontally with two cuts, thus splitting each person into three parts. The heads, bodies and legs can be mixed up to produce an incongruous result. When asked to write a story describing how a particular character came to be dressed in a particular way (for example, in a nurse's uniform, and wearing rugby boots and a miner's hat), no child, and I did this with several hundred, was ever at a loss. I feel certain, therefore, that we can assume pattern-making as a fundamental human trait, which needs no stimulus or explanation. It is de Bono's contention, however, that the pattern-making activity and its output can be improved - can be made to yield more utility, by the application of consciously developed and wilfully applied techniques. These techniques, which I will describe shortly, pay scant respect to their provenance, which is eclectic and disparate. For example, A J Cropley writing in the early sixties refers to a division between behavioural psychologists and cognitive scientists. The first maintain that creativity is based on the association of stimulus and response: the creative person makes unusual responses to given stimuli. Cognitive scientists, however, see creative thinkers as people who handle information in ways that produce novel combinations. They toy or play with ideas and objects for the sake of it. They actively manipulate the messages from the environment. For a behaviourist, thinkers are much more passive. Fundamental though this division might
be, it is too fine for a robust pragmatist like de Bono. He begins a paragraph in his
discussion of how creative thinking can be blocked by saying (p 232):

One is blocked because there is nothing in the way....Here a particular way of
looking at things leads one straight past a better way of looking at them.

This seems to suggest a behaviourist interpretation of blocking: a particular stimulus
elicits a given response. In the next-but-one paragraph he says:

If things are put together in a certain way to give one pattern then this prevents
them being put together in another way to give a different pattern.

A comparison of both statements shows that they make different assumptions about how
the mind works. The second is indeed cognitive in its approach, accepting thinking as an
active or re-inventing process, but a process which often becomes hackneyed. The first,
even in its imagery, shows the thought process being led. The juxtaposition of these
contradictory views does not seem to bother de Bono in the slightest, even though he
makes some commitment to a passive mind on the rare occasions he philosophises about
thought. One feels that de Bono cannot be bothered about subtle distinctions - no
concerns about angels and pinheads for him. He has the scent of a practical thinking
primer, and he follows it doggedly. This refreshing approach stands him in good stead
when he considers the nature of pattern-making and association, just as it did when he
considered ideation.

The Act of Juxtaposition

De Bono sees the creation of new patterns from the dismembered pieces of the old, as
well as any other material accumulated, as both essential to creativity and to his
particular method. As will become obvious shortly, this creation depends heavily upon
Paradoxes in Decision-making  
Living with Paradox

inconsistency: the juxtaposition of often opposing items. Once the new patterns are created by lateral thinking, then we can depend upon the innate tendency to vertical thinking to generate implications and applications based upon them. He says (p 259):

Lateral thinking is used to restructure the perceptual pattern which is the way a situation is looked at. Vertical thinking then accepts that perceptual pattern and develops it. Lateral thinking is generative, vertical thinking is selective. Effectiveness is the aim of both.

In a nutshell, de Bono accepts that we have minds hard wired to see the potential in patterns. We use this feature, hijack it almost, when working creatively, by feeding it unusual patterns to start with. Given the importance of these new patterns in his scheme, and given that he has many suggestions for breaking up the old patterns, he is decidedly reticent on schemes for creating new patterns: for recombining the shattered bits of the old patterns. Effectively, he suggests that any combination is acceptable providing it isn't the original one. Although ways of achieving this are not exhaustively or systematically pursued, certain basic techniques can be gleaned from an overall reading of the text. As with the techniques described above which are used to break down old patterns and acquire extra material, De Bono does not see the act of combination and its ensuing creativity as the sole prerogative of the individual. Many of his examples and exercises take place in groups, showing that combination is as much a meeting of minds as a meeting of items in an individual mind. Following my earlier lead, therefore, it might be appropriate to think of the group as a conducive environment for the act of combination rather than as a technique in its own right.

Addition

This is simply the combining of items, and is similar to Lull's technique for generating new ideas. Here de Bono suggests a continuing combination of items until a particular combination sparks a development from the vertical thinking mechanism. In particular,
several of his reflexive techniques, such as reversal, which are used, by elaboration, to generate new material, can immediately be extended to combination. So, in the illustration I used earlier, the idea of a policeman directing traffic can be immediately combined with the idea of the traffic directing him.

**Bizarre combination**

Although the use of bizarre combinations can be seen simply as a form of addition, de Bono seems so fond of it that he seems to see it as a separate technique and a way of making particularly insightful associations. In many ways, the bizarre combination stands as a paradigm for de Bono's whole approach. Suggesting that we think of a human with a nose on its leg, and then extrapolate the idea, is not only the essence of his approach, but the opposite of the sort of sensible, vertical thinking which he suggests we should deny. By advocating this deliberate search for the zany, he is advocating our deliberate denial of our own natural thought patterns. Certainly, when attempting thinking skills with my own classes, there seems to be a built in resistance to bizarre combination. I feel that de Bono is saying that the achievement of the grotesque combination is a sort of Nirvana for his approach.

**Random combinations**

De Bono does not especially advocate the use of random addition, but its use is implied by his emphasis on the bizarre. Effectively, there is no easier way of gaining novel combination that taking a series of attributes, ideas or objects and pairing them by the use of random numbers. Some of the techniques I have used with children and described above - the Misfits program, the creation of a new invention, the solving of a problem like finding your way in the fog - use random combination as an essential part of their action. Almost invariably, random combination produces an inconsistent pairing.
The varied choice of entry points

Finally, de Bono sees the direction from which we attack material as a way of providing a novel combination. His reference to the detective novel, noted above, is particularly appropriate here. If, for example, we entered the detective novel when the murderer was first planning his crime, there would be no puzzle for us to solve. We would have the answer. Because the detective fiction writer obliges us to enter later, traditionally just before the murder, we are forced to piece together the previous events from later information. The change of the entry point, therefore, forces us to look at the crime from an angle different from a simple narration of events. Likewise, a particular set of tasks might be put together into a timed and developing sequence by starting from the end and working backwards rather than starting from the beginning and working forwards. This is particularly appropriate where an end deadline is known, and it is necessary to work out the time to begin.

I have tried to show that de Bono, in a no-nonsense manner, attempts to address the questions raised by the act of creation, and to show in part that creativity can be trained. I have further tried to show that the tension and ambiguity generated by paradox can both in a natural and artificial way stimulate creativity and creative scenarios. Certainly, creativity sees inconsistency as perfectly rational. Given that creativity does depend upon the combination and association of ideas, and that the human brain is capable of trawling the insights that are generated for usefulness, it is tempting to ask if there are not more systematic or methodical ways of generating the associations. After all, time marches on, and the techniques the brain developed in the veldt six hundred thousand years ago for the survival of the species may no longer be appropriate to the information crunching world of the third millennium. Given the vast computing power we now have at our disposal, might it not be possible to make all the possible combinations of ideas, select the most useful by some consistent and definable means, and realise the dream of Lulls and the Cabalists to know everything about everything? Put another way, isn't creativity
too hit and miss for the modern world, even given the efforts of de Bono and others to make it more comprehensible and rigorous?

This issue was considered at length in *Labyrinths of Reason*, a recent book by William Poundstone. He asks us to consider a foolproof procedure by which we might combine ideas. The procedure allows ideas to be combined, providing they do not contradict each other. It is simple in this way to combine two ideas. For example the propositions:

1. *All hay is brown*
2. *Hay is grass*

can be fruitfully combined, and deductions drawn (created) from them. A third statement:

3. *All grass is green*

cannot be added because it contradicts the first two. This checking for consistency is the step which guarantees the validity of the combinations. As the list expands, this internally consistent set of combinations will eventually generate all knowledge. Although simple unchecked concatenation might be useful in de Bono's sense, in that the forced associations may produce useful creative insight, it is important to understand that we are trying to eradicate the need for this unpredictable creative leap. Only a consistent list allows all useful knowledge to be generated.

Unfortunately, the act of checking quickly runs into trouble. For example, suppose we substitute the statement:

4. *All fire engines are red*
Paradoxes in Decision-making
Living with Paradox

it can be added to the first two without contradiction. If we now add proposition 3, it only becomes inconsistent when compared with the subset 1 and 2: the other combinations (1,4 and 2,4) admit it. As the list of propositions lengthens, so the number of subsets with which comparison needs to be made increases. For example, in a list of three items, there are seven subsets. In a ten item list, there are 1023 subsets. To add a 101st item to a list of 100, $10^{30}$ comparisons must be made. Obviously, the number of comparisons being made requires an extensive and powerful calculating device. Poundstone proposes such a computer. Each independently acting component of the computer is an atomic proton. Further, every proton in the universe is dedicated to the computer, and messages pass between them at the speed of light. Finally, the computer is allowed to run for the entire lifetime of the universe. Poundstone undertakes that such a machine could complete $10^{186}$ independent operations. If each of these operations consisted of one check of an item in a list against a particular subset, then the total number of operations available would be sufficient to construct a self-consistent list (in the terms described above), of just 558 items! Even though the items will be more profound and important than the 'all hay is brown' variety, 558 items is evidently totally inadequate to capture all human knowledge. Given that 558 is the longest self-consistent list that can possibly be compiled, then the lengths of self-consistent lists that we can actually compile are much, much shorter. Our attempts to combine data items into meaningful items of knowledge merely skim the surface of all the possible combinations that could be made. It is quite likely, for example, that a cure for cancer or a design for a light-speed drive already exists within the gamut of our present knowledge but the correct connections between the numerous disparate bits of the cure or design have not yet been made. Given that the amount of human knowledge is burgeoning, and that connections and associations are the key to creation and progress, Poundstone's bleak logic suggests that creativity, hit and miss though it might be, is actually the optimal procedure. I hope I have shown that paradox can be a key ingredient of it.
DISSONANCE REDUCTION

A hungry fox tried to reach some clusters of grapes which he saw hanging from a vine trained on a tree, but they were too high. So he went off and comforted himself by saying: 'They weren't ripe anyhow.'

Aesop, Sour Grapes

In his book, *Logic and Society*, Jon Elster proposes that the *state of nature* described by Hobbes is, in fact, the Prisoner's Dilemma. He argues thus. First, Hobbes' society is based entirely on selfishness. This selfishness can be equated with the defection strategy of the Prisoner's Dilemma. Second, any spontaneous acts of co-operation are doomed, just as they are in the Prisoner's Dilemma. Third, Hobbes does not agree with or advocate this primitive society. Life, after all, is 'solitary, poor, nasty, brutish and short'. And in order to escape this natural predicament caused by wholesale defection, man must learn to co-operate. This is the basic lesson of the Prisoner's Dilemma. Whether we advocate Elster's interpretation, we can certainly advocate, as a way out of the Prisoner's Dilemma, Hobbes' solution to the state-of-nature problem: that is, we can advocate the formation of government. Where necessary, sovereign power can eradicate the Prisoner's Dilemma simply by making the defection choice less palatable than the co-operative. For example, if people are punished for walking on the grass, they will stick to the pavement even though it means a longer journey. In fact, as I hope to show, the application of law and order is but one way of achieving results by encouraging dissonance reduction.

This concept was first proposed by L Festinger in his book, *A Theory of Cognitive Dissonance*. Dissonance reduction operates in situations where we are psychologically discomfited, like the fox in the epigraph to this section. That is to say, where we are exposed to contradiction or inconsistency, we aim to reduce its effect: to smooth it away. It is, I would contend, as much an outcome of inconsistency as creativity is. If we deny the rationality of inconsistency, then we deny the need both for creativity and for
Paradoxes in Decision-making

Living with Paradox

dissonance reduction. I hope to show that dissonance reduction not only exists but like creativity is broadly rational and useful. If creativity harnesses paradox and inconsistency, dissonance reduction side-steps or dampens it. It operates in an opposite way to lateral thinking, by reducing options and emphasising serial thought. Just as it is easy and fashionable to enthuse about creativity, so it is as easy to revile dissonance reduction. However, this approach to both is not appropriate. Continued creative action can soon be perceived as either re-inventing the wheel or as breaking new ground instead of consolidating gains into solid achievement. Dissonance reduction can, alternatively, be seen as one means of acquiring certainty and sense. Neither need dissonance reduction be necessarily an uncreative act. An example, taken from Robert Martin's book cited earlier, will explain what I mean. He describes the well-known puzzle of the man and his son who are involved in a car crash, and end up at different hospitals. As the son is being prepared for surgery the surgeon appears, looks aghast and says: "I cannot operate on this person - it is my son." We are asked to explain this. Thinking creatively we could provide a number of solutions: the father makes a quick recovery and returns to his work as a surgeon; the surgeon mistakes the boy for his son; the crash sends the boy into a parallel universe; one father might be a step-father. Dissonance reduction demands we solve the problem because the puzzle causes us mental discomfort. Further, we know that it is a puzzle, and therefore has a simple but non-obvious solution. This need to apply Ockham's Razor, to find the elegant solution rather than any solution, creates further cognitive dissonance. We know, consequently, that we are compelled to solve the problem as it stands: that the boy is indeed the son of the surgeon who is not the father of the boy. We conclude that the surgeon, therefore, is his mother. This surprising result, the surprise being based on the stereotyping we apply to jobs, is lost if we apply the creative approach. This approach, with its numbers of solutions and avenues for exploration gradually demystifies the problem, so that it becomes almost bland and mundane. I described earlier my work with the Misfits program, where children had to explain the reasons for a person wearing an odd assortment of clothes. Instructive and enjoyable though it was, it had exactly this consequence of removing the mystery,
humour or piquancy from the situation. The occurrence of the person in miner's helmet, policeman's jacket and waders became explicable, commonplace and boring. There is something to be said, therefore, for the appropriate use of conventional thinking.

Festinger first describes dissonance reduction, its roots, and then goes on to give examples of it in operation. Dissonance arises where individuals are unable to reconcile their actual actions or beliefs with what they want in theory. So a smoker experiences dissonance if he wants to give up smoking. If these wants could be achieved then there would be no dissonance. It is the fact that they cannot be achieved that causes the problem. We have seen frequent illustration throughout this dissertation of the paradox of wanting, particularly of willing what cannot be willed; of obtaining something desired by not desiring it. Dissonance reduction operates within the orbit of this paradoxical state. Festinger proposes several reasons why the want or need cannot be directly achieved, reasons which have been discussed earlier. These include such issues as the direct wanting of by-products, and the acceptance of short term pleasure or easily achievable gain rather the consideration of long term advantage. Likewise, coping with the environment creates problems, particularly where complex social situations are to be manipulated. Again, we considered this issue extensively when we examined the problems involved in making wilful decision within the parameters of, for example, the Prisoner's Dilemma. In fact, incidents described throughout this dissertation tend to suggest that dissonance reduction is often resorted to in times of difficulty.

The sorts of responses to dissonance proposed by Festinger are likewise redolent of the sorts of responses to paradox we have encountered, some of which have been expressly related to dissonance reduction in the course of this dissertation. Rather than repeat these in detail, it will suffice to say that Festinger suggests three basic responses. First, we can attempt to change our own behaviour. In this way, our needs or wants can be aligned with what we actually obtain. In the situation proposed by Elster at the start of this section, for example, we might choose to obey a sovereign power and its laws, to put up
with coercion and loss of independence, in return for security and peace. By using this approach the Prisoner's Dilemma, and the stress it causes, can be circumvented. Second, we can attempt to change a dissonant environment. If we are not happy with the behaviour of children in school, for instance, we can introduce the sorts of rules discussed in the previous chapter. Finally, we can introduce other elements into area of dissonance. These might aim at supporting the act which is causing dissonance: to justify it. Smokers might, for example, look for the benefits of smoking. Similarly, we might look for elements which show the need for dissonance: to explain and excuse it. A smoker might suggest, thereby, that there is nothing really wrong with smoking. She feels uncomfortable because she is treated by others as a social outcast. So by a combination of side-stepping, selective attention and self-delusion we are able to cope with the inconsistent, ambiguous or paradoxical situations we encounter. I believe that the propensity for such behaviour is extensive and rational, given the kind of world we live in. More importantly, if we accept this world and these responses, then we have to accept that we are always in situations which are essentially unique, where individuals resolve their own personal dissonance in their own personal way.

CONCLUSION

I have tried to show in this chapter that the role of the individual in decision-making is central. Each decision made by any of us is a blend of individual style and response, acting upon a situation which is itself unique, and perceived differently by all. The common view of this milieu is that the useful decision is creative and original, and I have tried to show that this is true. Equally, however, decisional situations can be rendered manageable by an attempt to ignore or avoid them, or to deal with them in tangential ways. This process of dissonance reduction, I believe, is equally acceptable and successful. In fact, I would argue that both are simply a response to the many incomprehensible situations in which we find ourselves: decision-making equivalents of fight or flight.
A phenomenon noticeable throughout history regardless of place or period is the pursuit by governments of policies contrary to their own interests. Mankind, it seems, makes a poorer performance of government than of almost any other human activity... Elsewhere than in government, man has accomplished marvels.

Barbara Tuchman, *The March of Folly*

I do not intend to summarise this dissertation at length, since the several points I have wanted to make are contained within it. Equally, these points can lose their force in a work of this size, so some drawing together of ideas is necessary. In this brief summary, therefore, I will content myself with some brief conclusions about the nature of management in schools, and the challenges for management styles that my work implies. The quote from Tuchman above emphasises again the intractable nature of management, a theme that has recurred throughout this dissertation. Further, the title of the book from which it is taken emphasises the role of folly in this, it is the role of folly which I question. Although human frailty affects management, I do not believe that it can so decisively affect it, especially given that so many other fields of human endeavour have prospered. I have tried to show in explanation of this how the means of dealing with problems and taking decisions stand or fall by our ability to understand the difficult environments in which decisions are often made, and the nature of what constitutes a rational decision. I considered these matters extensively in Chapters 1 to 3, and do not intend to dwell on them further. I tried to show in Chapter 5 that creative thought and dissonance reduction are two successful responses to difficulty decisional environments. Game theoretical thinking, on the contrary, is not a successful approach. Creative thought and dissonance reduction accept that items which contribute to a decision - information, individuals, objectives, constraints, and so on - are often ambiguous and
inconsistent. Game theory, on the other hand, assumes complete consistency and perfect knowledge. My paradigm for inconsistency, imperfection and ambiguity is paradox. Because paradoxes are a product of the logic of situations, they cannot be eradicated by harder work, redefinition or sounder thinking. This inherent intractability of situations means that approaches to decision-making and rationality which assume perfection and perfectibility will tend to founder.

Of the paradoxes I have discussed several are of particular significance in that they cause problems for both management and managed. These paradoxes are:

The Sorites Paradox;
The Slippery Slope;
The Prisoner's Dilemma;
Newcombe's Paradox;
Willing What Cannot Be Willed;
Level Shifting;
Self Binding;
By-products and Collateral Outcomes;
The Paradox of Perfect Information;
Arrow's Theorem;

and I would argue that some knowledge of them and their repurcussions is essential to organisations and their management. Rather than look to coercion, control or systemic solutions to the problems posed by these paradoxes, I would argue for a solution in the spirit of this work. Both dissonance reduction and creativity imply individual responses to situations, where solutions vary from person to person. I would argue further, as I did in Chapter 4, that it is appropriate for management to encourage individual responses, to empower rather than dampen or control. Managers should aim to ensure that these responses should mesh and support each other, rather than attempt to set up systems or
hierarchies which proscribe or channel. The only real place for coercion is upon the managers themselves, to ensure that they operate within the system, and not around or above it.

This style of management is perfectly illustrated by the incident in the Peloponnesian War referred to in Chapter 3. The Athenian forces were facing annihilation by the Spartans and their allies at Syracuse. Nicias, the Athenian general, fearing for the situation, called each of his commanders to him in turn, and tried to bolster their morale by appealing to their honour, their patriotism, and the responsibility they held for the good names of their families and ancestors. Thucydides continues:

these words of encouragement seemed to Nicias himself to fall short of their mark and to be no more than barely adequate to the occasion. After speaking to them, he went back and led the infantry to the sea..... (VII, 69)

This incident well illustrates the relationship between empowerment, mutual support, and direction. In his final encouraging of his commanders Nicias tried to empower them, to give them all that he personally could to help them in the coming battle. During the battle, by mutual empowerment, they would attempt to do their best in as creative a way as possible. This ability to strive and succeed would be enabled by Nicias' battle plan, which would both direct and co-ordinate their efforts towards success, and indicate clearly the price of failure. The section quoted, however, sums up for me the central dilemma of management: Nicias, having done all he could do, now had to leave others to do what they had to do. For all their preparatory work, managers lose control of the actual actions of an organisation when it begins to undertake these actions. In a school, after all, it is teachers who actually teach the children, not managers. I believe that in the last several years educational management - DFE, LEA, School, Advisory service - has attempted to formalise and institutionalise the processes of education, in order to try to standardise and control the quality of service provided by teachers. Unlike Nicias,
managers have felt that in some way they can fight the battle for their soldiers. Unfortunately, I feel that this programme of formalisation has tended to create the sorts of problems I have described throughout this dissertation. Instead, I have tried to argue for empowerment, where decision-making, and the creative leap are made by the actual enablers and actuators of the decision. Where problems arise from the paradoxes of individual action, such as the Prisoner's Dilemma, networking or mutual empowerment would appear to provide a way forward. Here, the desire to freeload is mitigated by the knowledge of other people's circumstances, of how people can support and be supported, and of everyone's general place in the order of things. Finally, management's role in offering clear guidance and direction where appropriate provides both the framework for action and the necessary compulsion to action: a fertile ground for creativity and dissonance reduction.

In the end, I am arguing for a community where individual professionals have access to all the system and its information, to use in ways they feel will achieve a clear set of goals which they believe in and have a need to support. Such a body would not be easy to predict and, hence, almost impossible to control. It would grow like an organism, and change as the individuals within it changed: as they developed, as they responded to each other or situations, as they left or arrived. Within it there might be boredom, bad faith and disappointment, but there would also be a flexibility and durability which would deal with problems as they arose. If these problems were avoidable, then the flexible, organic approach might not be appropriate. I have tried to show in this work that, on the contrary, problems are not the detritus of organisations; ephemera which do not exist where reason and commitment do. Instead they are the norm, and paradox makes it so.
Bibliography


__________ *Logic and Society*, Chichester, Wiley, 1978


__________ *Sour Grapes*, Cambridge, Cambridge University Press, 1979


__________ *Metamagical Themes*, Harmondsworth, Viking, 1985


HOLMES, E, Life of Mozart including his correspondance, in VERNON, P E (Ed.), *Creativity*, Harmondsworth, Penguin, 1980
HOYLE, E, Micropolitics of education, in WESTOBY, A (Ed.), *Culture and Power in Educational Organisations*, Milton Keynes, Open University, 1988

JUVENAL, *Satires*


MARTIN, R M, *There are two errors in the the title of this book*, Peterborough Ontario, Broadview Press, 1992

MCCLENNAN, E, Prisoner’s dilemma and resolute choice, in CAMPBELL, R and SOWDEN, L (Eds.), *Paradoxes of Reality and Co-operation*, Vancouver, University of British Columbia Press, 1985


MITCHELL, M, *Gone With The Wind*, London, Pan, 1974


ORMELL, C, *Some Varieties of Superparadox*, Norwich, University of East Anglia, 1993


POINCARE, H, *Science and Method*, Thomas Nelson


__________ *Prisoner's Dilemma*, Oxford, Oxford University Press, 1993


__________ Game theory and the study of ethical systems, *Journal of Conflict Resolution*, XII No. 1, Mar. 1968


__________ *The Lady and the Tiger*, Harmondsworth, Pelican, 1983
What is the name of this book?, Englewood Cliffs, Prentice-Hall, 1987


SUTHERLAND, S, Irrationality, the Enemy Within, London, Constable, 1992


THUCYDIDES, The Peloponnesian War, Harmondsworth, Penguin Classics, 1979

TORRANCE, E P, Education and the Great Potential, Minneapolis, University of Minnesota Press, 1963


WALTON, D, Slippery Slope Arguments, Oxford, Oxford University Press, 1992

WASON, P C, and JOHNSON-LAIRD, P N (Eds.), Thinking, Cambridge, Cambridge University Press, 1977