BAYESIAN MODELS AND REPEATED GAMES

by

Simon Christopher Young, B.Sc.

This Thesis is submitted for the degree of Doctor of Philosophy at the University of Warwick

Department of Statistics
University of Warwick
Coventry

September 1989

TABLE OF CONTENTS

SUMMARY vi

1. INTRODUCTION 1

2. DISCUSSION OF VARIOUS TYPES OF GAMES

2.1 Types Of Games That We Are Considering 4

2.2 Solution Concepts and Classifications of Games 7

2.3 An Experimental Study 12

3. GAME THEORY LITERATURE

3.1 Introduction 16

3.2 Game Theory Literature 16

3.3 Experimental Game Literature 22

3.4 Conclusions 32

4. THE INFINITE REGRESS

4.1 Introduction 34

4.2 Example: Nuclear Disarmament 34

4.3 The Theory Of Metagames 36

4.4 Belief-Rational Strategies 41

4.5 Truncation of the Regress 44

4.6 Two Examples 54

4.6.1 A Prisoner's Dilemma Game 54

4.6.2 A Competition 57

4.7 Stability of the solutions 59

4.8 Conclusions 62

5. A REVIEW OF BAYESIAN GAME THEORY

5.1 Introduction 64

5.2 Review 64

5.3 Conclusions 76

6. REPEATED EXPERIMENTAL GAMES

6.1 Introduction 79

6.2 Assumptions and Notation 80
6.3 Bayesian Rationality in Repeated Games 82
6.4 A Pragmatic Solution To Modelling Games 88
6.4.1 An Advertising Model 88
6.4.2 A Bidding Problem 89
6.5 Rationality and Calibration in Symmetric Games 91
6.5.1 Example 1 — Repeated games with known utilities 93
6.5.2 Example 2 — A Model for the Prisoner's Dilemma 96
6.5.3 Example 3 98
6.5.4 Example 4 101
6.6 Diagnostics, Robustness and Beyond 102
6.6.1 Diagnostics 102
6.6.2 The stability of calibrated models and the efficacy of A1 & A2 103
6.6.3 Dominated societies and stability for rational games 103
6.7 Conclusions 104

7. DERIVING OPTIMAL STRATEGIES
7.1 Introduction 105
7.2 Wilson's Algorithm 106
7.3 Form of the Optimal Solution 109
7.3.1 Case when *p* is known 115
7.3.2 Case when *p* is unknown 118
7.4 Determining Optimal Play 118
7.5 Extensions 122
7.5.1 Discount Factor 122
7.5.2 Other Games 123
7.6 A Further Example 126
7.7 Testing the Appropriateness of the Model 129
7.8 Conclusions 130

8. SUBOPTIMALITY OF M–STEP BACK STRATEGIES
8.1 Introduction 131
8.2 Notation and Assumptions 132
8.2.1 Notation 133
8.2.2 Assumptions 134
8.2.3 Example 1 136

8.3 Results 139
8.4 Extensions and Implications 144
8.4.1 Extensions of the theorem 144
8.4.2 Example 2 146
8.4.3 Implications of dropping the calibration assumption 150

9. FURTHER RESEARCH 152

10. CONCLUSIONS 162

REFERENCES 165

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Jim Smith, for his constant help and guidance. He was always full of ideas and enthusiastic about all aspects of my work, and it was enjoyable working with him. I would also like to thank the staff and research students of the Statistics department, particularly Gareth, Duncan and Jo, for helpful discussions and for providing a stimulting atmosphere during my three years there. Thanks are also due to Dennis Lindley for showing an interest in my work, and pointing out a few problems.

I would like to thank my parents for constant support in all respects during my studies. Special thanks are due to Karen for her understanding and encouragement, and for always being there when I needed her.

This research was supported by a Science and Engineering Research Council studentship.

DECLARATION

I declare that this thesis is entirely the result of my own research during the past three years apart from sections 6.3 and 6.5 which are the result of joint work with Dr. J. Q. Smith.

Personal Pronouns

Rather than having to refer to "he or she" and "his or her" throughout this thesis, all singular personal pronouns are male. This is not in any way meant to be a sexist statement, but merely a reluctant way of avoiding awkward and long-winded phrases. I hope that nobody is offended by this.

To Karen

SUMMARY

A game is a theoretical model of a social situation where the people involved have individually only partial control over the outcomes. Game theory is then the method used to analyse these models. As a player's outcome from a game depends upon the actions of his opponents, there is some uncertainty in these models. This uncertainty is described probabilistically, in terms of a player's subjective beliefs about the future play of his opponent. Any additional information that is acquired by the player can be incorporated into the analysis and these subjective beliefs are revised. Hence, the approach taken is 'Bayesian'.

Each outcome from the game has a value to each of the players, and the measure of merit from an outcome is referred to as a player's utility. This concept of utility is combined with a player's subjective probabilities to form an expected utility, and it is assumed that each player is trying to maximise his expected utility. Bayesian models for games are constructed in order to determine strategies for the players that are expected utility maximising. These models are guided by the belief that the other players are also trying to maximise their own expected utilities.

It is shown that a player's beliefs about the other players form an infinite regress. This regress can be truncated to a finite number of levels of beliefs, under some assumptions about the utility functions and beliefs of the other players. It is shown how the dichotomy between prescribed good play and observed good play exists because of the lack of assumptions about the rationality of the opponents (i.e. the ability of the opponents to be utility maximising). It is shown how a model for a game can be built which is both faithful to the observed common sense behaviour of the subjects of an experimental game, and is also rational (in a Bayesian sense).

It is illustrated how the mathematical form of an optimal solution to a game can be found, and then used with an inductive algorithm to determine an explicit optimal strategy. It is argued that the derived form of the optimal solution can be used to gain more insight into the game, and to determine whether an assumed model is realistic. It is also shown that under weak regularity conditions, and assuming that an opponent is playing a strategy from a given class of strategies, S, it is not optimal for the player to adopt any strategy from S, thus compromising the chosen model.

1. INTRODUCTION

A game is defined to be a model of a group of players, each of whom is required to choose a move from a set of possible moves. The outcome for a particular player depends not only upon his choice, but also on the choices of all the other players of the game. There is therefore an interdependence between the players of such a game, and this is what makes the subject fascinating to study, but also complicated. Game theory is the collection of solutions to these games, and has concentrated on either the search for equilibria in a game, or finding the optimal actions for one particular player of the game. We shall concentrate on the latter of these.

In trying to find an optimal action for a player, we need to determine our beliefs about the actions of the other players. It is assumed that these actions are unknown to the player under consideration, or at least that he has some uncertainty about them, otherwise the problem is trivial. In one player games (against nature) the problem is to determine an optimal action against a random process. However, when considering games with two or more players (as we do here), we have the problem of uncertainty as opposed to randomness.

The approach that we will take here is to model this uncertainty probabilistically. These probabilities are assumed to be determined subjectively by the players. This approach is referred to as the 'Bayesian' approach, after the Reverend Thomas Bayes, who showed that for events A and B, the conditional probability

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(B|A)\mathsf{P}(A)}{\mathsf{P}(B)}.$$

This result makes it possible for extra information to be incorporated into the model, and therefore the subjective probabilities of a player can be updated when he receives more information. For a detailed formulation of Bayesian statistical decision theory, see DeGroot (1970).

We then assume that for each outcome of the game, each player can define a utility — a numerical value that describes the desireability of the outcome. These two concepts of subjective probability and utility are combined to form expected utility. For an excellent discussion of the use of expected utility to make decisions under uncertainty, see Lindley (1985). From the expected utility function we can define Bayes optimality, where a player is assumed to be maximising his expected utility, and also Bayes strategies, which are courses of actions that achieve the goal of expected utility maximisation.

Throughout this thesis we shall consider the ability of the opponents in a game, and shall mainly consider them also to be utility maximising (or 'rational'). A truely subjectivist approach would assume that the subjective probabilities already incorporate all such information about opposing players. However, this would trivialise the theory to a simple maximisation problem. How the player's beliefs about the rationality of his opponents are incorporated into the problem is a difficult and interesting problem that makes game theory such a stimulating topic. These arguments are developed in chapter 6 below.

Also it is debatable whether a player should model his beliefs about his opponents' actions on what a player *should* do, or on what players have been *observed* to do. There is a vast literature that suggests that the two approaches do not give the same results. Our approach is to determine what the opponents ought to do, but this is determined from beliefs about how people have been observed to play in the past. So, a player is assumed to take into account how his opponents are likely to actually play, in order to determine how he (the player) ought to play. So, in some respects, we are incorporating both approaches to determine an overall optimal approach.

We will incorporate the above features into the Bayesian models of games that we construct. From these models we will be able to determine optimal play for a particular player of these games. We shall discuss the approaches that have been considered in the past, and then extend these approaches, and prove results to show how these approaches can be improved upon. I shall now outline the areas that are covered, and the results that are obtained in the following chapters of this thesis.

Chapter 2 discusses the types of games that we shall be considering. We briefly consider the different types of games that have previously been analysed in the literature, and then state which of these we are going to explicitly concentrate on. We also state the general assumptions that we shall be making about the basic parameters of the games. To illustrate these points, an example of an experimental game is provided.

Chapters 3 and 5 provide a review of the extensive literature on game theory and experimental gaming. These literatures are extremely cross–disciplinary, and I apologise if I have missed any pertinent references. Chapter 3 is divided into two sections — the first section reviews the traditional game theoretic literature, and the second reviews the experimental gaming literature. Chapter 5 reviews the Bayesian game theory literature, and it is this area that is most related to the work in this thesis.

Chapter 4 investigates the 'infinite regress' that occurs in repeated games with incomplete information. Earlier work of another author (Howard) is discussed and then this work is generalised by a novel approach. This new approach is also shown to relate to previous work by other authors. By adopting this approach, the conditions that are required to truncate the infinite regress can be determined.

Chapter 6 considers the discrepancy between the dictates of traditional game theory and the results of experimental games. Bayesian models for such games are developed, and it is shown that the observed discrepancy exists because the traditional models do not have some necessary features. The following chapter shows how a known algorithm for calculating optimal next moves can be improved by the knowledge of the form of an optimal solution for a game. It is illustrated how this form of the optimal solution can be found, and how it enhances the algorithm. From the resulting approach we can not only determine how a player should play on all subsequent moves of the game, but we can also discuss the appropriateness of the assumed model.

In chapter 8 we consider a class of strategies for a game that, at any time point, depend only upon the previous m move pairs. We show that (under very unrestrictive conditions) when it is assumed that an opponent is playing such a strategy, it is never utility maximising for the player to play such a strategy himself. In chapter 9 we consider some areas of future research that there has been insufficient time to cover in the main body of this thesis. In chapter 10 we draw some conclusions from the work presented in this thesis.

2. DISCUSSION OF VARIOUS TYPES OF GAMES

2.1 Types Of Games That We Are Considering.

Before proceeding to analyse various games and the solution concepts that are applied to these games, we must first define the types of games that we are considering. The game theoretic literature has produced a vast array of different kinds of games and ways of representing them. We shall limit ourselves to only some of these.

It is common in the literature to dicuss various games in terms of their *extensive form* representation. This is a representation of the game by means of a game tree, where the vertices correspond to the choice points for the players and the branches represent the options open to the players. The terminal points of the tree give the outcomes of the game. Also information sets are given on these trees, which are the sets of vertices that a player cannot distinguish between when he makes his move. These information sets are not required if the game has *perfect information*, i.e. all previous moves that have been played are known by all players. Therefore games such as Chess and Backgammon have perfect information, but not Poker. Also, in all games in extensive form, the concept of a subgame can be defined. This is the game defined on the portion of the game tree starting at any point in the original game tree (other than a terminal point), and consists of all points and branches that can be reached from this given starting point. For a fuller description of this representation of games see Thomas (1984).

However, any finite game (i.e. each player only has a finite number of choices at each time point) in extensive form can be reduced to its *normal form* without losing any information. The normal form of a game is simply a rectangular array of numbers that form a *pay-off matrix*. Each row in this matrix represents a possible move for a player P_1 and each column represents a possible move for another player P_2. The entry in the matrix corresponds to the outcome of the game for the respective row and column choices by the players, and these choices are made simultaneously. The games that we are considering are also non-cooperative. By this we mean that no binding contracts or commitments can be made by the players, and the only communication permitted is through the moves played. Work has been carried out on various relaxations of this assumption, to determine the effect of communication, side-payments, threats, commitments, etc.

Consider the normal form of the game determined by the pay–off matrix given in Figure 2.1.1.

$$\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 1 & 2 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} (1,1) & (-1,2) \\ (2,-1) & (0,0) \end{pmatrix} \end{array}$$

Figure 2.1.1

The first entry in each outcome vector corresponds to the pay–off to P_1 and the second entry to P_2. Occasionally only one entry in the matrix is required, due to symmetry or because there is a relationship between the pay–offs to the players, but this will be made obvious by the context of the example. We can see from Figure 2.1.1 that if both players make move 2 then they will both receive a pay–off of 0, whereas if P_1 made move 1 when P_2 made move 2, then P_1 would receive a pay–off of -1 and P_2 would receive a pay–off of 2. The normal form is therefore an extremely simple representation of the game, and we shall concentrate on this particular form.

The games that we are considering are almost exclusively two player games. Now this is obviously a major restriction on the games that could be considered. However, a lot of the ideas that we develop can be extended to games involving $n \geq 3$ players, although the notation becomes very messy. If the game being considered explicitly involves more than two players, then the normal form of that game is less appropriate because many sets of matrices are required. In this case the extensive form representation is more appropriate, but obviously the game tree becomes more complicated. So as we are mainly considering two player games we shall concentrate on the simpler normal form representation, and the players will generally be labelled P_1 and P_2, as in Figure 2.1.1.

As we shall discuss in the next chapter, the theory of games was first developed for *zero–sum* games, i.e. games where if P_1 received a pay–off of q from a particular outcome of the game, then P_2 would receive a pay–off of $-q$. Because of this it is only necessary to give the pay–off to one player (by convention P_1) in the pay–off matrix. We are not going to restrict ourselves to such games. Indeed most of the games that we consider will be non–zero–sum. The theory that we shall develop will be applicable to these more general games and then zero–sum games can be considered as a special case. Because of this distinction, a lot of the results developed in the early literature are applicable to zero–sum games, but not to non–zero–sum games. Care must obviously be taken when applying results from the specific case to the more general case, and careless generalisations can produce inaccurate answers.

We shall consider the pay–off matrix to determine just that — the pay–offs. There seems to be some controversy as to whether the matrix determines the utility that a player gains from a particular outcome, or whether it simply determines the pay–off. If the matrix does represent the utilities obtained, then this must take into account the complete utility from the given outcome, including the outcome to the opponent and all aspects of the outcome in the given game context. This will be achieved very rarely by a simple pay–off matrix, and different people will then require different pay–off matrices. Also, this would appear to invalidate all experimental studies as these provide pay–offs as the outcomes of the games, and do not allow for any other ways that a player may obtain utility (e.g. receiving a higher pay–off than his opponent, receiving the highest overall pay–off, etc.).

Having to determine the utility function of the players then appears as a problem with the approach that we are advocating. It is often assumed in the literature that utilities are linear on pay–off (with perhaps a discount factor that discounts future pay–offs), but experiments very rarely confirm this to be realistic. Therefore players must be allowed to have more general forms of utility function. This is discussed further in chapter 6 of this thesis. Determining one's own utility function can be a problem, but more of a problem is that of determining an opponent's utility function, which is assumed to be unknown by the player concerned. This therefore makes determining optimal moves a decision problem with incomplete information, as the player is unsure about his opponent's utility function. The player is assumed to have some beliefs about his opponent's utility function, and these beliefs can be used to determine an optimal strategy. If, on the other hand, the player *is* assumed to know his opponent's utility function, then the decision problem is simplified as it becomes a game of complete information. The concept of complete information is distinct from, and should not be confused with, the concept of perfect information mentioned above.

Also, as will be discussed in the next chapter, the classical approach to game theory concentrated on the stationary, one–play game. It was argued that the theory for this should be determined before dynamic multi–play games were considered. In these multi–play games, we call the single game given by the pay–off matrix the *generating game*, and a generating game that is repeated many times is referred to as a *repeated game*. Individual plays of the generating game are referred to as *stages* of the repeated games. The theory for one–play games rarely carries over to multi–play games satisfactorily, and hence a new approach must be taken when considering repeated games. Our theory is directed at these repeated games, and uses the fact that future interactions are likely to occur, to determine how to play the present stage of the game. Having said this, our work can be used in the more specific area of games only played once.

At each stage of this repeated game, each player has a choice between moves, as determined by the rows (or columns) of the pay–off matrix. A *strategy* for a player is a decision rule that determines a move for that player at every stage of the game, that depends on the move sequence only through the previous outcomes of the game. A strategy that determines a specific move at every stage of the game is referred to as a *pure strategy*. Strategies that are probabilistic mixtures of these moves at any stage of the game are referred to as *mixed strategies*. Thus a move will be determined by a mixed strategy by some independent event such as the outcome of a toss of a coin, or the roll of a die. In games with complete information, a player may wish to use a mixed strategy so that his opponent cannot determine the strategy being used, but this may not be necessary for games with incomplete information. In some games, a particular solution concept may determine the optimal strategy to be a mixed strategy, but we have the following result.

THEOREM 2.1. (Harsanyi, 1977, pg. 102)

Let σ be a mixed strategy for P_1 that is a best reply to a strategy σ_2 for P_2 (i.e. no other feasible strategy obtains a higher utility against strategy σ_2). Then each pure strategy used in σ with positive probability is also a best reply to σ_2, and so is any arbitrary probability mixture of these pure strategies.

Mixed strategies can be used to determine equilibria and other solution concepts for various games, as we shall see in the next section.

2.2 Solution Concepts and Classifications of Games.

The main solution concept that has been used in the literature is Nash Equilibria, and is often given as the solution to any competitive game. An equilibrium is a pair of strategies, such that by unilaterally altering his strategy, a player will receive a smaller pay–off. Therefore neither player would have any reason to regret his strategy, if he found out the strategy of his opponent. Nash (see next chapter) proved the existence of such an equilibrium in all of the types of games that we are considering, but uniqueness is not guaranteed. In two player zero–sum games, the equilibria are easy to find by determining the *saddle point* — the element of the pay–off matrix that is the maximum in its row and the minimum in its column. Also in such games, the pay–off to both players from one equilibrium must equal the pay–off to both players from any other equilibrium, but such a result does not carry over to more general games.

Another concept common in the literature is that of a *maximin strategy.* The maximin strategy for a player is the strategy that maximises the minimum possible pay–off to that player. The strategy can be pure or mixed, and this maximum of minimum pay–offs is referred to as the player's *security value.* In two player zero–sum games, P_2 can hold P_1 to his (P_1's) security value by playing his (P_2's) maximin strategy. Therefore the maximin strategies determine an equilibrium point, as neither can unilaterally do better than their security value. In more general games, this result does not hold, as we can see from the following example.

$$\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 1 & \quad 2 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} (2,2) & (3,3) \\ (1,1) & (4,4) \end{pmatrix} \end{array}$$

Figure 2.2.1

Consider the game determined by the pay–off matrix given in Figure 2.2.1. We can see from this that P_1's security value is 2 and is obtained by playing move 1. Also P_2's security value is 3 and is obtained by playing move 2. But the outcome when both players play their maximin strategies is a pay–off to both players of 3, which is higher than P_1's security value. The result is however not an equilibrium, as the only equilibrium for this game is where both players make move 2, which results in a pay–off to both players of 4. This demonstrates that in general cases, the maximum pair (i.e. when both players play their maximin strategies) is not necessarily an equilibrium, and hence in this sense there is no obvious solution concept for the game.

Rappoport & Guyer (1966) give a complete classification of all two player, two move games. Similar studies for games with more moves or players have not been performed due to the large

Game	Pay–off matrix	Name (if any)	Pure Strategy Equilibrium Points
1	$\begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}$	–	(1,1) and (2,2)
2	$\begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}$	Trust	(1,1) and (2,2)
3	$\begin{pmatrix} 4 & 1 \\ 3 & 2 \end{pmatrix}$	–	(1,1) and (2,2)
4	$\begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix}$	–	(1,1)
5	$\begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix}$	Spite	(1,1)
6	$\begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix}$	–	(1,1)
7	$\begin{pmatrix} 3 & 1 \\ 4 & 2 \end{pmatrix}$	Prisoner's Dilemma	(2,2)
8	$\begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}$	Convergence	(1,1)
9	$\begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}$	Chicken	(1,2) and (2,1)
10	$\begin{pmatrix} 3 & 4 \\ 2 & 1 \end{pmatrix}$	–	(1,1)
11	$\begin{pmatrix} 2 & 3 \\ 4 & 1 \end{pmatrix}$	Leader	(1,2) and (2,1)
12	$\begin{pmatrix} 2 & 4 \\ 3 & 1 \end{pmatrix}$	Battle of the Sexes	(1,2) and (2,1)

Figure 2.2.2

Moves in the equilibrium points correspond to the strategies for P_1 and P_2 respectively, as in the pay–off matrix given in Figure 2.2.1.

number of possible games. Figure 2.2.2 presents a classification of the distinct two player, two move, symmetric games. A symmetric game is where the pay–off matrix for P_2 is simply the transpose of the pay–off matrix for P_1. Because of this we have only presented the pay–off matrices for P_1. Also we have only ranked the pay–offs on a scale 1 (lowest) to 4 (highest) as this is sufficient to classify them.

Of these 12 games, eight possess either single equilibrium points, or equilibrium points that strictly dominate (obtain greater pay–off for both players) than all other equilibrium points, and so are of only limited interest. The other four games (numbers 7, 9, 11 and 12) were described by Rapoport (1967) as the 'archetypes' of the two player, two move games, and they have all attracted a lot of interest in the literature. The game that has attracted by far the most interest in the game theoretic and experimental gaming literatures is game number 7 — the Prisoner's Dilemma game (PDG). This game is defined by the pay–off matrix given in Figure 2.2.3, where $C > A > D > B$ and $2A > B + C$.

$$\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 1 & 2 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} A & B \\ C & D \end{pmatrix} \end{array}$$

Figure 2.2.3

The name Prisoner's Dilemma comes from the following anecdote attributed to Albert Tucker. Two people have been arrested and charged with a serious crime. However the police do not have any firm evidence with which to convict them unless one or other of the accused confesses. The prisoners are held seperately and cannot communicate with each other. If neither confess (i.e. both make move 1) then they will both be charged with some minor offence. If both confess (i.e. make move 2) then they will both be convicted and sent to jail for a long time. However, if one confesses and the other doesn't, the person who confesses is set free and given a reward for giving Queen's evidence, whereas the other receives a very heavy jail scentence. So it is better for each prisoner to confess, irrespective of what the other prisoner does, but if both refuse to confess, they both obtain a better outcome than if they were to both confess.

This PDG will be the game that we shall concentrate on mainly in this thesis. We shall refer to move 1 as Cooperation, and move 2 as Defection, in common with the literature on this game. An interesting background to the dilemma presented in this game is given in Rapoport & Chammah (1965). As we stated earlier, we shall be considering repeated games, and all of the concepts discussed above are more applicable to one–play games. However, it would appear that repetitions of the PDG would not produce any different results from those for the one–play game, as one move *dominates* the other.

We say one move m in a pay–off matrix dominates another move m' for P_1, if for all strategies available to P_2, the pay–off from move m is greater than the pay-off from move m'. So we can see that Defection (move 2) dominates Cooperation (move 1) in the PDG as $C > A$ and $D > B$. Also a stronger argument than this exists for using continual cooperation in a repeated PDG. If the end point of the game is known, then Defection on the last move of the game must be optimal. But the opponent is likely to do the same, and as mutual Defection will therefore occur on the last move, Defection must be optimal for the penultimate move. This process (called 'extended rationality') can then be extended back throughout the game. Some sociologists (for example, Hamburger, 1979) find this argument convincing enough to advocate continual Defection at all stages of a repeated PDG. However, as we shall see later, players of experimental repeated PDGs almost always obtain higher pay–offs than the continual mutual Defection pay–off.

It has been argued (see the next chapter) that the key to effective play in repeated PDGs is to elicit Cooperation in your opponent, so that mutual Cooperation can be achieved. One of the most effective strategies for eliciting Cooperation has been Tit–for–Tat (TFT). This strategy makes a Cooperative move on the first stage of a repeated game, and then simply mimics the opponent's previous move on all subsequent stages of the game. It is claimed that TFT does so well in repeated PDGs because it never Defects before its opponent does, both Cooperation and Defection are immediately reciprocated, and it is clear to the opponent what strategy is being used.

TFT would appear to be good at eliciting Cooperation from the opponent and maintaining mutual Cooperation, but it does not take into account other effects such as the termination time of the game. In most experimental studies, the end point of the game is assumed to be unknown to the players and also to be independent of the choices of the players. Often the end point is determined by a probabilistic procedure, with the probability of the game terminating having a geometric distribution, or similar. If, on the other hand, the end point of the game is known, then various end–effects come into play. The effect of knowing the termination time of the game will depend upon the level of discounting of future pay–offs and the game conditions. As we shall discuss further in chapter 6 of this thesis, we shall make the assumption that the end–point of the game is unknown, to avoid these end–point effects.

2.3 An Experimental Study.

To demonstrate the types of experimental games that have been performed, we shall briefly discuss one experimental study. In common with most experimental studies, the subjects used as players were all drawn from a particular population — in this case a group of undergraduate students. These players were not allowed to communicate at all during the games, except to indicate their chosen move at each stage of the game, which the two players did simultaneously. Again, as is generally done, the players were informed that an amount of money equal to some increasing function of the sum of his pay–offs would be paid to each player when the experimental study was over.

This study comprised two experimental games — labelled here $G1$ and $G2$. As we mentioned above, in the majority of experimental games, the termination time of the game is unknown to the players. However, in the games presented here all players were told that each game would terminate after 20 stages. The subjects played the games facing their opponents, and so it was obvious that they were playing against a fellow undergraduate, as opposed to any unknown person, or indeed a computer. In between the games the players were changed around so that all players had a different opponent for the second game to the one that they had for the first.

The first game played ($G1$) was the game defined by the symmetric pay–off matrix given in Figure 2.3.1.

$$\begin{array}{cc} & P_2 \\ & \begin{array}{cc} 1 & 2 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \end{array}\begin{pmatrix} 2 & -1 \\ 3 & 0 \end{pmatrix} \end{array}$$

Figure 2.3.1 — $G1$

By checking Figure 2.2.3 and the given conditions, it is easy to verify that this game is a PDG. As stated above, the only equilibrium point is when both players choose their move 2 (which is also their maximin strategy). This would lead to an overall pay–off of 0 to both players. However, in the experimental study, only one of the 18 players received a pay–off lower than 0, with most players receiving a significantly higher pay–off. This would suggest that the players managed to communicate (through their moves) their willingness to cooperate with each other, rather than to continue with their dominating moves.

Another effect that has been previously discussed that is apparent in our example is a high incidence of *lock–in* effects (see Rapoport & Chammah, 1965). This occurs when the players seem to use a "training period" of only a few moves to learn about an opponent, and then play future moves accordingly with long runs of either mutual Cooperation or mutual Defection resulting. Many of the pairs that achieved a Cooperation lock–in, Defected on the last one or two stages of the game, presumably applying the 'extended rationality' argument outlined above. This would suggest that the players were prepared to Cooperate with each other to obtain a higher pay–off for themselves, but decided to Defect on the last couple of stages, either to prevent being exploited themselves, or in response to a Defection on the previous move by their opponent. Consider the results given in Figure 2.3.2 of two pairs of players playing this PDG.

STAGE		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
PAIR A	P_1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2
	P_2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2
PAIR B	P_1	1	2	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2
	P_2	2	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2

Figure 2.3.2

In both games the players had a training period of a few moves, far less for pair A than for pair B. After this, both pairs "locked–in" on mutual Cooperation, which was only broken towards the end of the game. Pair B both defected on the last two stages, therefore finishing equal on points, whereas P_1 of pair A gained the upperhand by defecting on move 18, before P_2.

The second game played ($G2$) was the game defined by the symmetric pay–off matrix given in Figure 2.3.3.

$$\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 1 & 2 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 4 & 3 \\ 6 & 0 \end{pmatrix} \end{array}$$

Figure 2.3.3 — $G2$

From Figure 2.2.2 we can see that this game is 'Chicken', and has two pure strategy equilibria, at $(1,2)$ and $(2,1)$. This game provided a good example of how tacit agreements can be made during a non–cooperative game. For instance, consider the results given in Figure 2.3.4 of a pair playing this game.

STAGE		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
PAIR C	P_1	2	1	2	2	2	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
	P_2	1	2	1	2	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1

Figure 2.3.4

In the above game, pair C settled into a run of alternating unilateral choices of move 1, after a training period of 5 moves. This alternating strategy produces the highest mutual pay–off of any feasible strategy. So in this game the players managed to "lock–in" to the alternating strategy, but there was no incentive to break out of this at the end of the game, as this would result in a loss in pay–off to the player who deviated.

It is interesting to note that the player who scored higher than all other players over both games employed a strategy (in both games) that was very similar to the TFT strategy mentioned above. This strategy's effectiveness to elicit Cooperation from its opponent in $G1$, and also its ability to maintain alternation in $G2$, enabled the player employing it to obtain higher pay–offs than the other players. Several players managed at some point to find the optimal mutual strategy combinations for the two games — mutual Cooperation in $G1$ and alternating unilateral choice of move 1 in $G2$. Obviously not all players did as well as the pairs given as examples above. These other players did not do so well for a variety of reasons — for example, playing only their dominating move in $G1$, or playing seemingly randomly in $G2$.

Now from the 'classical' game theoretic literature, it would seem unlikely that we would obtain the results that we did. It has been suggested that rational players should play their dominating move (move 2) in $G1$ and once one of the equilibrium pairs in $G2$ had been reached, there should be no reason to deviate from it. Yet this experimental study (and almost all others) show that players can easily do better than the prescribed "good play". As we have stated, the 'classical' game theoretic literature is mainly applicable to one–play, zero–sum games. The games in this study are neither one–play nor zero–sum, and so it is perhaps not too surprising that we have obtained the above results. So the theory that we shall concentrate on will be applicable to the more general repeated, non–zero–sum games, and therefore to one–play, zero–sum games as a special case.

In the next chapter we shall look more closely at the prescriptions of classical game theory, and the results of experimental games, in order to determine the discrepancies between them.

3. GAME THEORY LITERATURE

3.1 Introduction.

Now we shall consider how the theory of decision making in these games, i.e. game theory, directs the players to the move that they should make at each stage of a game. We shall discuss the various 'solution concepts' that have been proposed in the literature, and outline the goals that this theory has been directed towards. We shall also consider the results of the experimental game literature, i.e. from actual plays of the games under consideration. These games have been played by human against human, human against computer, and computer against computer, and a wide range of results has been obtained. These experimental findings can then be compared and contrasted with the recommendations of the game theory literature.

3.2 Game Theory Literature.

Game theory, as the mathematical approach to solving decision problems under conflict, began in the 1920s. Borel was perhaps the first to consider game theoretic problems, introducing the notions of mixed and pure strategies. Just after this, von Neumann proved the minimax theorem and created the theory of games with more than two players. These papers did not receive much attention at all, until the publication of the classic von Neumann & Morgenstern (1947) book. This lack of interest has been attributed to the high mathematical content of the early papers. The von Neumann & Morgenstern book was less mathematical, and therefore made game theory more available to other scientists and social scientists. It has also been speculated that this new theory also provoked more interest because of its possible application in the situations arising in the recent war.

Game theory was at this point considered to be a 'panacea' to solve all human conflict problems, and indeed researchers are still working on this 'classical' approach to game theory and determining new solution concepts. Luce & Raiffa (1957) was perhaps one of the earliest works to stress the limitations of game theory, and since then game theory has been considered to be a useful way of thinking about a given conflict problem. Research is now concentrated on devising techniques for analysing these problems, and also determining why people make the decisions that they are observed to.

The classic book of von Neumann & Morgenstern (1947) introduced not only the formulation for determining the solution of general n person zero–sum games, but also the theory of utility. This theory, although crucial to game theoretic solutions, is not explicitly part of game theory. A useful history of utility theory, and a full exhibition of the form of utility that we are assuming here, is given in Savage (1954). Von Neumann & Morgenstern argued that equilibria in static games needed to be determined before a useful dynamic theory could be developed. To initiate this they explicitly defined the complete concept of a game and then demonstrated the existence and method of determining a solution for all zero–sum games.

The two player zero–sum game is solved by the minimax theorem, and the n player zero–sum game ($n \geq 3$) is solved by considering a *characteristic function.*

THEOREM 3.1 — MINIMAX THEOREM. (von Neumann & Morgenstern, 1947)

In the domain of mixed strategies, every two player zero–sum game has at least one equilibrium pair, and where there are several, they are equivalent and the equilibrium strategies are interchangeable. The common utility of the equilibrium pairs is known as the value *of the game.*

An equilibrium pair in a game is a strategy pair where each player's equilibrium strategy determines an outcome that is the maximum entry in its column and the minimum entry in its row of the pay–off matrix. The characteristic function is a real valued set function that satisfies four simple conditions and determines the value of the game to various possible coalitions of the players. Von Neumann & Morgenstern also extend these arguments to more general games by considering the decomposition of games into simpler games, and by extending the concept of the characteristic functions to non–zero–sum games.

Luce & Raiffa (1957) parallels the work of von Neumann & Morgenstern, but from a less mathematical slant, and so concentrate more on the concepts rather than on the solutions of these games. Their work is presented from a social science point of view, and so they are mainly discussing the applicability of the mathematics as opposed to the mathematics itself. As well as being the first work to present the ideas of von Neumann & Morgenstern (1947) in an easy to follow style, Luce & Raiffa (1957) was amongst the first to consider a theory of repeated plays of a game (a *supergame* as they termed it), and also the importance of psychological factors. They show that in a supergame correlated joint strategies can emerge, even including dominated strategies. They also discuss what can be considered solutions to non–zero–sum games. Their approach relies mainly on the following theorem by Nash (1951).

THEOREM 3.2. (Nash)

Every non–cooperative game with finite sets of pure strategies has at least one mixed strategy equilibrium pair.

As a result of this theorem, equilibria in games are often referred to as 'Nash equilibria'. It is clear that solutions to non–zero–sum games can be found by this approach if all equilibrium pairs are interchangeable (i.e. have the same outcome for all of the players). Solutions can also be found if the equilibrium pairs are not interchangeable, by only allowing players to consider strategies that are not dominated by any other strategy, and then determining equilibria for this reduced game. Obviously in these non–zero–sum games, more factors have to be taken into account, as in these games an increase in pay–off to one player does not necessarily imply a decrease in pay–off to another player, as it does in zero–sum games. Concepts that were used in zero–sum games, such as maximin, can be extended to non–zero–sum games but the corresponding results cannot always be directly applied to non–zero–sum games.

Two player, two move games have been classified (see Rapoport & Guyer, 1966 and the symmetric classification in the previous chapter of this thesis), and whilst some specific named games have been considered extensively, e.g. Prisoner's Dilemma (PDG), Chicken, Battle of the Sexes, etc., others have not been considered in their own right at all. Of these named games, the game that has been considered by far the most is the PDG, and we too shall concentrate on this game. Often solutions have been determined for particular games, rather than the whole class of 2×2 games, although claims are frequently made that the results easily extend to the other games. One methodology that can extend across to the other games is that of Howard's metagames. This is an approach that was briefly dicussed by von Neumann & Morgenstern (1947) in terms of majorant and minorant games, and the theory was developed by Howard (1966a, 1966b, 1971) as discussed in chapter 4 of this thesis. A general discussion of two player, two move games is given in Colman (1982), and also from a more psychological aspect in Hamburger (1979).

When determining these equilibria, Shubik (1982) questions the use of mixed strategies in non–zero–sum games. Examples can be found where the players can obtain higher pay–offs by deviating from their equilibrium strategies, if the opponent recognises the possibility of this deviation. Therefore, the mixed strategy equilibria are not in general stable, and so he claims that equilibrium points in pure strategies are the only really significant non–

cooperative solution concepts. Shubik also considers the possibility of correlated strategies in non–cooperative non–zero–sum games. To be able to correlate strategies to mutual benefit, players must be able to 'communicate' with each other. As explicit communication is not permitted, it can only be achieved through laws or social norms, or through threats in a repeated game. We shall consider how a player can 'communicate' and 'learn' the strategies being employed, when developing our models.

The result of the minimax theorem can be extended to repeated zero–sum games, but the extension to repeated non–zero–sum games is not obvious. Equilibria can be found in these repeated games that obtain higher pay–offs than repeatedly playing the equilibria for the generating game, and these equilibria in repeated games exist because of the threat of lower pay–offs that each player has available. For example, in a repeated PDG, the mutual cooperation outcome achieves a higher pay–off than the mutual defection outcome, but it is not in equilibrium. It can be viewed as a sort of 'repeated equilibrium' as the players are aware that if they make a defection move, then their opponent is likely to respond with a defection, and mutual defection is likely to result. This type of rationale is extended in the next chapter.

Aumann (1981) gives a survey of repeated games that is divided into games with complete information and games with incomplete information. One important theorem for games with complete information is what has been referred to as the "Folk Theorem".

THEOREM 3.3 — FOLK THEOREM.

The pay–offs to Nash equilibrium points in a supergame G^* are the feasible individual rational pay–offs in the generating game G.

In this theorem, pay–offs are 'feasible' if they are a convex combination of the pay–offs to pure strategy n–tuples in G. This theorem therefore relates cooperative behaviour in a generating game to non–cooperative behaviour in the associated supergame. One other approach to multi–play equilibria is that of Selten (1975), which is termed *perfect equilibria,* and is based on the possibility of a player trembling and making a different move to the intended move. This is discussed in more detail in chapter 5 of this thesis, but its main characteristics are that perfect equilibria cannot include any weakly dominated strategies, and the pay–offs from perfect equilibria in the supergame are the same as repeated pay–offs from the Nash equilibria in the generating game.

In non–zero–sum games with incomplete information, Aumann (1981) states that a basic problem is that the actions of a player making use of his information can get mixed up with any signals to his opponent. Therefore it is difficult for a player to determine the intentions of his opponent from his opponent's moves, and so an infinite regress of the kind discussed in chapter 4 of this thesis is obtained. Taylor (1976) provides an interesting and amusing discussion of repeated plays of a PDG with incomplete information. Taylor's main argument is against the view that the only way to ensure mutual cooperation in such a game is to establish an authority that has enough power to make it in each player's interest to cooperate. He argues that mutual cooperation can be achieved without any outside authority, and it is because subjects are used to living under such an authority, that they try to get away with as much as possible when this authority does not exist. That all this can be determined from a simple PDG is doubtful, and there are problems with 'ecological validity'. We will try to develop a method that can take such factors into account, and this should be preferable to one that cannot.

An alternative angle of trying to achieve a resolution in conflict situations, i.e. trying to determine ways for players to achieve the mutual cooperation outcome is considered by Rapoport (1974). He discusses his own experiences with the PDG and reviews some of the main features of game theoretic reasoning. Rapoport claims that the effects of collective rationality (as opposed to individual) can only be attained by either changing non–cooperative games to cooperative ones, or by completely abandoning individual rationality. He also argues that the belief that the mutual defection outcome in a PDG is inevitable should lead people to try to find agreements to turn the game into a cooperative one, and thus achieve the mutual cooperation outcome. Rapoport's main concerns are that researchers should consider utility maximisation rather than simply pursuing equilibria, and that questions of rationality in non–zero–sum games should be considered as opposed to classical decision theory, and particularly *zero–sum mentality*. The work presented here is in accordance with these concerns.

Another author who believes that game theory can be used to reduce tension in real world problems is Osgood (1980). He puts forward a strategy named GRIT (Graduated and Reciprocated Initiatives in Tension–reduction), which is claimed to smooth a path toward negotiation. The strategy is specified by ten directives on the form of unilateral initiatives. These initiatives are designed to induce unilateral responses by the opponent, and thus inducing further rounds of reciprocations. Osgood claims that it was a strategy similar to GRIT that Kennedy and Khrushchev employed in 1963 in rounds of weapons reductions, and believes that such a strategy could be used to achieve the ultimate goal of getting rid of all nuclear missiles from the world. By using the players' beliefs, and knowledge of the problem, such strategies can be developed.

How these conflicts determine the evolution of species forms evolutionary game theory. The main author in this field has been Maynard–Smith, with most of the relevant work summarised in Maynard–Smith (1982). The theory of evolution depends on the evolution being defined upon the frequencies of the genes in the population, and that the frequency of a particular gene increases if it increases the *Darwinian fitness* of its possessors. Here Darwinian fitness is defined to be the expected number of surviving offspring from a particular genotype.

The concept of an *evolutionary stable strategy* (ESS) is central to Maynard–Smith's approach, which is a strategy such that if most of the members of a population adopt it, no mutant can 'invade' the population by natural selection. Maynard–Smith's achievement was to specify this mathematically, and then determine the existence and characteristics of ESS's in different populations. These models have stood up well to empirical evidence on a variety of animals and insects (such as dung flies and digger wasps). Also of relevance is the work of Axelrod (1980b), which proves the evolutionary stability of TFT in the tournaments that Axelrod organised.

It is clear that some games do not have an ESS and some have more than one. Work has continued by other authors to find simple ways of determining whether or not a strategy is an ESS. Bishop & Cannings (1978) showed that if a mixed strategy is an ESS, then all pure strategy components of this mixed strategy obtain the same expected pay–offs as the ESS. This proves very useful in the search for mixed strategy ESS. Haigh (1975) used the fact that pay–offs can be expressed in matrix form to produce a simple algorithm for determining whether a strategy is an ESS by simply checking the eigenvalues of a particular matrix. A small mistake in this work is corrected by Abakuks (1980), and Bishop & Cannings (1976) produce an essentially identical formulation.

ESS's in the first instance were set up in one–play games. However, there is obviously an underlying dynamic game, as the population will develop over time. This leads to the concept of dynamic equilibria in dynamic evolutionary games, and the stability of these equilibria (stable to perturbations of the strategies). It can be shown that in a continuous time evolutionary game, every ESS is asymptotically stable (all close strategies tend to the equilibrium strategy), but the converse is not true. Zeeman (1979) develops a dynamic system on the population in order to determine stability and classify games up to topological equivalence. Zeeman also shows that elementary catastrophes can occur in these systems. Zeeman (1981) considers animal conflicts in such dynamic systems, and the effect of stability upon these. These evolutionary stability concepts are related to the calibrated societies developed in chapter 6 below.

The concepts above form the classical approach to game theory, in terms of finding equilibria for mainly static games, although some work has been performed on dynamic, repeated games. Obviously there is a very wide and cross disciplinary literature on this subject and I have only considered a few, hopefully pertinent, references. A further and rapidly growing literature on a Bayesian approach to game theory, based on utility maximisation, is presented in chapter 5 of this thesis.

3.3 Experimental Game Literature.

After having discussed the literature on how people 'ought' to play the types of games that we are considering, we now turn to how people have been observed to play these games. Researchers have performed experiments to determine how people play several different sorts of games, and these have been widely published. This has produced a vast and widely spread literature with, as Rapoport (1974) points out, well over 200 experiments performed just on the PDG. These experiments have been performed for various reasons: to determine how different populations (e.g. sex) differ in play, responses to pre–programmed strategies, ability to predict the play of the opponent, etc. Because of these different reasons, experiments have been conducted in differing ways: in classrooms, in laboratories, or on computers, but the underlying priciples are similar. We shall now consider some of these experimental games, and point out their main conclusions.

Flood (1954a, 1954b) was one of the earliest authors to consider experimental games. His work has mainly considered one player games against 'nature' where a player makes a decision at every stage of a repeated game, and is then informed whether this decision 'won' or not. Flood (1954a) shows that players have the ability to 'learn' during such a game, and can therefore update and improve their strategies. This is comparable with the ideas of Wilson (1986) that are extended in chapter 7 of this thesis. Flood (1954b) conducted two experiments to attempt to show that the behaviour of the subjects would be a pure strategy if the subject was convinced that the game was stationary, and a mixed strategy if the subject believed that there might be non–stationarity. The experimentation did not conclusively show this to be true or false, but offered more evidence in favour rather than in contradiction. Flood also conducted several other experimental games (see Flood, 1958), and it is claimed that Flood was the first person to experiment with the PDG.

Rapoport & Chammah (1965) gives an extensive study of repeated PDGs. They mainly consider the changes in players' strategies during the game. They too are concerned with how players learn as the game progresses, and consider various differing populations, the effects of differing games, and also of concealing the pay–off matrix. They discovered a tendency for the rates of cooperation to decline initially, but then recover as the game continues. An interesting result was that over 90% of the pairs playing these games had matched strategies (i.e. the same) by the end of the game. Unilateral responses were ruled out by the 'conversion' of the defector, or the exploited deciding to give up hopes of converting the other. It was discovered that there were significant differences between the sexes, but these differences almost disappear when a male plays against a female. Rapoport and Chammah also claim that men are more prone to the TFT strategy than women. Their results are however based on aggregate behaviour, rather than trying to explain the reasoning of individuals. This thesis concentrates more on how an individual attempts to maximise his expected utility.

Experimental games just involving computers have also been performed. For example, Axelrod (1980a, 1980b, 1984) organised two tournaments in order to determine how to play a PDG effectively. People were invited to send in computer programs and these came in from researchers in many different disciplines. TFT won both of these tournaments, despite the fact that it can at best only do as well as its opponent. Axelrod claimed that these experiments show that a successful rule must be nice (never defect before the opponent), provocable (always defect after an 'uncalled for' defection) and forgiving (forgive defections on past stages of the game). Axelrod (1980b) also considered the stability of the strategies in an evolutionary sense, and Axelrod (1984) discusses collective stability (in terms of whether a group of strategies can be 'invaded' by another strategy). Rules for how to choose an effective strategy in a PDG in such a setting, and how to promote cooperation in this PDG are also given.

The difference between the data generated by these competitions analysed by Axelrod and most other experiments is that the programs were designed to achieve different aims to the aims that people typically appear to be playing for. The structure of the competitions would suggest that the programs were designed to play in a specific artificial environment, rather than in a general PDG. Therefore, the conclusions from the results of these competitions can not really be applicable to PDGs in general.

The aims of the programs were primarily to specify a strategy that would score more points than other strategies. The whole aspect of the two players collaborating to induce the mutual cooperation outcome is lost, and the interaction between the two players does not lead to an understanding of the other player, but merely how often the computer program chose to defect. Therefore, the players are not maximising utility on aggregate or even individual winnings, but on coming *first* in the tournament, and so a different form of rationality is in operation. The programs were tailored for a computer tournament setting and, especially in the second tournament, contained features that would cope with events known to occur in the competition. For example, a check to see whether the opponent was playing itself, or whether the opponent was in fact playing a random strategy — i.e. cooperating on each move with probability 0.5.

Because of these factors the results obtained in the tournaments do not appear to be typical of the data obtained in other experimental games, where subjects playing the games were human. From the results stated it seems that long runs of cooperation and defection were apparent in these tournaments. The games between strategies that are both 'nice' always produced games of nothing but mutual cooperation. However, long ruts of continual defection seem to be experienced by several programs when the opposing program was unresponsive to cooperation, or appeared to be.

In contrast, whilst pairs of human subjects can maintain reasonably long runs of joint cooperation occasionally and examples of joint defection are not impossible to find, much more typical are breaks in the sequences. If joint cooperation has been obtained, it is tempting for human players to try an occasional defection to see what happens, or what could be got away with. Some programs were set up to do this, although they would not have led to a new interaction learning position as humans typically do, but in most cases they led to mutual defection. If two human players were in a state of mutual defection there would typically be an occasional attempt to encourage the opponent to cooperate, which does not appear to be a general feature of the computer programs.

As mentioned above, the three attributes of successful programs in the tournaments were niceness, forgiveness and provocability. Forgiveness and provocability are often features of how the human players tend to play such games, but less so is niceness. However, that these three factors are essential properties for general successful strategies is dubious and that there are only three is also doubtful. Axelrod's work provides an interesting discussion on what makes a computer program that is designed to play another computer at a PDG successful. That it says anything about how people ought to play a PDG is doubtful and therefore it is unlikely that it is a "primer" on how to play the PDG effectively, except in the given setting.

By extending Axelrod's comments a more general strategy could be determined that will do well in such a game setting. I believe that such a strategy would contain the three attributes niceness, forgiveness and provocability, and also a fourth attribute: reciprocity, i.e. the ability to reciprocate the opponent's forgiveness (if any). These are all variable amounts, as they can be dependent on various lengths of past history, and effective for various numbers of future moves. Therefore a player can determine his prior beliefs about the optimal amounts of these attributes to maximise his utility. These can then be updated as more information is received. This model is considered further in chapter 9.

Behr (1981) shows that the results of Axelrod's experiments are changed by altering what a player's utility depends upon from best aggregate score to beating the opponent by the most, on average. Also results are compared when the random strategy is taken out of the analysis. TFT is no longer the best strategy, and in fact comes close to the bottom. Two non–nice strategies appear to perform best, although the nice/non–nice effect is not significant. Therefore correct specification of the utility function is crucial to the analysis of PDGs and games in general, and we agree with Behr that the objectives of the players must be of primary concern. So, realistic models must depend upon the game setting, and utility structure.

A comprehensive review of the literature on repeated PDGs and related games played by humans, is given in Colman (1982). This includes discussions on the proportions of players using minimax strategies or achieving lock–in of defection or cooperation, the effects of varying the relative magnitudes of pay–offs, the effects of circumstances of play, the effects of gender, and the player's beliefs about his opponent. Colman claims that these results show that, amongst other things, many players will reciprocate continual cooperation in a PDG, but a large proportion will exploit it, and that TFT does not necessarily elicit more frequent cooperative choices from subjects than other programmed strategies. Also Colman claims that many players have been observed to lock–in on mutual defection.

Colman also presents a set of experimental games of PDGs and of Chicken. The main findings here were that players made fewer cooperative choices in more life–like situations than in abstract variations of the games. Also in these life–like games, the strategies appeared to have utilities closer to the explicit pay–off structure given. One clear result that is echoed in several experimental games is that the effects of altering monetary pay–offs influences behaviour, but these effects vary considerably from player to player.

A further extensive review of experimental games is given by Pruitt & Kimmel (1977), where the games are split into four classes: matrix games, negotiation games, coalition games and trucking games. It is the matrix games that we are most concerned with here. Pruitt & Kimmel believe that more attention should be placed upon creative hypothesis building (in terms of how people devise strategies) and less upon hypothesis testing. They adopt a goal–expectations approach, whereby the outcomes are determined by the goals of each player and his expectations of the future actions of his opponent. This is loosely the basis on which our approach is based. It is claimed that simultaneous cooperation is the key to continual mutual cooperation, and this can be produced by experience of the mutual defection outcome. They also claim that attitudes, feelings and norms have little influence on behaviour in these games. I believe that such forces play a crucial role in determining optimal play, and they are central to the approach taken below.

The experimental evidence that cooperative and competitive players hold differing views over the actions of others when playing a PDG is examined by Kelley & Stahelski (1970). This is to say that the different views are caused by the players' personalities, and these have affected the players' experience of the interactions in the game. Experiments were then performed by playing people with cooperative or competitive 'goals' against other players with the same or different 'goals'. The results showed a behavioural assimilation of cooperators to competitors, and a competitor's misconception of the goals of the cooperators. The players expectations about their opponents' goals are summarised by the *triangle hypothesis* which is demonstrated in Figure 3.3.1.

This shows that a cooperative player believes his opponent to possibly have goals ranging from cooperative to competitive, whereas a competitive player believes that his opponent is similarly competitive. Therefore it would appear that nobody believes their opponent to be

$$\begin{array}{lrc} & & \text{Expectation of opponent's goals} \\ & & \text{Cooperative} \qquad \text{Competitive} \\ \text{Player's goals} & \begin{array}{r} \text{Cooperative} \\ \\ \\ \\ \text{Competitive} \end{array} & \begin{pmatrix} * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & & & * & * \\ & & & & * \end{pmatrix} \end{array}$$

Figure 3.3.1

more cooperative than themselves. Evidence of this kind can also be incorporated into a player's prior beliefs of how an opponent will play a PDG. This backs up our view that a successful model is dependent on the players' beliefs and the context of the game.

Other authors have considered games amongst people with differing goals. Terhune's (1974) approach was to seperate the players into four motive classifications: achievement, affiliation, power, or none of these. Terhune found that initial defensiveness needs to dissipate before personality effects can emerge, and so these personality effects are said to "wash-in" as the game progresses. Also mutual cooperation appears to be more likely to be experienced in the early stages of the game if the first move is cooperative, but these first stage effects "wash-out" as the game progresses. Terhune states that research should be performed to determine the interaction behaviour of the players of the game. I believe that this is attained quite simply by adopting a Bayesian approach, so that subjective beliefs are updated as more information is received.

Harford & Solomon (1967) also perform experiments to determine the effect of the initial moves of players upon the rest of a PDG. Two strategies are considered: "reformed sinner" which made 3 defections, then 3 cooperations and then played TFT, and "lapsed saint" which made 3 cooperations and then played TFT. Subjects played against these programmed strategies to determine the amount of cooperation each elicited. The reformed sinner was found to elicit more cooperation than the lapsed saint. It was claimed that this can be explained by the lapsed saint encouraging exploitation and not providing experience of mutual loss, whereas the reformed sinner is showing a willingness to cooperate as well as to fight. It is interesting to note that at the end of their experiment, Harford & Solomon asked the subjects to play one more stage of the game (known to be the last), when the (programmed) opponent had already chosen a cooperative move. This was to test trustworthiness, although it would appear that it has more to do with the player's utility function and understanding of the game rather than trust. However, over 70% of the subjects chose to cooperate on this final move, and there was no noticeable difference between the subjects that had played against the reformed sinner or the lapsed saint.

Similar experiments to those presented in Harford & Solomon (1967) were conducted by Wilson (1971), with a group of undergraduate students playing aginst pre–programmed computers. This was to determine the best strategy for inducing cooperation, to determine any intergroup bias and to study the effect of initial cooperation on later cooperativeness. The programmed strategies were TFT and three TFT deviates: firstly each defect provoked two defections in return, secondly a run of three mutual defections forced a cooperation, or thirdly a run of two mutual defections forced a cooperation. Players were led to believe that they were playing a member of an opposing team, and were regularly informed that their own play was more competitive than the average of the other players (irrespective of actual play). A naturally played game was performed as a control.

Wilson found that a period of double crossing increased the amount of cooperation in later stages, and that the programmed strategies achieved more cooperation from their opponents than the natural play. This could be an effect referred to by Axelrod (1984) as 'transparency'. A transparent strategy is where the opponent can determine the precise strategy after only a small number of stages. This can work negatively if the strategy is exploitable, or positively if the strategy encourages mutual cooperation, like for instance TFT. Obviously in this case the pre–programmed strategies are likely to be more transparent than the natural play, as the programs are playing fixed strategies. Also TFT was found to elicit more cooperation than the other strategies (in line with Axelrod, 1980a, 1980b), and there was a tendency to rate the 'in–group' higher than the 'out–group', which is in accordance with other studies (e.g. Kelley & Stahelski, 1970), suggesting that an opponent is believed to be at least as competitive as the player concerned.

Laskey (1985) conducted a similar experiment to determine the ability of subjects to predict the cooperativeness of his opponent at the next stage of a PDG. Subjects were also asked to report their overall strategy. Many of the subjects had a belief that their opponent was likely to alternate between cooperation and defection, which a model designed to predict the players' cooperativeness failed to recognise. This was probably a function of the pay–off matrix used, as alternation achieved a pay–off close to the mutual cooperation pay–off. The actual games exhibited features commented on above, such as an amount of 'lock–in' and retaliating to defections. The subjects managed a higher pay–off than the predictive model of the players achieved, and were better at predicting the cooperativeness of the opponents than the model. This backs up the notion that a player uses his subjective beliefs to determine his own strategy.

The concept of *utility* represents the relative amount of satisfaction that a player attains from given outcomes of a game. It is central to our approach that some people can have differing utility functions to other people, and this can be seen to be a factor in the above experiments. This could explain why in Colman's (1982) experiments, different results are obtained when the phrasing of the instructions of the experiments is different, yet the pay–off matrix is the same. Rapoport & Chammah (1965) shows that by altering the magnitude of the pay–offs but keeping the ordering of the pay–offs the same, the subjects are observed to play differently. Laskey (1985) asked subjects what their aims over the whole game were, and these were highly variable, ranging from trying to achieve mutual cooperation, to trying to exploit the opponent as much as possible.

So it is clear that the possibility of non–equal utility functions must be allowed for, and also the possibility that these utility functions are not linear in the pay–off received. Also the actual pay–offs that players receive at the end of such an experiment is normally very small, due to the limit on resources of most research establishments. With only small losses occurring if a 'bad' outcome occurs, players might well be tempted to try obviously suboptimal strategies for no better reason than to "see what happens". Therefore, as the monetary pay–offs are so small, more utility may be obtained by different means than the monetary gain. Colman (1982) comments that the monetary incentives influence behaviour, but not in a manner that is consistent across the players, or across different games. It could be argued that pay–offs in the pay–off matrix are supposed to specify utilities completely, but in this case why do players appear to have differing utilities over the set of outcomes? Also, this would make all the results from experimental games wrong, as all the outcomes determine in a typical experimental game is the pay–off, not the complete utility.

The difference in the utilities obtained from the different pay–offs will have an effect upon the players' choice of moves. If the utility obtained from the mutual defection outcome in a PDG is very similar to the utility obtained from the mutual cooperation outcome, then the defection move would appear to be a more likely result than if there had been a large discrepency between the utilities. Very similar utilities could be a result of the pay–offs being very small in comparison to the current wealth of the players. In order to rationalise another player's moves, an understanding of that player's utility function is therefore required. So this knowledge is required to determine the likelihood of mutual cooperation or defection after any sequence of moves. However such information about an opponent is not usually available to a player. So the player must determine his subjective beliefs about this utility function by considering earlier games, previous interactions, his beliefs about the population from which his opponent comes from, etc. A probabilistic structure achieves this most efficiently.

Also any discount factor that players might use to discount future pay–offs (perhaps due to inflation) relative to current pay–offs is important, and must be included in a game model. For example, if the discount factor is such that the utility from a unilateral defection move on the first move of a PDG plus the discounted utility from mutual defection on all subsequent moves is greater than the discounted utility from mutual cooperation throughout the game, then it would make sense to defect on the first stage of the game, irrespective of all other factors.

One way of classifying the utilities of the players of experimental games is in terms of their goals or intentions. This type of classification is considered by, for example, Kelley & Stahelski (1970) and Terhune (1974). In this case, when playing a repeated PDG, the players involved will have a goal that they aim to achieve in the long term. They will hope to do this by attempting to manipulate the responses of the opponent so that the goal is achieved. To be able to manipulate these responses in such a way, a player needs to be able to assess the likely ***degree of alienation*** (DOA) of his opponent. From this, an optimal class of strategies to play against this opponent can be obtained in order to achieve the desired goal.

It can be seen from this argument that whatever goals are desired, information about the opponent's behaviour pattern is of value. Also any information conveyed to the opponent about the player's own behaviour pattern will also be of value to the player when trying to determine his optimal strategy. For the repeated PDG, one classification of goals is into five categories of degrees of alienation for each player (P_1) as is summarised in Figure 3.3.2.

DOA	Behaviour	Description
0	Cooperative	P_1 plays completely cooperatively.
1	Egocentric	P_1 maximises his own pay–off, given that his opponent (P_2) independently maximises his pay–off.
2	Paranoic	P_1 maximises his own pay–off, given that P_2 independently maximises P_1's losses.
3	Competitive	P_1 maximises the difference between P_1's pay–off and P_2's pay–off.
4	Punitive	P_1 maximises his opponent's losses.

Figure 3.3.2

Obviously the higher that P_1's degree of alienation is, the less cooperative P_1 is. It is possible for P_1 to determine which category he belongs to himself, and also his beliefs about the category that his opponent belongs to. Given these beliefs P_1 can determine his optimal strategy. As the game progresses, more information will be obtained about the classification of P_2 (although this is dependent on the actions of P_1), and P_1's beliefs can be updated.

How P_1 decides to play in the initial stages of a game such as a repeated PDG can be crucial to the achievement of both short and long term goals, even if there is no discounting. The effect of this may well be exaggerated when playing a pre–programmed strategy as in several experimental studies. The literature does not seem to suggest which move is optimal on the first stage of a repeated PDG even in the most idealised and simple setting. This is to be expected, as the optimal move must depend not only on the context, but also on the beliefs of the players and also their utility functions, and can be a very complicated problem, as we shall see in chapter 4 of this thesis.

One criticism that is often levied at experimental games is that they lack ecological validity — they are purely experiments performed in a laboratory and give no indication as to what would happen in an identical game in real life. This is often a fair criticism, and it should be recognised that great care must be taken when applying experimental results. Very few direct comparisons of games in abstract and lifelike forms have been performed, one such comparison being in Colman (1982). Colman's results suggest that the players were more competitive in the lifelike game than in an identical abstract game. Care must again be taken here, as obviously different lifelike games will have different effects on the players.

Great care must also be taken when experimental games have been performed using pre-programmed strategies. This is particularly so in the case of the tournaments conducted by Axelrod (1980a, 1980b). Clearly such things as attitudes, feelings and norms can take on vastly different connotations in real life games to what they would in an abstract game. Also one problem in real life games is that players will not necessarily be equal in power, and thus the symmetry is lost. It has been speculated (Pruitt & Kimmel, 1977) that relative weakness is likely to produce a tendency to reciprocate the behaviour of the other player. The problem of ecological validity can, to a certain extent, be narrowed down to the problem of correctly specifying the utility functions of the players — the factors that players gain utility from in a laboratory setting are possibly quite different from those that players gain utility from in a real life game.

Therefore, a variety of results has emerged from these experimental studies. This should not be too surprising given the differences in the experiments, and the differing aims of the subjects. Various underlying results do seem to hold with a degree of generality, such as the lack of consistency of the effect of monetary pay–offs, the dominance of 'TFT' as a good strategy in PDGs and differences in players' goals or utilities in the games. In determining a Bayesian model for how a player believes his opponent will play, and therefore how he should play, this information about the opponent's beliefs and utilities can be incorporated. We shall consider Bayesian models for these experimental games in chapter 6 of this thesis.

3.4 Conclusions.

From the above we can see that the game theoretic literature determines a mainly normative theory of how people should play particular games. However, the experimental game literature presents a theory of how people actually play these games. These two theories often give widely differing results, the difference being more marked in some games (e.g. the PDG), than in others. The basic approach of game theory is to determine equilibria for the players in the game, so that they achieve an outcome that no one player has any incentive to move away from. Experimental evidence, on the other hand is more concerned with determining the various goals of the players, and the players' expectations of how their opponents will play.

Experiments into how people play these games have usually concentrated on the effects of differing personality characteristics of the players, the amount of information they have received, how they react to different strategies by the opponent, and their responsiveness to changes in pay–off. Often these effects produce significant results, whereas from a game theoretic viewpoint, no change should occur as the equilibria are unaltered. Results of this kind lead to the desire to model these games in such a way that these personality and psychological effects can be incorporated. Also, in order to determine a model that is capable of making reasonable predictions of future play, we must base any such predictions on a similar basis to that which the players appear to be using. I feel that features like these are best incorporated probabilistically, using the subjective beliefs of the players. Therefore, it would seem that a method using utility maximisation would be more appropriate than one using equilibria such as those discussed above. It is a utility maximising model based upon the players' subjective beliefs that we hope to develop in the following chapters.

4. THE INFINITE REGRESS

4.1 Introduction.

In trying to determine a Bayesian model for how a player should play a particular game a problem arises that has traditionally been called *the infinite regress.* Essentially this regress arises because of the way in which a player determines his next move. In order to determine his optimal next move he must determine his beliefs about how his opponents will play on the next stage of the game. In an analogous way, the opponents' next moves will depend on what they think about the player and how they think he will play on the next stage of the game.

Therefore to make a rational decision about which move to make on the next stage of the game, a player must determine his views about his opponents' future play, how they think he will play, how they think he thinks they will play, etc. This can readily be seen to extend to an infinite number of views, and is what is referred to as the infinite regress.

A number of authors have tackled this problem. Some (e.g. Mertens & Zamir, 1985) approximate the spaces of beliefs that the players could hold by a finite set and then introduce a probability distribution on these beliefs in order to determine equilibrium points of the game. Others (e.g. Harsanyi, 1967) summarise the uncertainty in terms of vectors which can then be incorporated in the games as chance moves. This game can then be transformed into a game with known solutions. A different approach is taken by Howard, and this approach is discussed in section 4.3 below.

Our approach is related to all of these, and can be seen to be a generalisation of Howard's approach. We define the process that determines the regress in terms of how the players will play the game, and how they think their opponents think about the game, given their prior beliefs about their opponents. We then show that, defined in this way, the regress can be curtailed to a finite number of levels, by imposing constraints on the rationality of the players and by making assumptions about the players' beliefs about their opponents' utility functions.

4.2 Example: Nuclear Disarmament.

To demonstrate the notion of the infinite regress we now consider a highly simplified example. Suppose there are two potentially hostile countries, A and B, who have equal stockpiles of nuclear weapons. At any stage of a repeated game the presidents of these countries can either decrease their stockpile (move D), or they can leave them as they are (move N). If both countries decrease their stockpiles, then the world is considered a safer place; if only one country decreases its stockpile then this country is weaker than the other and in danger of being attacked; and obviously if neither country decreases then there is no change from the current position. We define the utilities obtained by the two countries to be as those given in the pay–off matrix in Figure 4.2.1.

$$\begin{array}{ccc} & & \text{Country } B \\ & & \begin{array}{cc} D & \quad N \end{array} \\ \text{Country } A & \begin{array}{c} D \\ N \end{array} & \begin{pmatrix} (10,10) & (-30,5) \\ (5,-30) & (0,0) \end{pmatrix} \end{array}$$

Figure 4.2.1

We shall assume for simplicity that D and N are the only two options that are available to the two countries. How should the president of country A play this 'game'? Should he order his military to dispose of a given number of weapons and possibly expose his country to an attack from a now relatively stronger country B? Or should he order no change in the stockpile and possibly lose the chance of a substantial increase in the stability of world peace?

To determine which move to make, country A must consider how country B is going to play. If it is believed that country B will reduce its stockpile at the next stage, then it would be better for country A to disarm as well. If, on the other hand, it is believed that country B will maintain current stockpiles, and would exploit any weapon advantage over country A if country A disarmed, then it would be better for country A to maintain its current level of weapons too.

Country B must be in the same situation. So in deciding which move to make, country A must think about how country B will think about the game, and therefore how country B thinks country A will play. Again, country B can be seen to be have this identical problem, and so how country B thinks country A thinks country B will move at the next stage must also be taken into account. Obviously this argument can be continued to an infinite number of levels of thoughts.

So, in order to determine how to play the 'game' given above, a president must consider an infinite number of levels about what the countries think about each other, and about how they think each other will play at the next stage of the game. This is called the infinite regress.

4.3 The Theory Of Metagames.

Howard (1966a, 1966b, 1970, 1971) proved some interesting results for a regress of this kind. His interest in this field was that, when playing PDGs like that above, players were observed to find an equilibrium at the mutual cooperation outcome. He explained this in terms of larger games, called *metagames*. These metagames allow players' strategies to be functions of their opponents' strategies, and therefore also allow strategies of strategies of strategies, thus inducing an infinite regress. What Howard showed was that, provided only a finite number of players were participating, only a given (finite) number of levels of this regress need be considered to find all equilibrium points of this regress.

In this analysis Howard only considered pure strategies. This is based on the arguments that players will not make serious decisions 'on the flip of a coin', and that a player can always find a pure strategy that will do as well as a given mixed strategy. Therefore the players are only allowed to use pure strategies, so they cannot, or will not, use mixed strategies. Because of this, and since the theory is directed to finding only equilibrium moves, it is only necessary to consider ordinal preferences. That is, we are only concerned with how the players' preferences for the outcomes are ordered, not by how much one outcome is preferred to another. Again we shall concentrate on games with only two players, P_i and P_j.

Definition. A *rational outcome* for P_i is a strategy pair $(\bar{a}_1, \bar{a}_2)$ such that

$$U_i(\bar{a}_1, \bar{a}_2) \geq U_i(a_1, \bar{a}_2) \qquad (4.3.1)$$

for all strategies a_1 available to P_i, where U_i is P_i's utility function.

Definition. An *equilibrium* in a game G is an outcome that is rational for both players.

This takes us on to consider what Howard calls the rationality axiom and the existentialist axiom. The *rationality axiom* states that a player will always choose his most preferred outcome, provided he believes that he will actually achieve this outcome. The *existentialist axiom* states that if a player is aware that a theory predicts how he should play, then this player can decide whether to obey or disobey this theory. Therefore a theory about the first theory is required for when the player chooses to disobey the theory.

Now from the rationality axiom P_j will choose his most preferred outcome *given* his beliefs about how P_i will play. So this can be seen as a function from the set of P_i's moves to the set of P_j's moves. But then the existentialist axiom says that P_j can choose any function from

P_i's moves to P_j's moves. These functions define a new game with an outcome (a_i, g_j), where a_i is a move for P_i and g_j is a function from P_i's moves to P_j's moves. Obviously from this we can determine the corresponding outcome of the original game, which is $(a_i, g_j(a_i))$.

At this stage it will probably be instructive to consider an example. Suppose P_i and P_j are playing the PDG given by the pay-off matrix in *Figure 4.3.1*, which will be labelled game G.

$$\begin{array}{cccc} & & \multicolumn{2}{c}{P_j} \\ & & C & D \\ P_i & C & (5,5) & (0,8) \\ & D & (8,0) & (1,1) \end{array}$$

Figure 4.3.1

Now in this basic game the only equilibrium is at the mutual Defection outcome, (D, D). As we have just discussed, by applying the two axioms for P_j we obtain the game given in Figure 4.3.2.

$$\begin{array}{cccccc} & & \multicolumn{4}{c}{P_j} \\ & & CC & CD & DC & DD \\ P_i & C & (5,5) & (5,5) & (0,8) & (0,8) \\ & D & (8,0) & (1,1) & (8,0) & (1,1) \end{array}$$

Figure 4.3.2

The move YZ for P_j refers to the function that determines move Y for P_j if P_i plays move C, and move Z for P_j if P_i plays move D. These four functions (CC, CD, DC, DD) from P_i's moves are the g_j referred to above, and the outcome from these functions is easy to determine. Again the only outcome that is in equilibrium is the mutual Defection outcome, $(D, DD) = (D, D)$.

This extension from the original game is viewed as P_j considering how P_i will play. P_j is believed to be thinking along the lines of "If P_i were to play move X, then I would play move Y," etc. This can be seen to be identical to the situation where P_j is being threatened by P_i, and is determining a response to this threat.

Then we reapply the rationality axiom to this new, extended game to construct functions from the functions $\{g_j\}$ to the set $\{a_i\}$, of which there are 16 in the 2 player, 2 move game. Reapplying the existentialist axiom we obtain a second level metagame with an outcome

$(f_i(g_j), g_j)$. Again, from this outcome we can determine the corresponding outcome from the original game. In terms of the above example, this generates the metagame with pay–off matrix given in Figure 4.3.3. Here the move $WXYZ$ for P_i means play move W against CC, X against CD, Y against DC and Z against DD.

$$\begin{array}{cccccc}
 & & & P_j & & \\
 & & CC & CD & DC & DD \\
 & CCCC & (5,5) & (5,5) & (0,8) & (0,8) \\
 & CCCD & (5,5) & (5,5) & (0,8) & (1,1) \\
 & CCDC & (5,5) & (5,5) & (8,0) & (0,8) \\
 & CCDD & (5,5) & (5,5)^{*} & (8,0) & (1,1) \\
 & CDCC & (5,5) & (1,1) & (0,8) & (0,8) \\
 & CDCD & (5,5) & (1,1) & (0,8) & (1,1) \\
 & CDDC & (5,5) & (1,1) & (8,0) & (0,8) \\
P_i & CDDD & (5,5) & (1,1) & (8,0) & (1,1) \\
 & DCCC & (8,0) & (5,5) & (0,8) & (0,8) \\
 & DCCD & (8,0) & (5,5) & (0,8) & (1,1) \\
 & DCDC & (8,0) & (5,5) & (8,0) & (0,8) \\
 & DCDD & (8,0) & (5,5)^{*} & (8,0) & (1,1) \\
 & DDCC & (8,0) & (1,1) & (0,8) & (0,8) \\
 & DDCD & (8,0) & (1,1) & (0,8) & (1,1) \\
 & DDDC & (8,0) & (1,1) & (8,0) & (0,8) \\
 & DDDD & (8,0) & (1,1) & (8,0) & (1,1)^{*}
\end{array}$$

Figure 4.3.3

Again this can be phrased in terms of threats, as P_i could play strategy X if P_j threatens to play strategy Y, etc. The equilibria for this game are marked with a $*$. Note that the outcome with pay–off $(5,5)$, i.e. the mutual cooperation outcome in the original game G, appears as an equilibrium in this game.

One point to bear in mind is that it is not possible to expand over both players' choices simultaneously, as this can lead to mutually inconsistent strategies. For instance P_i could decide to play the same move as P_j, when, at the same time P_j decides to play the opposite move to P_i.

The game given in Figure 4.3.2, which we derived by considering how P_j might play, given P_i's choice in the original game G, we obtain a first level metagame which is denoted jG. Similarly, taking P_i's choices over P_j's moves in jG we have the second level metagame given in Figure 4.3.3, which is labelled ijG. Higher level games are labelled obviously. Thus iG corresponds to the game that would be played if P_i knew exactly how P_j would play the game G, and ijG corresponds to the game that would be played if P_j knew how P_i would play the game G and P_i knew how P_j would play the game jG. Therefore, in the game ijG, P_i is trying to determine the equilibria for both players, given that P_j has determined the equilibria for how both will play the game G.

It is possible to consider games where the players are not taken alternately when determining the strategies, e.g. iiG, $jjiG$, etc. These games are included for completeness, but are only trivial extensions of the other 'alternate' games, as no further functions are being considered. In the game $k_1 \ldots k_r G$, the sequence $k_1 \ldots k_r$ is referred to as the *title of the game.*

Definition. If an outcome is rational for P_i in the metagame $k_1 \ldots k_r G$, then the corresponding outcome in the basic game G is called *metarational* for P_i.

As we consider larger and larger metagames (i.e. the title of the game is increased) then the number of metaequilibria cannot decrease, because any metaequilibrium from a given game must also be a metaequilibrium from all higher level games derived from the first game. So these metaequilibria can be viewed as additions to the equilibria of the original game G, and as higher level games are considered, the set of metaequilibria will not decrease.

Now we define three sets for P_i: A_i, B_i and C_i, which depend on the last occurence of i in the title of the game. P_j $(j \neq i)$ belongs to A_i if j appears in the title *after* the last occurence of i (or just appears if i doesn't appear). P_j belongs to B_i if j appears in the title only *before* the last occurence of i. P_j belongs to C_i if j doesn't appear in the title at all. We then have the following result.

THEOREM 4.3.1. (Howard)

An outcome $(\bar{a}_i, \bar{a}_j)$ is metarational for P_i in the game $k_1 \ldots k_r G$ if and only if

$$U_i(\bar{a}_i, \bar{a}_j) \geq \begin{cases} \max\limits_{a_i} \min\limits_{a_j} U_i(a_i, a_j) & \textit{if } P_j \in A_i, \\ \min\limits_{a_j} \max\limits_{a_i} U_i(a_i, a_j) & \textit{if } P_j \in B_i, \\ \max\limits_{a_i} U_i(a_i, \bar{a}_j) & \textit{if } P_j \in C_i. \end{cases} \tag{4.3.2}$$

The proof follows by induction on the number of terms, r, in the title of the game, i.e. $k_1 \dots k_r$. The interested reader is directed to either Howard (1966b, pg. 191), or Howard (1971, pg. 89).

COROLLARY 4.3.2. (Howard)

For any given metagame $k_1 \dots k_r G$ we can delete all except the last occurence of each player in the title of the game, and not affect the set of metarational outcomes.

Thus, in the two player game we need only consider up to the second level metagames ijG and jiG to determine all metarational outcomes, as in all higher level metagames we will only obtain the same metarational outcomes as we have already obtained. In general, in n player games we need only consider up to the nth level metagames to determine all metarational outcomes. From these metarational outcomes we can obviously determine all metaequilibria. In the two player case, the metaequilibria are simply the intersection of the metarational outcomes for the two players. For games involving more than two players, at all points that are metarational outcomes for all players we can construct a further metagame (from the original) which is at equilibrium at this point.

So we can construct all metaequilibria by considering a finite number of metagames. However we are left with the problem of determining which of these equilibria a player should adopt. Howard then proves various results about different types of metagames. A *complete* game $k_1 \dots k_r G$ is a game such that each player occurs in the title of the game once.

Definition. If an outcome is a metaequilibrium in all complete games for a given set of players, then it is a *symmetric equilibrium.*

It is not difficult to see that every complete game has a metaequilibrium, but the set of symmetric equilibria may be empty. If the game being played is symmetric (i.e. the pay–off matrix for P_j is the transpose of the pay–off matrix for P_j), then *all* metaequilibria are symmetric equilibria. This follows because the matrix for the metagame ijG is simply the transpose of the matrix for the metagame jiG, due to the symmetry of the original pay–off matrix. This result obviously holds for a symmetric game with any number of players.

It has been argued that these symmetric equilibria are the most natural to be considered, especially if there is no reason to suppose that P_i and P_j are different, but it is not clear which should be chosen if there is more than one symmetric equilibrium. Howard also gives some applications of this theory (see especially Howard, 1970, 1971).

So the infinite regress that is induced by the rationality and the existentialist axioms given above, can be reduced to a small number of levels in order to determine certain equilibria. Whether the basis on which these equilibria are calculated is acceptable, and where the theory leads to if we change this basis is considered in the next section.

4.4 Belief–Rational Strategies.

For notational simplicity we shall now concentrate on a two player game where both players (labelled P_i and P_j in this chapter[1]) have two moves available at each stage. Later in this chapter we shall discuss the generalisation to the case where each of $n \geq 2$ players have $m \geq 2$ moves avialable at each stage, and show that the following results extend to these cases.

Therefore we shall consider the general 2 × 2 pay–off matrix given in Figure 4.4.1.

$$\begin{array}{ccc} & & P_j \\ & & \begin{array}{cc} n_1 & n_2 \end{array} \\ P_i & \begin{array}{c} m_1 \\ m_2 \end{array} & \begin{pmatrix} A, A' & B, B' \\ C, C' & D, D' \end{pmatrix} \end{array}$$

Figure 4.4.1

In the first instance we shall only be considering the game from P_i's perspective. By this we mean that we shall determine the optimal strategy for a given player, P_i, to adopt, given his beliefs about how P_j will play. It has been argued (e.g. Terhune, 1974) that this is unnecessarily restrictive, and that to gain real insight into the problem we must consider the situation from all players' perspectives. We discuss the significance of this perspective later.

We label the utility functions that P_i believes the players to have as U_i and U_j for P_i and P_j respectively. P_i is assumed to know U_i, but not U_j. In line with the rest of this thesis, we shall assume that P_i believes P_j has be drawn, at random, from a known population. P_i is assumed to have a distribution over the players in this population, about their utility functions and about the strategies they adopt. Now every strategy or move available to the opponent has an associated expected utility, after we have averaged over the population that P_j has been drawn from. For the rest of this chapter we shall make the simplifying assumption that

[1] The players have been indexed i and j as opposed to 1 and 2 in this chapter to distinguish from the levels of the regress, which will have numeric subscripts.

P_i believes P_j's expected utility function is found by taking expectations over this population. Therefore, P_i will assume P_j's preferences are determined by $\overline{U}_j$, the expected utility function defined above. So, P_i can substitute for P_j a 'typical opponent' (P'_j) whose utility function is $\overline{U}_j$.

Now the equilibria for such games have traditionally been calculated by determining an outcome where no player has any incentive to make a different move. In many games this equilibrium concept leads to outcomes that most players believe they could improve upon by making a different move. However they are bound to choose an equilibrium move to avoid obtaining a smaller pay–off even if, in their opinion, the probability of obtaining this smaller pay–off is close to zero. The Bayesian approach to such games suggests that rather than trying to find an outcome that no–one will move away from, a player is simply trying to achieve the utility maximising outcome with respect to how he believes his opponent will play, and how he believes his opponent thinks he (the player) will play. Thus an outcome may well be stable in this sense and not in the former sense, if for example, both players attach a very small probability to their opponent exploiting them. For an illustration of this, see section 4.6.

So we are trying to determine a strategy that is utility maximising for P_i given his beliefs about P'_j. If P_i were to adopt strategy Q and P'_j were to adopt strategy R, then we denote the expected utilities to P_i and P'_j by $\overline{U}_i(Q,R)$ and $\overline{U}_j(Q,R)$ respectively. We shall assume that P_i has some initial beliefs about how P'_j will play this game. Let the strategy that P_i believes P'_j will initially adopt be labelled R_1. The assumptions that P_i makes in order to determine the strategy R_1 are discussed in the next section. Initially, the strategy is determined through P_i's subjective beliefs about how P'_j will play. These beliefs are then guided by considerations of P'_j's rationality. That is, whether P'_j is utility maximising with respect to the expected utility function $\overline{U}_j$, and whether P'_j believes P_i to be utility maximising, etc.

We let $\mathbf{Q}$ be the set of all possible strategies for P_i, and $\mathbf{R}$ be the set of all possible strategies for P'_j. In the two player, two move game under consideration these sets must be equal because if a strategy is available to P_i, it must also be available to P'_j. Firstly we need to define what we mean by a belief–rational strategy and a belief–rational player.

Definition. A *belief-rational strategy* Q^* for P_i is such that

$$\overline{U}_i(Q^*,R) = \max_{Q\in\mathbf{Q}}\{\overline{U}_i(Q,R)\} \qquad (4.4.1)$$

where $\overline{U}_i$ is the expected utility function for P_i, taking expectations over the set of strategies for P'_j, and R is the strategy that P_i expects his 'typical opponent', P'_j, to play.

Definition. A player is said to be *belief-rational* if he always uses a belief-rational strategy, given his beliefs about his opponent.

This is obviously very similar to the notion of a metarational outcome given in the previous section, extended from the one outcome case to the case where strategies for the rest of the game can be considered. This is then restricted somewhat, because a belief-rational strategy is a strategy that is utility maximising based upon how P_i believes P'_j will play. This differs from the method given above in section 4.3, where any move was acceptable, provided it determined an equilibrium outcome.

As we are considering this problem from P_i's point of view, we shall always assume P_i to be belief-rational. However, P'_j is not necessarily assumed to be belief-rational. Once we have determined what each level of the infinite regress represents, we can impose various degrees of rationality upon P_i's beliefs about P'_j. At first we can assume that P_i does not consider P'_j to be belief-rational, then P_i assumes P'_j to be belief-rational, then P_i assumes that P'_j believes P_i to be belief-rational, etc. The level to which P_i thinks about P'_j's belief-rationality obviously affects the strategy that is utility maximising.

It is interesting how closely this ties in with the work of Howard discussed in the previous section. The infinite regress occurs for precisely the same reason — how a player believes his opponent is thinking about the game. Also, as we shall see in the next section, the method that we propose increases the dimension of the problem at each level of the regress, as does Howard's model.

There are, however, crucial differences. Firstly there is a different interpretation of rationality, as defined above. Associated with this is the different way that utility is handled. Howard's method requires only ordinality, whereas our method depends strictly on the cardinality of the utility function. Also, we have a different interpretation of the initial level of the regress. Classically (and in line with Howard's method) one starts with the concept of Nash equilibria. From a Bayesian point of view this is not necessarily the most sensible strategy to employ as a starting point. As well as this, Howard's metagames only permit the use of pure strategies by the players. The theory developed below permits the players to employ any strategy — pure or mixed. One further distinction is that the metagame approach is applicable only to one–play games, whereas the theory that we shall develop can be applied to more general repeated games.

Our method uses the fact that P_i has beliefs as to how a typical player, P_j', will play. Therefore P_i will choose a strategy to maximise his utility with respect to these beliefs. This method does not have the stability of the Nash equilibrium solution at the initial level of the regress, as there is nothing to constrain the initial strategy to be in equilibrium. However, as discussed in the next section, this stability may be imposed at the higher levels of the regress by different means. It should also be noted that the method that we develop is a generalisation of Howard's method, as we shall show in the next section.

One other feature that the two methods have in common is that the regress can be curtailed in both, but whilst the truncation follows in Howard's method as a consequence of the model, extra assumptions are required in our method. We discuss this in the next section.

4.5 Truncation of the Regress.

Firstly we shall discuss what each level of this infinite regress represents, and then determine how P_i should calculate a belief–rational strategy for each level. We shall continue to concentrate on the game given in Figure 4.4.1, and this game shall be laballed G.

We now need to consider how P_i thinks about P_j''s rationality. We can see that there are various levels of P_j''s rationality that P_i could assume: P_j' is not necessarily belief–rational, P_j' is belief–rational, P_j' assumes that P_i is belief–rational, P_j' assumes that P_i thinks P_j' is belief–rational, etc. Initially, we shall consider the first level of rationality, so we shall assume that P_i does not make any assumptions about P_j''s belief–rationality.

At the first level of the regress, P_i assumes that P_j' will adopt strategy R_1. Thus P_i has a simple problem of maximising over the set of strategies available to him (P_i), given his beliefs about how P_j' will play. As stated above, a classical method would advocate the use of an equilibrium strategy for P_i at this stage. It is easy to find examples where an equilibrium strategy is not obviously the best strategy for P_i to adopt (any PDG, for example). Our method dictates a much more general approach, that simply ensures that P_i chooses a strategy that will maximise his utility, given his beliefs about P_j''s strategy. Obviously in some cases P_i's strategy will be an equilibrium strategy, but not in *all* cases.

We shall label the strategy that P_i decides to play as Π. This will have a subscript which will indicate the level of the regress that determines the strategy Π as the utility maximising strategy for P_i. So, at the first level of the regress, the belief-rational strategy Π_1 is defined to be

$$\Pi_1 = \left\{ Q^* \,\middle|\, \overline{U}_i(Q^*, R_1) = \max_{Q \in \mathbf{Q}} \{\overline{U}_i(Q, R_1)\} \right\}. \tag{4.5.1}$$

Now we consider the second level of the regress. At this level P_i takes into account how he believes P_j' thinks P_i will play. We let Q_2 be the strategy that P_i believes P_j' thinks P_i will play. As P_j' takes strategy Q_2 into account, P_j' is expected to play strategy $R_2(Q_2)$. So, at the second level of the regress, the belief-rational strategy Π_2 is defined to be

$$\Pi_2 = \left\{ Q^* \,\middle|\, \overline{U}_i(Q^*, R_2(Q_2)) = \max_{Q \in \mathbf{Q}} \{\overline{U}_i(Q, R_2(Q_2))\} \right\}. \tag{4.5.2}$$

We can then consider the third level of the regress. Here P_i believes that P_j' thinks about how P_i thinks P_j' will play. We let R_3 be the strategy that P_i thinks P_j' thinks P_i thinks P_j' will play, $Q_3(R_3)$ the strategy that P_i thinks P_j' thinks P_i will play as a result of R_3, and $R_3(Q_3)$ the strategy that P_i thinks P_j' will play as a result of $Q_3(R_3)$. Hence, the belief-rational strategy Π_3 for P_i is defined by

$$\Pi_3 = \left\{ Q^* \,\middle|\, \overline{U}_i(Q^*, R_3(Q_3)) = \max_{Q \in \mathbf{Q}} \{\overline{U}_i(Q, R_3(Q_3))\} \right\}. \tag{4.5.3}$$

The fourth level of this regress is defined similarly on the game where P_i believes that P_j' thinks about how P_i thinks how P_j' thinks P_i will play. This process continues to further levels of the regress in a similar manner, ad infinitum.

Now we consider the second level of rationality, i.e. what effect the assumption of P_j' being belief-rational has for each of the levels of the regress. The utility maximising strategy for the first level depends only on the strategy R_1 that P_i believes P_j' will adopt initially. As P_j' is assumed to be utility maximising, the strategy R_1 must be a utility maximising strategy. Therefore, given this strategy, Π_1 can be determined by equation (4.5.1) as before. At the second level of the regress, P_j' is thought to take into account how P_i will play the game. So we can see that the strategies are defined by

$$\begin{aligned} R_2(Q_2) &= \left\{ R^* \,\middle|\, \overline{U}_j(Q_2, R^*) = \max_{R \in \mathbf{R}} \{\overline{U}_j(Q_2, R)\} \right\} \\ \Rightarrow \Pi_2 &= \left\{ Q^* \,\middle|\, \overline{U}_i(Q^*, R_2(Q_2)) = \max_{Q \in \mathbf{Q}} \{\overline{U}_i(Q, R_2(Q_2))\} \right\} \end{aligned} \tag{4.5.4}$$

Therefore P_i is choosing a utility maximising strategy, given that he believes P_j' thinks P_i will play strategy Q_2, and P_j' will play his utility maximising strategy given this belief.

The third level of the regress incorporates the strategy that P_i believes P_j' thinks P_i thinks P_j' will play. In a similar fashion to the above equations (4.5.4) we define

$$\begin{aligned} R_3(Q_3) &= \left\{ R^* \,\middle|\, \overline{U}_j(Q_3(R_3), R^*) = \max_{R \in \mathbf{R}}\{\overline{U}_j(Q_3(R_3), R)\} \right\} \\ \Rightarrow \Pi_3 &= \left\{ Q^* \,\middle|\, \overline{U}_i(Q^*, R_3(Q_3)) = \max_{Q \in \mathbf{Q}}\{\overline{U}_i(Q, R_3(Q_3))\} \right\} \end{aligned} \tag{4.5.5}$$

Belief–rational strategies for the fourth and higher levels of the regress can be determined in an identical manner.

The next level of rationality that P_i could consider P_j' to have, is where P_i assumes that P_j' believes P_i to be belief–rational. Again we consider the effect of this assumption on P_i's belief–rational strategy. As before, Π_1 can be determined by equation (4.5.1), as this still only depends upon P_i's initial beliefs about how P_j' will play. At the second level of the regress, the strategies are defined by the equations given in (4.5.4).

At the third level of the regress, the strategies are defined by

$$\begin{aligned} Q_3(R_3) &= \left\{ Q^* \,\middle|\, \overline{U}_i(Q^*, R_3) = \max_{Q \in \mathbf{Q}}\{\overline{U}_i(Q, R_3)\} \right\} \\ \Rightarrow R_3(Q_3) &= \left\{ R^* \,\middle|\, \overline{U}_j(Q_3(R_3), R^*) = \max_{R \in \mathbf{R}}\{\overline{U}_j(Q_3(R_3), R)\} \right\} \\ \Rightarrow \Pi_3 &= \left\{ Q^* \,\middle|\, \overline{U}_i(Q^*, R_3(Q_3)) = \max_{Q \in \mathbf{Q}}\{\overline{U}_i(Q, R_3(Q_3))\} \right\} \end{aligned} \tag{4.5.6}$$

So at this level, P_j' is thought to believe that P_i thinks that P_j' will play strategy R_3, and as P_i is thought to be utility maximising, P_i will play strategy $Q_3(R_3)$. Hence the belief–rational strategy for P_j' is $R_3(Q_3)$, and we can therefore determine the belief–rational strategy for P_i to be Π_3.

The fourth level of the regress is defined by

$$\begin{aligned} Q_4(R_4) &= \left\{ Q^* \,\middle|\, \overline{U}_i(Q^*, R_4(Q_4)) = \max_{Q \in \mathbf{Q}}\{\overline{U}_i(Q, R_4(Q_4))\} \right\} \\ \Rightarrow R_4^{(2)}(Q_4) &= \left\{ R^* \,\middle|\, \overline{U}_j(Q_4(R_4), R^*) = \max_{R \in \mathbf{R}}\{\overline{U}_j(Q_4(R_4), R)\} \right\} \\ \Rightarrow \Pi_4 &= \left\{ Q^* \,\middle|\, \overline{U}_i(Q^*, R_4^{(2)}(Q_4)) = \max_{Q \in \mathbf{Q}}\{\overline{U}_i(Q, R_4^{(2)}(Q_4))\} \right\} \end{aligned} \tag{4.5.7}$$

At this level, P_i believes that P_j' thinks that P_i thinks that P_j' thinks P_i will employ strategy Q_4, so P_j' is thought to play strategy $R_4(Q_4)$. So being utility maximising, P_i is thought (by P_j') to play strategy $Q_4(R_4)$, and hence P_j' is thought (by P_i) to play strategy $R_4^{(2)}(Q_4)$. Therefore Π_4 is the belief–rational strategy for P_i. Belief–rational strategies for the fifth and higher levels of the regress can be determined in precisely the same way.

The fourth level of rationality that P_i could assume is where P_i believes that P_j' thinks that P_i thinks that P_j' is belief–rational. The first, second and third levels of the infinite regress in this case are defined in precisely the same way as for the third level of rationality, i.e. from equations (4.5.1), (4.5.4) and (4.5.6) respectively. The fourth level of the regress is defined by the equations given in (4.5.7) in addition to

$$R_4(Q_4) = \left\{ R^* \;\middle|\; \overline{U}_j(Q_4, R^*) = \max_{R \in \mathbf{R}} \{\overline{U}_j(Q_4, R)\} \right\} \tag{4.5.8}$$

which gives rise to the other equations in (4.5.7). Again belief–rational strategies for higher levels of the regress can be determined in the same manner. Also, belief–rational strategies for further levels of assumptions about P_j''s rationality can be calculated by allowing higher levels of Q and R to be determined by the expected way that P_i and P_j' respectively are thought to think about their opponent.

So, at each further level of the regress we are increasing the overall size of the problem. In the same way that the size of the game that Howard is considering is increased by adding an extra player to the title of the metagame, so the size of the game that we are modelling increases. This is because at each higher level, a further set of thoughts about the opponent is considered, and therefore the complexity of the problem increases. By imposing the various levels of rationality on P_i's beliefs about P_j', we are restricting the set of strategies that P_i believes P_j' will employ. This in turn restricts the set of strategies that P_i has to choose from.

It is clear that however many levels we consider, we will still determine just a single strategy for P_i based on a strategy that P_i believes P_j' will employ. This is because any number of levels of beliefs that P_i considers, will all simply determine one strategy that is utility maximising over P_j''s strategies and P_j''s beliefs about P_i's strategy. This corresponds to the Howard method above where, whatever the size of the metagame and complexity of the metaequilibria, the corresponding move for each player in the original game can be determined. However, the solution under our approach is sensitive to the initial strategy R_1. By considering all possible initial strategies we can determine all possible belief–rational strategies for P_i.

One problem that exists with this approach is the ability of a player to determine his opponent's utility function. Now P_i's beliefs about the form of P_j''s utility function are assumed to be those stated above, and P_i can calculate his belief-rational strategy given these beliefs. Also, if P_i knows U_j with probability one, P_i can identify P_j with P_j' (at least mathematically), as the required expectations commute. However, P_j' does not necessarily know U_i. If we make no assumptions about the ability of P_j' to determine U_i, then there is not necessarily any method of truncating the infinite regress. We shall show that if P_j' is assumed to know U_i, and P_j' assumes P_i knows U_j, both with probability one at all stages of the game, then we can truncate the regress. If P_j' is assumed to only have a limited knowledge of U_i, then the possibility of truncating the regress, and the required number of levels to do so, depends upon the assumed beliefs. If the given set of assumed beliefs are not sufficient to truncate the regress, then an approximation to within a given bound can be determined by a finite number of levels, as is demonstrated by Mertens & Zamir (1985).

Before we prove any results, we shall show that the Howard metagame methodology is a special case of the above approach. Suppose that in the above approach, the strategy space is restricted to pure strategies, that the game is only a single play game, and that both players explicitly know each other's utility function. These utility functions are assumed to be simply linear in the pay-off achieved by the player concerned. Further to this, we assume that both players are utility maximising and assume their opponent to be so.

For each level of the infinite regress we can determine the moves that are available to the two players, and hence we can write down a pay-off matrix for the game being played. As the utility functions of the two players are assumed to be linear on pay-off, and both players assume each other to be utility maximising, Nash equilibria can be seen to be utility maximising strategies.

Now we must consider what the levels of the regress represent. The first level of the regress (for P_i) corresponds simply to the game iG (as defined in section 4.3 above) as P_i considers how P_j' will play the game and then plays accordingly. Similarly, the first level for P_j' corresponds to jG. The second level of the regress (for P_i) corresponds to the game ijG, as P_i considers how P_j' thinks P_i will play in order to determine how P_j' will play, and hence how P_i should play. The third level of the regress (for P_i) corresponds to the game $ijiG$, as P_i considers the game where he believes that P_j' thinks about how P_i thinks P_j' will play the game. These levels of the regress for P_j', and higher levels for both players, can be seen to define other metagames in exactly the same way.

From the utility assumptions, pure strategy equilibria is the relevant solution concept, and these can be determined by considering the maxima and minima over the moves available to each player, as was shown by Howard (see section 4.3 above). It is intuitively obvious that the regress can be truncated when the title contains all n players in an n player game, as otherwise the same maximum (or minimum) of a particular set is calculated more than once. So the truncation of the regress here is simply a consequence of the assumptions of the model. By making the same assumptions in our model, we can also truncate the regress, and we obtain precisely the Howard metaequilibria. In this sense our approach is a generalisation of Howard's approach, as we can make other utility assumptions, permit mixed strategies and consider multi–stage games.

In the method that has just been outlined, we initially take expectations over the population of players, to form the average opponent P_j'. This is not the only way of considering the problem. One other possible way is to carry out the analysis using the player's beliefs about this population throughout, and then take expectations at the end, in order to determine the expected utility maximising strategy. There are also other ways which could be employed, but I feel that the one presented is the most intuitive.

We would now like to determine under what beliefs it is possible to truncate the infinite regress that is defined above. To do this we need to define a notion of stability, which is defined on the belief–rational strategies for each level of the regress.

Definition. For a given level v of P_j''s rationality, an infinite regress has a *stable strategy* $\tilde{\Pi}_v$ at the rth level of the regress if

$$\tilde{\Pi}_v = \Pi_r = \Pi_s \quad \text{for all } s \geq r \tag{4.5.9}$$

for all values of $r = 1, 2, \ldots$, and all values of $v = 1, 2, \ldots$.

What we are saying here is that a regress has a stable strategy at a given level, if the belief–rational strategies from all higher levels of the regress are identical to the belief–rational strategy at the given level. Therefore if this is known to be the case, then a player need only calculate the belief–rational strategies up to the level where the stable strategy is defined, in order to determine all belief–rational strategies for that particular level of rationality. This can be seen to be a kind of equilibrium concept as, by employing this stable strategy, a player believes that his utility maximising strategy will be $\tilde{\Pi}_v$, when he believes his opponent to believe him to be playing $\tilde{\Pi}_v$. To prove the next three results we make the heroic assumptions that P_j' knows U_i, and that P_j' assumes P_i knows U_j, both with probability one at all levels of the regress. This assumption can be loosened later.

LEMMA 4.5.1.

At the third level of rationality, the stable strategy

$$\tilde{\Pi}_3 = \Pi_3. \tag{4.5.10}$$

PROOF: We begin by showing that, under the given conditions, $\Pi_3 = \Pi_4$, and then prove the lemma by induction on the level of the regress, where Π_3 and Π_4 are defined by the inequalities (4.5.6) and (4.5.7) above. Now we know that as P_j' is assumed to believe P_i to be utility maximising, and to know U_i with probability one, so at the third level of the regress, $R_3(Q_3)$ is such that

$$R_3(Q_3) = \left\{ R^* \;\middle|\; \overline{U}_j(Q_3^*(R^*), R^*) = \max_{R \in \mathbf{R}} \{\overline{U}_j(Q_3^*(R), R))\} \right\} \tag{4.5.11}$$

where $Q_3^*(R)$ is the utility maximising strategy for P_i given third level beliefs about the strategy R that P_j' will employ. Similarly, at the fourth level of the regress,

$$R_4^{(2)}(Q_4) = \left\{ R^* \;\middle|\; \overline{U}_j(Q_4^*(R^*), R^*) = \max_{R \in \mathbf{R}} \{\overline{U}_j(Q_4^*(R), R))\} \right\} \tag{4.5.12}$$

where $Q_4^*(R)$ is the utility maximising strategy for P_i given fourth level beliefs about the strategy R that P_j' will employ.

Now suppose that strategy $R_3(Q_3)$ achieves a higher utility than $R_4^{(2)}(Q_4)$, i.e. $\overline{U}_j(\Pi_3, R_3(Q_3)) > \overline{U}_j(\Pi_4, R_4^{(2)}(Q_4))$. However, $R_3(Q_3)$ is a strategy that is available to P_j' at the fourth level of the regress, and so P_j' could achieve a higher utility than was achieved through $R_4^{(2)}(Q_4)$. That is, there exists a strategy R such that

$$\overline{U}_j(Q_4^*(R), R) > \overline{U}_j(\Pi_4, R_4^{(2)}(Q_4)), \tag{4.5.13}$$

which contradicts the definition of $R_4^{(2)}(Q_4)$.

Now suppose that $\overline{U}_j(\Pi_3, R_3(Q_3)) < \overline{U}_j(\Pi_4, R_4^{(2)}(Q_4))$. Similarly, strategy $R_4^{(2)}(Q_4)$ is available to P_j' at the third level of the regress, and so P_j' could achieve a higher utility than was achieved through $R_3(Q_3)$. That is, there exists a strategy R' such that

$$\overline{U}_j(Q_3^*(R'), R') > \overline{U}_j(\Pi_3, R_3(Q_3)), \tag{4.5.14}$$

which contradicts the definition of $R_3(Q_3)$. Therefore we must have that the utilities for P_j' from $R_3(Q_3)$ and $R_4^{(2)}(Q_4)$ are always the same, and so they must always define the same set of strategies. So it is clear that the utility maximising strategies for P_i from the third and fourth levels of the regress, i.e. Π_3 and Π_4, must also determine the same set of strategies.

Further to this, suppose that $\Pi_3 = \Pi_4 = \cdots = \Pi_{s-1}$ for some value of $s \geq 3$, where Π_5, Π_6, etc. are defined as obvious extensions of the series Π_1, Π_2, Π_3, Π_4 already defined. For notational convenience, suppose that Π_s is calculated given the belief that P_j' will play strategy $R_s(Q_s)$, which is calculated given the belief that P_i will play strategy $Q_s(R_s)$, and that P_j' knows U_i with probability one. Again, from the rationality assumptions we have

$$R_s(Q_s) = \left\{ R^* \,\middle|\, \overline{U}_j(Q_s^*(R^*), R^*) = \max_{R \in \mathbf{R}} \{\overline{U}_j(Q_s^*(R), R))\} \right\} \tag{4.5.15}$$

where $Q_s^*(R)$ is the utility maximising strategy for P_i given sth level beliefs about the strategy R that P_j' will employ.

Now suppose that $\overline{U}_j(\Pi_3, R_3(Q_3)) > \overline{U}_j(\Pi_s, R_s(Q_s))$. However, $R_3(Q_3)$ is a strategy that is available to P_j' at the sth level of the regress, and so P_j' could achieve a higher utility than was achieved through $R_s(Q_s)$. That is, there exists a strategy R such that

$$\overline{U}_j(Q_s^*(R), R) > \overline{U}_j(\Pi_s, R_s(Q_s)), \tag{4.5.16}$$

which contradicts the definition of $R_s(Q_s)$.

Now suppose that $\overline{U}_j(\Pi_3, R_3(Q_3)) < \overline{U}_j(\Pi_s, R_s(Q_s))$. Similarly, strategy $R_s(Q_s)$ is available to P_j' at the third level of the regress, and so P_j' could achieve a higher utility than was achieved through $R_3(Q_3)$. That is, there exists a strategy R' such that

$$\overline{U}_j(Q_3^*(R'), R') > \overline{U}_j(\Pi_3, R_3(Q_3)), \tag{4.5.17}$$

which contradicts the definition of $R_3(Q_3)$. Therefore we must have that the utilities for P_j' from $R_3(Q_3)$ and $R_s(Q_s)$ are always the same, and so they must always define the same set of strategies. So it is clear that the utility maximising strategies for P_i from the third and sth levels of the regress, i.e. Π_3 and Π_s for any $s \geq 3$, must also determine the same set of strategies. Hence $\widetilde{\Pi}_3 = \Pi_3$. □

So this shows that at the third level of rationality, a player need only determine his beliefs about the first three levels of the infinite regress to calculate a belief–rational strategy for all higher levels. This is provided he is prepared to believe this typical opponent to be belief-rational and that this opponent believes him to be belief-rational. From this we quickly obtain the next result.

COROLLARY 4.5.2.

At the vth level of rationality, $v \geq 3$, the stable strategy

$$\tilde{\Pi}_v = \Pi_3. \tag{4.5.18}$$

PROOF: This follows directly from Lemma 4.5.1, when we notice that the proof depends only upon P_i believing that P_j' is utility maximising, and that P_j' believes P_i to be utility maximising. Any higher levels of rationality are superfluous, as the arguments only require the first three levels, and these are included in all higher levels. Therefore the lemma holds for all levels of rationality at or above the third level. □

We now use these last two results to prove the main result about the truncation of the infinite regress.

THEOREM 4.5.3.

For any levels of rationality, $v, s \geq 3$, the stable strategies

$$\tilde{\Pi}_v = \tilde{\Pi}_s. \tag{4.5.19}$$

PROOF: For convenience in this proof we shall place the level of rationality for a strategy Π as a superscript, e.g. $\Pi_4^{(3)}$ denotes the strategy that P_i adopts at the fourth level of the regress and at the third level of rationality.

From Lemma 4.5.1 we have that at the third level of rationality, $\tilde{\Pi}_3 = \Pi_3^{(3)}$ and hence,

$$\Pi_3^{(3)} = \Pi_4^{(3)} = \cdots = \Pi_r^{(3)} \tag{4.5.20}$$

for all $r \geq 3$. Also, we know from Corollary 4.5.2 that $\tilde{\Pi}_v = \Pi_3^{(v)}$, and $\tilde{\Pi}_s = \Pi_3^{(s)}$. So therefore all we need to show is that $\Pi_3^{(3)} = \Pi_3^{(s)}$ for all $s \geq 3$.

We also know from Lemma 4.5.1 that $\Pi_3^{(3)}$ is determined as the utility maximising strategy against $R_3(Q_3)$, which is calculated as the utility maximising strategy for P_j', given that P_i will play his (P_i's) belief-rational strategy. Therefore,

$$\tilde{\Pi}_3 = \left\{ Q^* \;\middle|\; \overline{U}_i(Q^*, R^*(Q^*)) = \max_{Q \in \mathbf{Q}} \{\overline{U}_i(Q, R^*(Q))\} \right\} \tag{4.5.21}$$

where $R^*(Q)$ is the utility maximising strategy for P_j', given that P_j' believes P_i will play strategy Q. But we also know from Corollary 4.5.2 that for all $s \geq 3$,

$$\tilde{\Pi}_s = \Pi_3^{(s)} = \left\{ Q^* \;\middle|\; \overline{U}_i(Q^*, R^*(Q^*)) = \max_{Q \in \mathbf{Q}} \{\overline{U}_i(Q, R^*(Q))\} \right\} \tag{4.5.22}$$

and the utility maximising strategy for P_i is $\tilde{\Pi}_s$ if he believes P_j' will play $R^*(\tilde{\Pi}_s)$. So any strategy that satisfies (4.5.21) must also satisfy (4.5.22), and the converse. Higher levels of rationality do not affect the belief–rational strategy, as the third level determines a stable strategy. Thus neither player can achieve a higher utility from a strategy determined by considering a higher level of rationality.

Therefore, the first three levels of the infinite regress at all levels of rationality (greater than three) are all that are required to determine the stable strategy. As the equations for these three levels are identical for all levels of rationality, $s \geq 3$, we must have that

$$\tilde{\Pi}_3 = \Pi_3^{(3)} = \Pi_3^{(s)} = \tilde{\Pi}_s \tag{4.5.23}$$ □

It should be noted that the stable strategy at the first level of rationality $\tilde{\Pi}_1$ is not necessarily equal to $\Pi_3^{(1)}$, and $\tilde{\Pi}_2$ is not necessarily equal to $\Pi_3^{(2)}$. This is because of the lack of assumed rationality, a player may not have the same beliefs at different levels of the regress.

So, by assuming that the typical opponent, P_j', is utility maximising, and also that he (P_j') believes P_i to be utility maximising, P_i need only consider up to the third level of the infinite regress to determine all stable strategies. Also there is no need to consider any higher levels of rationality about the opponent, than the third level.

Now these results were based on the assumptions that P_j' knows U_i and P_j' assumes P_i knows U_j with probability one, at all levels of the regress. It is clear that as long as these assumptions hold at one level of the regress, k, and all subsequent levels, then these results will hold in the same way. The only modification would be that instead of the stability holding from the third level onwards, it would be from the $(k+3)$rd. level onwards. Note that the results would still hold for the third level of rationality.

As we have seen, some sets of assumptions lead to the truncation of the infinite regress. It should also be pointed out that we have only considered simple, idealised sets of assumptions. It is likely that other similar results could be derived for different sets of assumptions, but these would depend crucially on the beliefs of the players concerned. Comparisons of the stable outcomes determined here with the outcomes found by other authors (e.g. Aumann's correlated equilibria) can be made, and these will be discussed in the next chapter.

It is also possible that no truncation of the regress will be possible, due to lack of explicit assumptions about the forms of the players' utility functions. As we have said above, this leads to the necessity to consider all levels of the infinite regress. However, it would seem unlikely that players of such games would consider more than a few levels of such a regress. A natural route to take in this case is to determine a distribution for each of the players over the number of levels of the regress that they believe their opponent will consider. In most parlour games (e.g. chess, bridge) players rarely go to levels higher than four or five, despite a whole infinite regress existing. Other factors, such as the familiarity of the players may be important here.

So it would seem that the players of such games naturally limit the number of levels considered, presumably due to time constraints, or limits to the memory or intelligence of the player concerned. Therefore it would seem logical for a player to put a distribution over the number of levels believed to be considered by an opponent. Unfortunately, I have not had time to develop a model with such a feature here, or to determine the likely results of such a methodology.

The infinite regress that has been detailed above can be limited to a small, finite number of levels by appealing to the notion of belief-rationality, and by making assumptions about the players' beliefs about their opponent's utility function. If these assumptions are unreasonable for the game in question, then the whole infinite regress need not be considered, as a finite approximation can be found. We shall question the existence and uniqueness of the belief-rational strategies for P_i, as well as the complication of more than two players, later in this chapter. Before that we consider a couple of examples.

4.6 Two Examples.

We shall now explore two examples of the type of infinite regress that we are considering, and calculate the belief-rational strategies that are determined by the method given in the previous two sections.

4.6.1 A Prisoner's Dilemma Game.

Suppose that two players are playing the PDG given by the pay-off matrix in Figure 4.6.1.

$$\begin{array}{ccc} & & P_j \\ & & \begin{array}{cc} C & D \end{array} \\ P_i & \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} (5,5) & (-5,10) \\ (10,-5) & (0,0) \end{pmatrix} \end{array}$$

Figure 4.6.1

As we have discussed before, the only Nash equilibrium for this game is the (D, D) outcome. Also Howard showed that the metaequilibria for this game are the (D, D) outcome and the (C, C) outcome. We shall again concentrate on a 'typical opponent' P_j'. To conform with the results of the previous section we shall assume that P_j' knows P_i's utility function U_i with probability one at all levels of the regress.

Our method begins by assuming a strategy R_1 that P_i believes P_j' will play. We shall assume this strategy to be such that P_j' will make move C at a proportion α of all future stages of the game, and will make move D at a proportion $(1-\alpha)$ of all future stages, for some $\alpha \in [0,1]$. It is easy to see that whatever this strategy is, i.e. whatever the value of α is, the utility maximising strategy at the first level of the regress (for any utility function that is increasing with pay–off) for P_i is to play move D with probability one. For the purposes of this example we shall consider the utility functions for both players to be (discounted) linear functions of the pay–off from the whole game to the player concerned. So we therefore have that, irrespective of α, Π_1 is the continual Defection strategy. Also, if we make no assumptions about the rationality of P_j', it is clear that a belief–rational strategy at any level of the regress must be the strategy that plays move D with probability one.

Now we consider the second level of P_j''s rationality, i.e. where P_i assumes that P_j' is belief–rational. At the first level of the infinite regress, the belief–rational strategy must again be the continual defection strategy. At the second level, P_i considers how P_j' is thinking about the game. As P_j' is assumed to be utility maximising, P_i will expect P_j' to play the continual defection strategy. Therefore the belief–rational strategy for P_i, Π_2, must also be the strategy that makes move D with probability one.

If we then consider the third level of P_j''s rationality, i.e. where P_i assumes that P_j' believes P_i to be belief–rational, then we obtain different results. At the third level of the regress, P_j' takes into account how P_i thinks P_j' will play. Therefore to calculate any belief–rational strategy for P_i we must consider how P_j' will play, which depends upon how P_j' thinks P_i will play. So we must consider all strategies Q_1 available to P_i. We will denote by $Q_1(\beta)$ the strategy that makes move C at proportion β of all future stages, and makes move D at a proportion $(1-\beta)$ of all future stages.

Therefore we need to question how P_j' will play if he believes P_i will play strategy $Q_1(\beta)$. To determine this we must consider the three possible cases: $\beta < \alpha$, $\beta = \alpha$ and $\beta > \alpha$. Firstly if $\beta < \alpha$, then P_j' is receiving a lower pay–off than P_i, and so P_j' will think that P_i is exploiting him. From this it is assumed that P_i would play $Q_2(\beta')$ if he believed P_j' would play $R_2(\alpha')$, where $\beta' < \alpha'$. Therefore to pre–empt this further exploitation P_j' is expected to play the continual Defection strategy, with the expected outcome (D, D) and pay–off $(0, 0)$.

Secondly, if $\beta = \alpha$, then P_j' has three basic choices of strategies to employ, as summarised in Figure 4.6.2, together with the respective expected utilities

R_2	$\alpha' < \alpha$	$\alpha' = \alpha$	$\alpha' > \alpha$
exp. utility	0	5α	$< 5\alpha$

Figure 4.6.2

The expected pay–offs for each of these is calculated as follows. If it is thought that P_j' will play strategy $R_2(\alpha)$ then P_i is expected to play strategy $Q_2(\alpha)$, giving the outcome

$$\alpha^2(C,C) + \alpha(1-\alpha)(C,D) + \alpha(1-\alpha)(D,C) + (1-\alpha)^2(D,D) \qquad (4.6.1)$$

with expected utility $5\alpha^2 - 5\alpha(1-\alpha) + 10\alpha(1-\alpha) = 5\alpha \geq 0$ to both players.

If it is thought that P_j' will play strategy $R_2(\alpha')$ where $\alpha' > \alpha$ then the expected utility must be less than 5α, as P_i will be expected to play strategy $Q_1(\alpha)$. If it is thought that P_j' will play strategy $R_2(\alpha')$ where $\alpha' < \alpha$, then P_i will be expected to play the continual defection strategy, with expected utility to both players of 0. Therefore the utility maximising strategy for P_j' to employ is $R_2(\alpha)$.

Thirdly, if $\beta > \alpha$, then P_j' has five choices of strategy $R_2(\alpha')$ to employ. These are summarised in Figure 4.6.3, with the expected utility from all future moves calculated in the same manner as was described above. From this we can see that the utility maximising strategy for P_j' to employ is $R_2(\alpha)$. Therefore P_i's utility maximising strategy is $\Pi_3 = Q_1(\alpha)$, and the expected strategy from P_j' is $R_2(\alpha)$, and so the expected utility is 5α to both players. From the theorem of the previous section we know that this is a stable strategy, given P_i's prior beliefs about P_j'.

R_2	$\alpha' < \alpha$	$\alpha' = \alpha$	$\alpha' \in (\alpha, \beta)$	$\alpha' = \beta$	$\alpha' > \beta$
exp. utility	0	$5\beta + 5(\beta - \alpha)$	$5\beta + 5(\beta - \alpha')$	5β	$< 5\beta$

Figure 4.6.3

This example shows how the solution is sensitive to the prior setting of R_1. For any given prior belief about R_1, a different utility maximising strategy for P_i will result. Also this example demonstrates that Nash equilibria and Howard's metaequilibria are determined by our method. We obtain the former by having a prior setting of R_1 on the strategy that makes move D with probability one (i.e. $\alpha = 0$), and the latter by having the prior setting of either one pure strategy or the other (i.e. $\alpha = 0$ or 1).

We ought to note at this point that the stable strategy for this PDG is not the stable strategy for all PDGs. For example, consider the PDG given by the pay-off matrix in Figure 4.6.4.

$$\begin{array}{ccc} & & P_j \\ & & \begin{array}{cc} C & D \end{array} \\ P_i & \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} (5,5) & (-90,10) \\ (10,-90) & (0,0) \end{pmatrix} \end{array}$$

Figure 4.6.4

It can be checked that the only belief-rational strategies in this game when the prior is $R_1(\alpha)$, are $\Pi_3 = Q_1(\alpha)$ for $\alpha > \frac{16}{17}$, as well as the strategy that plays move D with probability one ($\Pi_3 = Q_1(0)$).

This example is provided to illustrate how the above approach can be used to determine stable strategies, and in this case the beliefs are in terms of the players' overall strategies. Obviously more specific strategies that determine particular moves can also be incorporated into the above methodology, as is demonstrated in the next example.

4.6.2 A Competition.

Suppose that two players (P_i and P_j) enter a competition whereby the winner is the first player to achieve a given number of points, M (> 20). The players receive points by playing the repeated game given by the pay-off matrix in Figure 4.6.5.

$$
\begin{array}{ccc}
 & & P_j \\
 & & \begin{array}{cc} n_1 & n_2 \end{array} \\
P_i & \begin{array}{c} m_1 \\ m_2 \end{array} & \left(\begin{array}{cc} (2,0) & (5,5) \\ (0,8) & (6,2) \end{array}\right)
\end{array}
$$

Figure 4.6.5

The game is assumed to finish as soon as one player achieves M points. So in this game the utility function for the two players over the final outcomes of the game is not linear, but P_i's utility function is assumed to be of the form

$$
U_i = \begin{cases} 1 & \text{if} \sum_k x_k \geq M \text{ and} \sum_k y_k < M, \\ \frac{1}{2} & \text{if} \sum_k x_k \geq M \text{ and} \sum_k y_k \geq M, \\ 0 & \text{if} \sum_k x_k \geq M. \end{cases} \tag{4.6.2}
$$

where x_k is the pay-off to P_i at stage k of the game, and y_k is the pay-off to P_j at stage k of the game. Again we shall construct a typical opponent, P_j'. $\overline{U}_j$ is defined similarly for P_j' (but with the x_k's and y_k's reversed). Also we shall assume that P_j' knows U_i with probability one.

From the results of the last section we know that to determine a stable strategy for P_i for this game we need to assume that P_j' is belief-rational, and believes P_i to be belief-rational. We then need to calculate Π_3, given P_i's prior beliefs about how P_j' will play the game. Suppose it is assumed that P_j' will play move n_1 with probability one. From this we can calculate that the only belief-rational strategy for P_i is to play move m_1 and hence the utility maximising strategy for P_j' is to play move n_2. Therefore the only stable outcome is the move pair (m_1, n_2).

This outcome will obviously give the same pay-off to P_i as it will to P_j'. If it is believed that on the later stages of the game, the players will continue with these strategies, then equal totals are to be expected. How P_i plays this game when the sums of pay-offs approach the critical value M, will depend on his subjective beliefs about how P_j' will play that next stage. End-game effects such as whether P_i should play move m_2 on the last stage of the game, or on the penultimate stage, and how P_j will play in such a situation can only be determined at that stage, and would be highly dependent on the actual value of M.

4.7 Stability of the solutions.

First of all we consider the existence and uniqueness of the belief–rational strategies given in section 4.5, i.e. when is it possible to construct such a strategy, and if we can, is it unique? To answer these questions we shall concentrate on the case where P_j' is believed to know U_i with probability one at all levels of the regress.

Now if the prior beliefs that P_i has over P_j''s future play are that P_j' will play a Nash equilibrium strategy, and will continue playing this strategy irrespective of P_j''s beliefs about P_i's strategy, then the utility maximising strategy for P_i must be the corresponding strategy for this Nash equilibrium. This Nash equilibrium strategy is therefore a stable strategy. All Howard metaequilibria can be seen to be stable strategies as well, in the same sense. If P_i's beliefs about P_j' are that he will play the move determined by a metaequilibrium at each stage of the game, then the belief–rational, stable strategy for P_i must be to play the corresponding move from the metaequilibrium at each stage of the game. Also, there can be outcomes that are determined by the stable strategies other than the metaequilibria. This is demonstrated by example 4.6.1 in the previous section.

From this we can deduce that we can always determine a belief–rational strategy for a given game. This follows from the theorem of Nash given in section 3.2 of chapter 3 above, which states that there is always an equilibrium outcome in all games of the type we are considering. Therefore, we can always find at least one belief–rational strategy for P_i for any given game.

Now we consider the uniqueness of a belief–rational strategy for given prior beliefs in a particular game. We can see that such belief–rational strategies will not necessarily be unique by considering the trivial example given by the pay–off matrix in Figure 4.7.1.

$$\begin{array}{ccc} & & P_j \\ & & \begin{array}{cc} n_1 & n_2 \end{array} \\ P_i & \begin{array}{c} m_1 \\ m_2 \end{array} & \begin{pmatrix} (3,1) & (3,2) \\ (3,3) & (3,4) \end{pmatrix} \end{array}$$

Figure 4.7.1

Suppose that U_i is any increasing function of P_i's pay–off. Then for any strategy R_1 that P_i believes P_j' might play, P_i will obtain exactly the same utility for a particular strategy Π as he will for any other strategy Π'. Therefore, for any given prior beliefs about P_j' in this game, any strategy that P_i plays will be belief–rational.

In general, if we have a belief–rational strategy Q for P_i, given a prior belief that P_j' will play strategy R_1, it may be possible to construct another strategy $Q' \neq Q$ such that Q' is also a belief–rational strategy for P_i, given R_1. Therefore the strategy determined by P_i to be stable is not necessarily unique. However, as both these strategies are utility maximising with respect to R_1, they must obtain the same expected utility for P_i. If we then assume the third level of rationality, both of these strategies must also be stable, and thus it makes no difference to P_i which of these utility maximising strategies he chooses to play. This is in direct contrast to multiple Nash equilibria, where if a player chooses one equilibrium when his opponent chooses another, a disastrous result could occur.

Next we consider the stability of these belief–rational strategies to changes in inital beliefs about P_j'. As stated above, there is always a belief–rational strategy for any given prior belief. If this is unique then there is a unique outcome that P_i believes will occur. If it is not unique then, as we have just argued, P_i will be indifferent between the alternative outcomes. It is also quite likely that several different prior strategies for P_j' will lead to the same belief–rational strategy. Indeed, all priors could lead to the same belief–rational strategy, like in the game given by the pay–off matrix in Figure 4.7.2, where whatever the prior beliefs about P_j' are, the only utility maximising strategy for P_i is to play move m_1 with probability one.

$$\begin{array}{ccc} & & P_j \\ & & \begin{array}{cc} n_1 & n_2 \end{array} \\ P_i & \begin{array}{c} m_1 \\ m_2 \end{array} & \begin{pmatrix} (5,5) & (5,5) \\ (0,0) & (0,0) \end{pmatrix} \end{array}$$

Figure 4.7.2

These above examples have all concentrated on the case where P_j' is believed to know U_i with probability one. If other assumptions are made about the beliefs that P_j' has about U_i, then different results may occur. It is, however, unlikely that these beliefs will always determine a unique belief–rational strategy for games such as that given by Figure 4.7.1, and more general results for the existence of belief–rational strategies could also be proved. Explicit results can be derived, but these will depend upon the assumed beliefs.

So far we have only considered the two player game. The problem is complicated further by the presence of more than two players. Suppose there are n players: $P_i, P_j, \ldots P_k$, playing a particular game G. Again we shall consider the problem from P_i's point of view. In this game, P_i must determine his prior beliefs about how each of the other players will play the game. From these beliefs P_i can determine his utility maximising strategy. We assume that the opponents $P_j, P_k, \ldots$ are replaced by 'typical opponents' $P'_j, P'_k, \ldots$ by averaging over the populations from which they have been drawn.

Then, as in the two player game, we consider the rationality of the opponents, and assume them to be utility maximising with respect to how they believe all the other players will play. Here we assume that, unless there is evidence to the contrary, that P'_j $(j \neq i)$ has the same view about how P'_k $(k \notin \{i, j\})$ will play as P_i does. P_i can then determine a utility maximising strategy against the resulting combined strategies of the other players.

Then we must consider how any particular opponent will think P_i and the other players will play. Again, unless there is evidence to the contrary, we make the assumption that P'_j has the same view about how P'_k will play as P_i does. The infinite regress is then formed in the same way as for the 2 player game, i.e. on the number of levels of thoughts that any player takes into account. By imposing belief–rationality on the opponents, we determine a belief–rational strategy for P_i for each level of this regress. Further to this we assume that P_i believes that P'_j $(j \neq i)$ thinks that all other players P'_k $(k \neq j)$ are belief–rational. We can then determine a stable strategy for P_i for each level of the regress, provided that we make an assumption regarding the beliefs of the other players about each others' utility functions.

By making these simplifying utility assumptions, this can be seen to be identical to the 2 player game considered earlier, but notationally much more complex. If we make the same assumption about the beliefs about the utility functions as before, the result of Theorem 4.5.3 also goes through in the same way as in the 2 player game. Therefore the belief–rational strategy from the third level of the regress is stable. P_i is therefore determining the belief–rational strategy by considering the expected combined strategies from $n - 1$ belief–rational players, as opposed to one belief–rational player. The fact that it is more than one other player does not affect the form of the analysis. If, however, the simplifying assumptions are not made, then the problem soon becomes very complicated and the above analysis will not necessarily hold.

Also, to determine all belief–rational strategies, all possible combinations of prior strategies for the other players must be considered. Thus the problem is much more complicated than that for two players due to the higher dimensionality. It is therefore a lot harder to determine all belief–rational strategies, but it is obviously mathematically possible. Also, the existence of at least one such belief–rational strategy is guaranteed, by the existence of at least one equilibrium in any n player game, due to Nash (1951). The availability of more than two moves for each of the players gives rise to similar complex, but not insurmountable problems.

The problem in the two player game of the beliefs of P_j' about U_i are magnified in the n player analogue. If it is unrealistic to make the assumption that P_j' knows U_i with probability one, then other results need to be proved. It may be possible that insufficient is known about these beliefs to be able to truncate the regress, but this would depend upon the individual case.

We have considered this problem from the point of view of one particular player — P_i. As mentioned earlier, it has been said that a problem cannot be considered fully, unless all players' views are taken into account. This depends to a large extent on what problem is trying to be solved. If, as Howard claims to, one is trying to explain *how* people play a given game then, unless the game is symmetric and all players are considered to be identical, each player's views must be incorporated into the model. If, on the other hand, one is trying to determine how a player *should* play a particular game, given his views about the other players, then it is acceptable just to look at one player in that position. This is the angle that we are taking. However our approach can be extended to the case where a rationale of why players made particular moves in a game is required. This is done by determining the belief–rational strategies for all players, with all possible prior beliefs about their opponents, in order to find the set of all outcomes that are belief–rational for all players simultaneously.

It has been argued (by, for example, Kadane & Larkey, 1982 and Laskey, 1985) that any rationality assumption of the kind in this chapter is contrary to the ideas of a subjective Bayesian methodolgy. What we are saying here, is that in determining how a player believes his opponent will play, a player will use his subjective probabilities, but these are guided by the fact that his opponent is utility maximising with respect to some (not necessarily linear) utility function. As the game progresses and the player receives more information about his opponent, he will update his beliefs about how his opponent will play. This argument is developed in chapter 6 of this thesis.

How this relates to the work of other authors in this area is discussed in the next chapter. Also, the various perspectives that these authors have adopted are discussed there.

4.8 Conclusions.

We have developed a framework that allows us to consider the infinite regress, by defining each level in terms of the players' thoughts and the rationality of the players. By making assumptions about the utility functions of the players we can truncate this regress, and therefore determine stable solutions. Using the derived framework we can discuss the effect of various beliefs about the utility functions on the stability of the regress.

By making given assumptions we can determine the models used by other authors, such as Howard. Other assumptions lead to stable solutions to games that have not previously been discussed. These solutions must be treated as a priori optimal, as beliefs will be updated as more information is received.

5. A REVIEW OF BAYESIAN GAME THEORY

5.1 Introduction.

We shall now give a review of Bayesian game theoretic results. Like the 'classical' game theory that was reviewed earlier, the Bayesian game theory has been developed from the work of von Neumann & Morgenstern (1947) and Luce & Raiffa (1957). Indeed, it could be argued that Bayesian game theory is an extension of the classical game theory, and therefore draws upon all previous results from this area. As mentioned at the end of chapter 3, Bayesian game theory is concerned more with determining the utility maximising choice of strategy for a player, as opposed to determining equilibria for a game. In finding this utility maximising choice, one can take into account the subjective beliefs of the player about the game, or about his opponent. Because of this, it would appear that Bayesian game theory lends itself more to games with incomplete information than those with complete information, although obviously the latter can be considered as a special case of the former.

Of considerable importance in Bayesian game theory models is the rationality of the opponent. If no assumptions about the rationality of the opponent are made, then the decision problem becomes just a maximisation over a number of variables. However, it seems a sensible modelling assumption that the opponent is as intelligent as the player under consideration, until any evidence to the contrary has been received. This creates problems in determining optimal play, due to an infinite regress of beliefs of how the player will act, of the kind considered in the previous chapter. Because of problems such as these, Bayesian game theory has produced a number of different methods of determining solutions, on the basis of different sets of assumptions.

5.2 Review.

One of the most influential writers in Bayesian game theory is Robert Aumann. He has produced a number of stimulating papers on topics of current concern in the subject, and has certainly been instrumental in advancing the theory of Bayesian games. His work covers a large area, including a formulation of subjective probability, work on common knowledge, cooperative games, games with infinitely many players, and also the correlation of strategies.

By common knowledge I mean that the players both know some fact, know that each other know it, know that each other know that each other know it, etc. Aumann (1976) shows that when two players have equal priors over some parameters, and their posteriors over these parameters are common knowledge, then these posteriors must be equal. It is argued that information will continue to be exchanged until the posteriors of the players are equal. This therefore gives a theoretical foundation for the reconciliation of subjective probabilities. However, this analysis is based on the assumption of equal priors for the players, or that all differences in subjective probabilities can be explained by differences in information that has been received. This is a neat and useful result if the conditions are met, but it is not clear when such conditions will be met in the context of a game. In some games, such as experimental games, players may hold similar initial beliefs to their opponents, and so this result gives some credence to symmetric models of the type discussed in chapter 6 of this thesis.

Aumann et al. (1983) shows that under weak conditions on the probability measure, any mixed strategy can be replaced by a pure strategy that achieves an expected pay–off within a specified bound of the original strategy. This result is useful, as then the players do not have to randomise to determine which strategy to play. An exact corresponding pure strategy is not always possible to find, even in simple cases. In games with complete information it has been considered necessary to randomise, so that the opponent cannot determine the strategy that the player is employing, but with incomplete information this is not necessary. This result ties in with the well known result that in a game with complete information, if a mixed strategy is a best reply to a given strategy by the opponent, then all pure strategy components of the mixed strategy are also best replies to the same strategy by the opponent. So it is not necessary for our models to incorporate randomised strategies.

The main concepts that Aumann has developed are correlated strategies, and correlated equilibria. In Aumann (1974) the idea of correlation of randomised strategies in non–cooperative games is introduced through the use of differing subjective probabilities. It is shown that the set of Nash equilibrium pay–offs correspond to the set of objective mixed equilibrium pay–offs, but by using correlated strategies, one can achieve equilibria with higher expected pay–offs than any Nash equilibrium. Also zero–sum games can achieve equilibria with positive expected pay–offs to both players by the use of subjectively randomised strategies. The differing subjective probabilities required for this method to work may appear to be in contrast to the result in Aumann (1976) that differences in subjective probabilities can be explained by differences in information (referred to as the *Harsanyi Doctrine* after Harsanyi, 1967). However, the irreconcilable priors assumed here can be allowed due to the "complex informational situation" that the players are in.

Aumann (1987) continues with the concept of correlated equilibria, and shows that if each player is Bayes rational (i.e. utility maximising), then a correlated equilibrium distribution will result. The results here are based on the Harsanyi doctrine mentioned above, but it is shown that this is not necessary. In line with Aumann et al. (1983) it is noted that there are problems with using randomised strategies, and hence such mixed strategy equilibria are rather unnatural. Aumann (1974) also shows that any convex combination of Nash equilibria can be viewed as a subjective correlated equilibrium. Later in this chapter we compare these correlated equilibria with the ideas that we shall pursue in this thesis.

A useful survey of repeated games is given in Aumann (1981) which includes a Bayesian view of equilibrium, showing again that by the assumption of Bayes rationality, an equilibrium results. It is claimed that this dispenses with the usual dichotomy between Bayesian game theory and classical game theory — i.e. between utility maximisation with respect to subjective probabilities, and equilibria. These results are only obtained by assuming a Bayesian methodology similar to that used in this thesis.

Another very important author in Bayesian game theory is John Harsanyi. His papers in the late 1960's on games with incomplete information provided a novel theory that has provoked a considerable amount of research into Bayesian game theory. In this sense, Harsanyi's work has been crucial to the development of the subject. Also Harsanyi's (1977) book provides a general theory of rational behaviour in cooperative, non–cooperative and bargaining games.

Harsanyi (1967) introduces the notion of games with incomplete information and discusses the causes and implications of such a game. One problem with games of incomplete information is that a player's strategy choice depends on an infinite regress of the kind discussed in the previous chapter of this thesis. Harsanyi describes such a model as a ***sequential–expectations*** model. Players are assumed to be determined by random events called attribute vectors, and the players are assumed to know the joint distribution of these events. This can be seen to be the same as the idea used in this thesis of players being drawn from a known population. Players are assumed to know the strategy spaces of all players, and games are considered in normal form. It is postulated that for every game of incomplete information, G, an equivalent game of complete information (or *Bayesian game*), G^*, can be found.

A Bayesian equilibrium point is defined in Harsanyi (1968a) to be where the strategy for a particular player maximises his expected pay–off over the other players' normalised strategies. It then follows that if a set of strategies form a Bayesian equilibrium point in the game G, then it also forms a Nash equilibrium in the Bayesian game G^*. Therefore, in every finite game there is at least one Bayesian equilibrium point. An example is given to show that the optimal strategies will not necessarily have minimax or maximin properties. It can be seen that under this model a player will be in a position to exploit any mistaken beliefs by his opponent. Games are also classified into 'immediate commitment' and 'delayed commitment' games, depending on whether a player must determine his strategy before or after the chance move determining attributes has occured. Bayesian games are delayed commitment games and so the normal form of such a game does not fully represent it. As we shall discuss below, the theory developed in this thesis is not dependent on the outcomes of chance moves such as these, but simply on a particular player's beliefs. Transformations to other games are not necessary in this case.

In Harsanyi (1968b) it is shown that when a joint probability distribution over the players' attribute vectors exists, then the game is *consistent* and this distribution is unique. Harsanyi then uses an 'outside observer' perspective to argue that players should only use information common to all players, and in which case any player can determine a consistent distribution over all players' attribute vectors. Overall consistency is, however, not possible because of the players subjective probability distributions. So Harsanyi argues that players should make any estimates as independent as possible of their own personal prejudices when attempting to construct a consistent distribution, in line with the 'Harsanyi Doctrine'. If this is accepted, then other players' inconsistent subjective distributions can be represented by a larger set of consistent distributions. If information suggests that mutual inconsistencies exist, then no Bayesian game will exist, but a game with immediate commitment will (termed a *Selten game*). The method that we shall develop will permit consistent and inconsistent distributions.

Harsanyi (1977) tries to determine a general theory of rational behaviour in three specific areas of study: decision theory, game theory and ethics. It is claimed that rationality postulates in game theory only usually address themselves to individual decision theory. Harsanyi defines rational behaviour in terms of goal–directedness, as an extension of human beviour, and then conflicts of interest and/or common interest are easy to incorporate. The theory is based upon subjective beliefs of players and the principle that they are trying to maximise some expected utility function. This is very much the approach that we are taking. From this basis, eight rationality postulates are determined in terms of how people should play in game situations.

The rest of the sections of the book that are devoted to non-cooperative game theory are chiefly concerned with unprofitable games and maximin solutions. Unprofitable games are such that no equilibrium exists that obtains more than a players security (maximin) pay-off. It is claimed that the solution to such games is independent of a player's expectations, as all players must always adopt their maximin strategies. We find it strange that after defining rationality in terms of subjective beliefs, Harsanyi should be so preoccupied with equilibria. The arguments presented imply that the mutual cooperation outcome of a PDG is not feasible. It is hoped that this thesis takes the rationality ideas of Harsanyi away from the context of equilibrium, and toward a more subjectivist, utility maximising concept.

Harsanyi's work is therefore amongst the most important in Bayesian game theory, if not the most important. The work on rationality, and on games with incomplete information has provoked considerable research, and still remains the starting point for new theories in the area. Harsanyi was also one of the first authors to consider non-linear utility functions, that are not known precisely by the other players. However, Harsanyi concentrates on obtaining equilibria for the particular game in question. This leads to a belief that his work could be extended by using his theory to determine optimal moves, and explanations of suboptimal moves. This is particularly important in repeated games, where equilibria in the generating game are of limited interest. Therefore it would appear that it is in repeated games that Harsanyi's work could be best extended.

An author that has collaborated with Harsanyi is Reinhard Selten, although this joint work has mainly been in the bargaining area (e.g. Harsanyi & Selten, 1972). Selten has produced a number of excellent works in Bayesian game theory, such as Selten (1978) on the Chain Store paradox, and Selten (1964) on n-person games, but his main contribution is the concept of *perfect equilibria* or *trembling hand equilibria*.

A perfect equilibrium on an extensive form game is defined by Selten (1975) by assuming complete rationality to be the limiting case of incomplete rationality. An equilibrium is defined to be subgame perfect if it induces an equilibrium on every subgame of an extensive form game, and therefore is in equilibrium irrespective of past moves. Then the possibility of slight mistakes (or *trembles*) is considered, whereby there is a small probability of a player choosing any available strategy, which is determined by an unspecified psychological mechanism. A perfect equilibrium of a game is defined to be a limit of a sequence of games that are perturbed by probability ϵ of these slight mistakes, as ϵ tends to zero. It is then shown that these perfect equilibrium points are subgame perfect equilibrium points. Also, every normal form game, and every extensive form game with perfect recall, has at least one perfect equilibrium point. However, a perfect equilibrium point of a normal form game does not necessarily correspond to a perfect equilibrium point of the corresponding extensive form game. The advantages of perfect equilibria are that they do not permit weakly dominated strategies to be included in equilibrium points, and the possibility of threats is excluded. As this concept is defined on extensive form games we shall not be using these results directly, but related concepts for normal form games can also be found.

Acceptable correlated equilibria are related to perfect equilibria in the sense that they are correlated equilibria that are stable against trembles, and are considered by Myerson (1986). Again they are determined as the limit of a sequence of perturbed games as the probability of the tremble tends to zero. It is shown that the set of acceptable correlated equilibria is a non–empty subset of the set of correlated equilibria, and that every perfect equilibrium is an acceptable correlated equilibrium. Myerson also defines an acceptable response to be where all unacceptable actions (actions that a player can not use rationally when the probability of trembling is arbitrarily small) have been eliminated. Then a 'predominant action' is acceptable in all residues as the probability of trembling tends to zero, and it is shown that the set of predominant correlated equilibria include at least one Nash equilibrium, and is therefore always non–empty.

Also, the notions of perfection and domination can be seen to be logically similar. By moving from perfection to domination we determine another concept for normal form games, that of rationalizability as considered by Pearce (1984). We agree with Pearce that it is misleading to study repeated games by just considering Nash equilibria, and rationalizability provides one strategic solution to a game. Pearce defines a rationalizable strategy by iterative means, as a strategy that is a utility maximising response to a combination of strategies that were previously thought to be rationalizable. This is obviously a decreasing set, and can be seen to converge to a set of rationalizable strategies, that it is assumed all players will use. Nash equilibria are obviously included in such a set. Pearce also defines a stronger concept of 'cautious' rationalizability which is where there is always a strategy combination that gives positive probabilities to *all* pure strategies available.

Another author that has worked with the concept of rationalizability is Bernheim (1984). Bernheim focuses mainly on the properties of rationalizable strategies, rather than refinements of them. Again the rationalizable strategies are defined iteratively, and it is shown that there is a non–empty set of rationalizable strategies for any game. Bernheim shows this only for pure strategies, as any pure strategy that is a component of a rationalizable mixed strategy is also rationalizable as a pure strategy. However, unless the Nash equilibrium is unique, globally stable, and satisfies a strict form of local stability, there will be an infinite number of rationalizable strategies, and so multiplicity is recognised as a problem. Bernheim then considers modifications of rationalizability in order to eliminate certain undesireable strategies. He considers ***perfectly rationalizable*** strategies that are the limit of ϵ–rationalizable strategies as ϵ tend to zero, and ***subgame rationalizable*** strategies that are a best response in every proper subgame of an extensive form game. Note that perfect rationalizability is very similar to Pearce's cautious rationalizability, but not identical to it.

The rationalizability ideas presented in Pearce (1984) and Bernheim (1984) are similar to those presented as belief–rational strategies in the previous chapter of this thesis, and are based on very similar assumptions to those that we are using. The two methods are working in opposite directions, in that our method determines the best of an increasing number of possible strategies, whereas Pearce and Bernheim find decreasing sets of strategies that are rationalizable. No work has yet been done in comparing the two methods under the same assumed utility structure, but such work may well provide interesting results.

An interesting extension to the work of Aumann (1987) and Selten (1975) is provided in Shin (1988a). He shows that Aumann (1987) formulates his concept of Bayes–rationality in a different way to that of Savage (1954), that Aumann claims is the same. Shin shows that Aumann's formulation allows players to choose a probability distribution over the state space, and each state specifies the actions taken by the players (which are therefore fixed). In Savage's (1954) framework, the probability distribution is fixed, and each player then chooses a function to maximise his expected pay–off. Essentially this is due to the difference between considering the problem from the view of a third party, or an actual player. Our approach, as is discussed in the next chapter, is more in line with Savage's framework than Aumann's, as we are assuming a player's probability distribution to be of a given form, and then the player chooses a strategy to maximise his expected utility.

Shin (1989) provides a more general notion of equilibrium, which is termed *ratifiability* after Jeffery (1983) (or *perfect correlated equilibria* in Shin, 1988a). It is shown that a concept (modestly ratifiable) equivalent to Aumann's correlated equilibrium can be determined from the Nash equilibrium by allowing 'trembles' by just the player under consideration (i.e. not the opponent). Also a second ratifiable concept that is equivalent to Selten's perfect equilibrium can be determined by allowing both players to tremble, but independently of each other. Ratifiability is then defined to be this concept without the independence condition, and this is equivalent to distributions that are modestly ratifiable *and* robust to perturbations. So by imposing independence of the 'trembles' and relaxing the robustness to perturbations (in either order) we obtain Nash equilibria from ratifiable distibutions. This concept of ratifiability would appear to be very close to the concept of acceptable correlated equilibria as developed by Myerson (1986) as they are both defined to be correlated equilibria that are robust to perturbations. However, all perfect equilibria are acceptable correlated equilibria, and Shin shows that equilibria can exist that are perfect, but do not satisfy the robustness assumptions required for ratifiability. The essential difference is that acceptable correlated equilibria are robust to any arbitrary perturbation, whereas ratifiability is defined relative to perturbations that are uniform across all possible strategies. Shin's work in this area is therefore helpful in clarifying the relationships between the different forms of equilibria.

Shin (1988b) gives a characterisation of the concept of common knowledge as used by Aumann (1976) in terms of a topology on a state space. This is done by using 'provability' as knowledge, i.e. an individual only knows something if he can prove it that it is true. This approach does not, however, permit individuals to have partitions over the state space (as this would require individuals to prove that they couldn't prove a statement) and relies on all individuals sharing a common state space. In game contexts (especially experimental games), a common state space would normally appear to be a reasonable assumption, as both players are usually aware of the moves available to all players, but this is not always the case (see especially Bennett, 1977, 1987). These results are not directly applicable to this thesis.

Mertens & Zamir (1985) fix upon Harsanyi's (1967) sequential expectation model which requires an infinite hierarchy of beiefs. Spaces over all possible types of player (the 'attributes') that are resticted so that at least one player believes that an element of the space is possible, are termed *Belief spaces.* It is shown that any of these belief spaces can be approximated by a finite space that is arbitrarily close to it on the Hausdorff metric. Therefore, any infinite regress associated with the beliefs of the players can be approximated by a finite number of levels of beliefs about the attributes of the players. This is in essentially the same vein to the approach taken in the previous chapter, where we reduced an infinite regress to a finite process by fixing beliefs about opponents' utility functions. It does, however, differ slightly because Mertens & Zamir consider the regress on the attributes of the players and then only determine Nash equilibria for the resulting game, whereas we consider the infinite regress on the beliefs of the players, thus determining optimal strategies.

Variations of repeated games called *stochastic games* are considered by Mertens & Neyman (1981). These are games whereby all players choose an action at every stage of the game, which determines the pay–off to each player, but these pay–offs can vary at different stages. Explicitly, the game can be in a number of states and the pay–offs depend on the state which the game is in. A referee determines the state, and the probability used by the referee to select the next stage depends upon the actions of the players, but is independent of past states. The players are informed at each stage what state the game is in, and then choose one of their available actions. Mertens & Neyman show that for a finite number of states and actions, all undiscounted games have a value. This implies that there exists a strategy that is ϵ–optimal (i.e. achieves an average pay–off within ϵ of the value of the game) in an infinitely repeated, or sufficiently long finite game. They also show that for games where there are infinitely many states and actions available, the same result holds, provided three simple conditions are met. A number of other authors (e.g. Shapley, 1953 or Shubik & Sobel, 1980) have considered stochastic games, and it can be seen that they are related to repeated games with incomplete information. However, in this thesis we shall only be concentrating on repeated games rather than stochastic games.

Another form of equilibria that has been considered in Bayesian game theory is that of *sequential equilibria.* This concept was first considered by Kreps & Wilson (1982). A sequential equilibrium is such that every decision made by a player must be part of an optimal strategy for the rest of the game. Therefore a player needs to determine his beliefs as to the situation that he is in, and to conjecture (predict) what will happen in the future (and as a result of his own next move). Sequential equilibria can be compared to, and are almost equal to Selten's perfect equilibria, but where perfect equilibria eliminate weakly dominated strategies, sequential equilibria do not.

A sequential equilibrium is defined to be a system of beliefs and a strategy, such that the beliefs are consistent (i.e. are in accordance with Bayes rule) and given these beliefs, the strategy is sequential rational (i.e. no player is able at any point to change his part of the strategy profitably). It is shown that for every extensive game, there is at least one sequential equilibrium. Also Kreps & Wilson show that sequential equilibria and perfect equilibria coincide at all perfect equilibrium points, except those that are not upper-hemicontinuous. By considering weaker sequential equilibria when player i has a small uncertainty about a player j's pay-offs, then these precisely equal the perfect equilibria. The main difference between Nash equilibria and sequential equilibria is that players' beliefs about events off the 'equilibrium path' can be used to determine optimal strategies in response to unspecified events. I feel that sequential equilibria play an important role in Bayesian game theory, especially when the games can be represented by their extensive form. In some games, sequential equilibria are of only limited use, like for example a repeated experimental game, and we shall not be using them here.

Smale (1980) considers repeated games where only some summary of the previous moves is used to determine a player's strategy. Solutions are in terms of undiscounted asymptotic solutions. For example, a 'good' strategy for a PDG when only the average pay-off is remembered from past stages, is shown to be one that defects when a player is being exploited, but cooperates more often than the opponent in order to encourage mutual cooperation. Smale shows that by introducing dynamics into the problem, a stable strategy can be found that is a uniquely optimal strategy, given the summary of the past stages. Also, any strategy that determines a Nash equilibrium under a given averaging system receives at least as high a pay-off as any Nash equilibrium strategy in the repeated game. It is useful that sufficient statistics can be found, such as average pay-off of the game to date. This idea is considered further in the next chapter of this thesis, and also in chapter 9. What might be useful is a guide as to which averaging or summary procedure is the most efficient for the game and/or beliefs about the other players. It is curious that Smale uses his results to determine only Nash equilibria for the players rather than any utility (or pay–off) maximising solution.

Blad (1986) obtains results equivalent to those of Smale, but avoids the assumption of bounded memory. This is achieved by, at every stage of the game, choosing two players at random to play the game. Therefore each player is assumed to play a pure strategy, and the dynamic link between the stages of the game relies on the evolution of the distribution over possible strategies. As new players are (probably) playing the game at each stage, no memory of earlier outcomes is required. Given a dynamic structure, solutions are determined as fixed points on a 2 simplex. 'Good' solutions that belong to a locally stable set are then found for a PDG. Blad then extends the model to permit mixed strategies. These are interesting results for determining solutions to a PDG, but when the conditions on the continual replacement of players at each stage will hold is not obvious. These results will not be used in this thesis.

This raises the question of how a Bayesian analysis of the type we are considering suggests rational players should play a PDG. Shubik (1970) presents differing attempts to resolve the PDG by three authors: Aumann, Howard and himself. Aumann's approach is that any outcome that achieves a pay-off higher than the mutual defection pay–off in an infinitely repeated game will be maintained as an equilibrium. This is close to the example 4.6.1 presented in the previous chapter. Shubik's approach is to consider 'sensible' and 'plausible' threats that are available to the players of such a game. However, this produces complex problems in its own right. In non–cooperative games, such threats can only be implicit, and will therefore suffer problems of communication. I agree that threats such as these can induce a sort of equilibrium in a repeated game.

The third approach (by Howard) is discussed in detail in section 4.3 of chapter 4 of this thesis. As we show there, a method based upon subjective beliefs and utility maximisation can be found as an extension of the Howard metaequilibria. I agree with Shubik that care must be taken when applying results from one context to another, especially when applying experimental game results.. However, I believe, unlike Shubik, that there are situations where Prisoner's Dilemma structures exist, and that the paradox *can* be solved by the use of subjective probabilities, and the notion of Bayes rationality (as we shall discuss in the next chapter).

Shubik (1981) again considers equilibria, and tries to define the properties that determine equilibria that are robust. He suggests that one should start with the set of Nash equilibria and then devise appropriate desireable properties. These desireable properties are classified into aesthetic properties, goals and limitations of the players, and communication structures. Various types of previously defined equilibria are then discussed. Shubik claims that the non-cooperative equilibrium solution is not a good candidate as a normative solution, but is more plausible as a behavioural solution. He then advocates a combination of the behaviouralistic and normative approaches as that which should be pursued in non-cooperative games. Shubik does not however indicate how optimal strategies should be determined under this approach.

Words of caution are offered to all of this by Kadane & Larkey (1982, 1983). In Kadane & Larkey (1982) it is argued that game theory must be considered in terms of maximising expected utility over the player's subjective probabilities of what the opponent will do on future stages of the game. Thus it follows that minimax solutions are not necessarily optimal for any game. All of this is in accordance with subjectivist Bayesian methodolgy, and I agree with it whole-heartedly. However, Kadane & Larkey do not insist upon the players making any rationality assumptions, as this will affect a player's subjective beliefs (which, it is supposed, have already taken such matters into account). Any solution concepts are considered to be simply a basis for a prior distribution.

Now this is fine as far as it goes, but Kadane & Larkey are making no use of how the opponent is viewing the player under consideration. It could be argued that this has already been taken into account in the subjective probabilities, but in this case, how has this been performed? We are in agreement with Harsanyi's (1982) comments on this paper. By simply stating that everything is already included in the subjective probabilities, Kadane & Larkey are reducing game theory to a simple maximisation problem. Rational behaviour is required to determine rational expectations of future play. How rationality is taken into account in subjective beliefs is an interesting and demanding pursuit, that must be performed in order to determine how people actually play, and given their beliefs, how they ought to play to maximise their expected utility.

Here also Kadane & Larkey have a warning. In Kadane & Larkey (1983) it is claimed that most of game theory has concentrated on how a player ought to play rather than how people actually play. They say that a proper understanding of the distinction is required, so models should be determined for a particular game setting, and then validated in accordance with their use. In a comment on this paper, Shubik (1983) agrees with this model specification and validation. But then he disagrees with Kadane & Larkey that a dichotomy actually exists between the subjectivist viewpoint that they are advocating, and normative game theory. He believes that both could be taken further to advance the theory of games. These views are in common with the work in this thesis. Shubik says that he is in favour of researchers determining a dynamic positive theory, but cautions that a complete and adequate static theory has not yet been determined.

Kadane & Larkey also suggest that a Bayesian perspective should be adopted, and then a player's understanding of his opponent's problem can be modelled and used to predict future play, with an updating of beliefs when information is received. This is essentially the approach that we have adopted here. A theory of how people ought to play the game (including all information about social norms and restrictions on play) is used to determine a player's beliefs about how he thinks his opponent will play, and therefore determine his own optimal move. Then, when observations about how the opponent actually plays the game are received, these can be incorporated into the player's subjective beliefs about future stages of the game. We shall continue this discussion in the next chapter.

Wilson (1986) uses the ideas of Kadane & Larkey (1982, 1983) by using subjective probabilities to model a player's beliefs about an opponent's play in order to determine an algorithm to calculate an optimal next move for the player. This algorithm is discussed and extended in chapter 7 of this thesis.

5.3 Conclusions.

It can be seen from this review that even the relatively specialised area of Bayesian game theory has a widely spread literature. Also, authors in this area have employed a large number of differing techniques, and differing sets of assumptions to determine how players should use any information or beliefs that they have about the game. Because of this diversity, the work presented in this thesis has more to do with the work of some of the above authors than others. My work here is mainly concerned with games of incomplete information, along the lines of Harsanyi (1967, 1968a, 1968b), which is then developed so that a player can determine his optimal play in the game, given various assumptions about his opponent. We shall go on to discuss a method of determining a precise optimal strategy, and show that a particular set of strategies is suboptimal for a player to adopt under given conditions. These will be discussed in the following chapters.

Our approach will concentrate on one particular player of these games (usually labelled P_1). We shall consider the game from his point of view, incorporating his beliefs and knowledge of the game, in order to determine *his* strategy. There have been objections that this approach can only have limited success (e.g. Terhune, 1974) as the game interactions are determined by the interdependence of the players. Other methods have been considered, such as the idea of an outside observer (or external observer) where all the players are assumed to be playing rationally as it would appear to an outside observer, i.e. given only the information common to all players. The approach that we are taking here can be defended by the fact that we are considering all a particular player's relevant knowledge and beliefs, and are incorporating this player's beliefs that his opponent is doing similarly. This leads to complex mathematical problems in terms of an infinite regress, but will determine an optimal strategy for the player. This is more in line with what Kadane & Larkey (1983) were arguing, because instead of determining the overall outcome of the game, we are considering one player and his beliefs about the game, which is more likely to be closer to what actual players of such games do.

Due to similar modelling assumptions, the work presented in this thesis is related to the work of Aumann and Harsanyi. However, it soon becomes apparent that there are differences between the approaches. The main difference is that Aumann and Harsanyi present the problem in terms of how it appears to an external observer. From this viewpoint, any decisions become states in the model, and the rationality of a player is judged by this third party, given the information available to him (i.e. that available to *all* players). Harsanyi achieves this by chance moves with a known distribution; Aumann by assuming the Harsanyi doctrine and all players are Bayes rational. Our approach considers how an actual player views the game. Rationality is then in terms of his beliefs and expectations. When rationality is assumed on the same information sets, and distributions over the players are assumed known, then the approaches must produce identical results, as the third party is fully informed. Our approach also facilitates the updating of beliefs and the possibility of a sequential decision rule, whereas Aumann and Harsanyi determine equilibria, that it is assumed will be repeated. This complicates our approach, but I feel that it makes it more adaptable, and more realistic.

Several authors have considered the possibility of the players being able to correlate their strategies. Optimal solutions can then be found in terms of a joint distribution over the possible outcomes of the game. Questions could be asked of the achievability of these joint distributions, as the development of such a distribution requires communication, which is not permitted in non-cooperative games. Communication can be developed in terms of the moves that the players make (like, for instance, the ability to communicate alternating strategies). However, in games with incomplete information, such communication (by moves) of more complicated joint distributions by a particular player, might be confused by the other players. Allied with this is the problem of the player's individual strategy. If the player simply plays the appropriate marginal distribution to achieve the intended joint distribution, then other players will not necessarily be able to determine the precise joint distribution if independence is not assumed. Therefore, without the presence of a 'deity', or similar, that is informing each player exactly what to do, there are problems with this approach.

There does, however, seem to be a strong link between the correlated equilibria considered by Aumann (1974, 1987) and the calibrated societies that we shall develop in the next chapter. In the case where utilities are known to be of the same form, and under the assumptions that all players are Bayes rational (and assume each other to be so), the calibrated societies must determine correlated equilibria. The two approaches have different formulations. Our approach is based upon the distribution that a player has over his opponent's utilities and actions, whereas Aumann's approach is based upon a known joint distribution determining how the players should play. The connection between the two concepts is not obvious when the utility functions are unknown, or little is known.

Our approach assumes that the players have been drawn from a particular population that they have some knowledge of (or at least beliefs about). A player can then calculate his optimal strategy, by taking expectations over this population to determine his beliefs about his opponent. This would appear to be reasonable if the players know something about their opponent (e.g. know that he too is a student), and symmetry can be used if little is known. When a player has received some actual information about his opponent, through plays of the game, these beliefs can be updated. This is in common with the type of argument used by many researchers in the area, following Harsanyi's (1967) notion of *attributes* of the players. Various limits on the behaviour of the opponents have been placed by some authors, in line with the assumptions placed upon the player under consideration, so as to achieve a particular concept or result.

In the following chapters we shall explicitly develop the ideas alluded to here.

6. REPEATED EXPERIMENTAL GAMES

6.1 Introduction.

How people play against each other in repeated games performed under experimental conditions has been extensively studied and recorded. However, there is still a large gulf between theoretical models, which are largely based on how players *should* play were they both "rational", and simple models constructed to fit the data from experiments.

Good players of certain types of experimental games can consistently achieve better results by choosing strategies which seem to be suboptimal in a game theoretic sense. In this chapter we shall construct Bayesian models of these games, guided by considerations of rationality and calibration. We argue that our model needs to correspond to a game of incomplete information where utilities are not necessarily linear in time. We will illustrate how our models can give insight into the success or otherwise of a species or group of players with different types of beliefs about each other.

One example of the dichotomy between game theory prescription and the results of experimental games is found in the study of the *Prisoner's Dilemma* game. Traditional game theoretic arguments dictate that both players in such a game should employ their maximin strategies and therefore *defect* at every stage of the game. However, players consistently achieve a higher pay–off than they would by employing these strategies. We will see how a model can be built which is both faithful to the observed common sense behaviour of the subjects of an experiment, and is also rational (in the Bayesian sense).

Another example illustrates how different behaviour can be explained by a player's utility structure relative to that of his opponent. We will see how a strategy can be considered optimal when a player's model of his opponent's responses is of a particular form, and he believes that his opponent is drawn at random from a population which has a given distribution over utilities.

In this chapter we use a modelling approach to game theory incorporating subjective beliefs, as advocated by Kadane & Larkey (1982, 1983). However, in contrast to these two authors, we stipulate that any realistic model of experimental games must have certain features that arise from game theory.

Firstly, I believe that the player should assume that his opponent is "rational". From a practical point of view, "rationality" can only be defined relative to players' beliefs about each others' play and about each others' utilities. This is most elegantly achieved by modelling these beliefs probabilistically. By not making this assumption we would be throwing away vital information and reducing the problem to a simple maximisation. Here we are in agreement with such authors as Harsanyi, Aumann and Mertens & Zamir.

Also, the model must correspond to a game of incomplete information. This is because it will hardly ever be appropriate to assume, for example, that an opponent's utility function is known precisely.

Finally, I believe that the model should not constrain us to make the unrealistic assumption that players' utilities are linear in time, as most models of repeated games do. That is, the players' utilities are not restricted to be equal to the (possibly discounted) sum of the pay–offs at each stage of the game. This is given explicitly in equation (6.3.1).

We shall first of all introduce some terminology and state our assumptions, and then set up the basic Bayesian model. We shall discuss some of the pitfalls that abound if one does not assume rationality on the part of one's opponent. Then we introduce the idea of mutual rationality through "Bayes calibration" and we indicate through some simple examples how "rational" models might be constructed to explain observed behaviour. We finally discuss some problems with the approach that we have adopted.

6.2 Assumptions and Notation.

A *repeated game* is one where each of 2 players, P_1 and P_2, plays a *move*, m_k and n_k respectively, on a sequence of *stages*, k, of the game ($k = 1, 2, \ldots$). The monetary pay–off to P_1 at stage k resulting from a move pair (m_k, n_k) is given by a pay–off matrix. For simplicity, throughout this chapter we will only consider games where the pay–offs at each stage of the game are symmetric for the players. Because of this symmetry, P_2's pay–off matrix is just the transpose of P_1's. We assume that both players have perfect information about the game being played, i.e. they both know the pay–off matrix for both players, all previous outcomes of the game, and also that only the information that is known by P_1 is known by P_2.

Let T denote the random variable after whose value t no further moves are played. Let $\mathbf{m}_k = (m_1, m_2, \ldots, m_k)$, $\mathbf{m}^{(k)} = (m_k, m_{k+1}, \ldots)$, $\mathbf{m} = \mathbf{m}^{(1)}$, and $m_k = 0$ if $k > t$. Similarly $\mathbf{n}_k = (n_1, n_2, \ldots, n_k)$, $\mathbf{n}^{(k)} = (n_k, n_{k+1}, \ldots)$, $\mathbf{n} = \mathbf{n}^{(1)}$, and $n_k = 0$ if $k > t$. We shall use the usual convention of capitalisation to denote the corresponding random variables. Let x_k

denote P_1's monetary pay–off at the kth stage of the game and $\mathbf{x} = (x_1, x_2, \dots)$. Note that, by the definition of t

$$x_k = 0 \qquad \text{for all } k > t. \tag{6.2.1}$$

We shall assume

(X1) x_k is a function of $(\mathbf{m}, \mathbf{n})$ only through $(\mathbf{m}_k, \mathbf{n}_k)$.

Definition. A *strategy* $\mathbf{d}$ for P_1 is a decision rule $\mathbf{d} = (d_1, d_2, \dots)$ where P_1's decision d_r for his rth move m_r is a (possibly randomized) function of only the moves to date $(\mathbf{m}_{r-1}, \mathbf{n}_{r-1})$ and any prior information that P_1 had before the game started.

Note that each player can *observe* directly his opponent's moves made so far but can only make *inferences* about his opponent's strategy and beliefs. This brings up the problem referred to in the game theory literature as the *infinite regress*, that we discussed in chapter 4. The notion of rationality through which the regress may be truncated, is central to the ideas in this chapter. Aumann's (1976) concept of *common knowledge* is also of importance here. This says that if the players have the same prior beliefs and that their updated beliefs are known by both players, and known to be known, etc., then these updated beliefs must be the same. This gives rise to assumptions such as (A2) given later.

Throughout this chapter, P_1 assumes that P_2 is playing a strategy. This implies that P_1 believes that P_2 is implicitly or explicitly using a decision rule d_k at time k which depends only on P_2's prior information before the game, and the past move sequence $(\mathbf{m}_{k-1}, \mathbf{n}_{k-1})$. Assuming P_2 plays a strategy enables P_1 to assert that P_2 can learn about P_1's next move only through the moves that P_1 has made so far. So, in particular, P_1's distribution over $N_k | \mathbf{m}_{k-1}, \mathbf{n}_{k-1}, \mathbf{d}$ is a function of $\mathbf{m}_{k-1}, \mathbf{n}_{k-1}$ only.

Throughout this chapter we shall assume that the game will terminate in finite time and T is independent of both P_1's and P_2's moves. Explicitly,

(T1) $T \perp\!\!\!\perp \mathbf{M} | \mathbf{d}$ for all strategies $\mathbf{d}$ for P_1, and $T \perp\!\!\!\perp \mathbf{M}$.

(T2) $T \perp\!\!\!\perp \mathbf{N}$.

A move m_k^* is said to *dominate* on the kth stage of the game if, for all possible moves n_k by P_2 and all other moves m_k by P_1,

(D1) $x_k(m_k^*, n_k) \geq x_k(m_k, n_k)$

and *strictly dominate* if

(D2) $x_k(m_k^*, n_k) > x_k(m_k, n_k)$.

In the PDG (see, for example, Figure 2.2.3 of chapter 2), move 2 strictly dominates as $C > A$ and $D > B$. Therefore the maximin strategy for the PDG is move 2, and for the repeated PDG is continual defection. This would seem to suggest that the optimal move for both players is move 2 at every stage of the repeated game, i.e. continual defection. However, as the game is clearly non–zero sum there is no reason to suppose that the maximin strategy is optimal in any broader sense, and as we saw in chapter 3, players of experimental PDGs often do better than if they had played *continual* defection.

So, game theoretic models of how people play these experimental games are not faithful to the way that players actually play them. We now consider the features of these models that are creating this discrepancy, in an attempt to find a set of models that are consistent with how people play these games.

6.3 Bayesian Rationality in Repeated Games.

Under the usual Bayesian definition of rationality, P_1 needs to choose a strategy, which henceforth we shall assume exists, that maximises his expected utility, this expectation being taken across his distribution over future relevant variables. In the context of experimental games, the experiment is usually designed so that these "relevant variables" are just those which will determine how P_2 will respond to P_1's chosen sequence of moves. Let $U(t,\mathbf{x})$ define P_1's utility on the outcome of a game terminated at time t, when he obtains a vector of pay–offs $\mathbf{x}$. Also at any stage of the game let Π_1 be P_1's distribution over P_2's future moves. Then, to fully specify our model we need to determine what constitutes sensible choices of $U(t,\mathbf{x})$ and Π_1. Firstly we discuss what constitutes reasonable choices by P_1 of $U(t,\mathbf{x})$.

In the past, usually for reasons of expedience, it has been usual to restrict attention to utility functions of a very specialised form. One form which regularly recurs in the economic and the psychological literature uses linear discounting in time — i.e. it sets

$$U(t,\mathbf{x}) = \sum_{i=1}^{\infty} \lambda^i U_i(x_i) \qquad (6.3.1)$$

where $0 < \lambda \leq 1$, and U_i is strictly increasing in x_i with $U_i(0) = 0$, $\quad i = 1, 2, \ldots$.

Although the mathematics becomes easier if a utility function of the form (6.3.1) is used, it can hardly be justified under criteria of "rationality" (see Luce & Raiffa, 1957). For instance, in the experimental games reported in the example of section 2.3 above, it would appear quite reasonable for P_1 to have a utility function U which was a function of the pay–off that he aggregated over the whole game. This is also true of all games where the players know that the game will only last a short length of time. So,

$$U(t,\mathbf{x}) = U_o\left(\sum_{i=1}^{\infty} x_i\right) \tag{6.3.2}$$

where U_o is increasing in its argument. After all, the game will usually only last at most a couple of hours. However, unless U_o is linear in pay–off, $U(t,\mathbf{x})$ cannot be written in the form given in equation (6.3.1), and it is known that a person's utility from a financial pay–off is rarely linear.

Also several other features may well be taken into account in a player's utility function. For example, a player might gain utility from achieving a higher pay–off than his opponent, or from gaining the highest pay–off in the group of players. Alternatively, a player may gain utility from appearing 'tough', and not conceding to a forgiving strategy after, for example, exploiting the opponent. Different possible features of utility functions are classified in Figure 3.3.2 above.

Because we want to encompass in our approach as wide a class of "rational" behaviour as possible we shall initially assume only:

(U1) $U(t,\mathbf{x})$ is increasing in each of its pay–off components x_i, or

(U2) $U(t,\mathbf{x})$ is strictly increasing in each of its pay–off components x_i, $\quad i = 1,2,\ldots$

Clearly (6.3.1) and (6.3.2) are special cases of (U1) and (U2). The specification of more explicit forms of U is left until later in the modelling process, and will be linked to the particular application of the game model in hand. Also incorporating other features than pay–off into the utility function will be left until later, as these are obviously context dependent.

Our first theorem relates the form of P_1's optimal strategy to the structure of the relationships within his specification of Π_1. We show that an optimal strategy for P_1 need only depend on P_1's "sufficient statistics" for P_2, in the sense that P_1's optimal kth move need only depend on s_k, defined below. We say $\mathbf{s} = (s_1, s_2, \ldots)$ is ***sufficient*** for N under P_1's model if P_1 assumes that:

(S1) P_2 makes his kth move in the light of the value of the (vector) function s_k of $(\mathbf{m}_{k-1}, \mathbf{n}_{k-1})$,

(S2) s_k is a function of $(\mathbf{m}_{k-1}, \mathbf{n}_{k-1})$ only through the value of $(s_{k-1}, m_{k-1}, n_{k-1})$

(S3) Given $(\mathbf{m}_k, \mathbf{n}_k)$ and $T = k+1$, $\quad U(k+1, \mathbf{x})$ is only a function of $(\mathbf{m}_k, \mathbf{n}_k)$ through s_{k+1}, where without loss of consistency with (S1) and (S2), we can define

$$s_k = s_t \qquad \text{if } k > t. \tag{6.3.3}$$

Several articles have appeared related to the Prisoner's Dilemma game which relate to models implicitly or explicitly using sufficient statistics for an opponent's strategy. For example, Smale (1980) uses relative aggregate pay–off to determine how well a player is doing relative to his opponent, and therefore if he is being optimal. On the other hand, Grofman & Pool (1975) use one–step ahead transition probabilities together with the current move pair in order to simplify the decision process. The utility maximising strategies from a given class of strategies can then be determined. We can conclude that when (S1), (S2) and (S3) hold in a model, the form of P_2's optimal strategy can be a fairly simple one.

LEMMA 6.3.1.

Suppose that the game terminates at time $k+p$, and assumptions (S1), (S2), (S3) and (T1) hold. Then, given the past move sequence $(\mathbf{m}_k, \mathbf{n}_k)$, the expected utility

$$\overline{U}(\mathbf{d}|\mathbf{m}_k, \mathbf{n}_k, T = k+p) \tag{6.3.4}$$

is a function of $(\mathbf{m}_k, \mathbf{n}_k)$ only through s_{k+1}, for each strategy $\mathbf{d}$ available to P_1. This is true for any $k \geq 1$ and for any $p \geq 1$.

PROOF: We shall prove this by induction on p. To prove for $p = 1$ note that

$$\overline{U}(\mathbf{d}|\mathbf{m}_k, \mathbf{n}_k, T = k+1) = \underset{M_{k+1}, N_{k+1}}{\mathsf{E}} [U(k+1, \mathbf{x}|\mathbf{d}, \mathbf{m}_k, \mathbf{n}_k)]$$

$$= \underset{M_{k+1}, N_{k+1}}{\mathsf{E}} [U(k+1, \mathbf{x}|\mathbf{d}, s_{k+1}, \mathbf{m}_k, \mathbf{n}_k)]$$

which, by (S1), (S3) and (T1)

$$= \underset{M_{k+1}, N_{k+1}}{\mathsf{E}} [U(k+1, \mathbf{x}|\mathbf{d}, s_{k+1})] \tag{6.3.5}$$

Clearly this expectation is a function of $\mathbf{d}$, $k+1$ and s_{k+1} only. So our assertion is certainly true for $p = 1$. Now assume the assertion is true for $p \geq 1$ so that, for all k and strategies $\mathbf{d}$,

$$\overline{U}(\mathbf{d}|\mathbf{m}_k, \mathbf{n}_k, T = k+p) = f(\mathbf{d}, s_{k+1}, k+p) \tag{6.3.6}$$

for some function f. So for any fixed d,

$$\overline{U}(\mathbf{d}|\mathbf{m}_{k+1},\mathbf{n}_{k+1},T=k+p+1)=f(\mathbf{d},s_{k+2},k+p+1)$$

which gives

$$\overline{U}(\mathbf{d}|\mathbf{m}_k,\mathbf{n}_k,T=k+p+1)=\mathop{\mathsf{E}}\limits_{M_{k+1},N_{k+1}}[f(\mathbf{d},S_{k+2},k+p+1)|\mathbf{d},\mathbf{m}_k,\mathbf{n}_k]$$

which by (S2) & (T1)

$$=\mathop{\mathsf{E}}\limits_{M_{k+1},N_{k+1}}[f(\mathbf{d},g(s_{k+1},M_{k+1},N_{k+1}),k+p+1)|\mathbf{d},\mathbf{m}_k,\mathbf{n}_k]$$

for some function g. This, by (S1)

$$=\mathop{\mathsf{E}}\limits_{M_{k+1},N_{k+1}}[f(\mathbf{d},g(s_{k+1},M_{k+1},N_{k+1}),k+p+1)|\mathbf{d},\mathbf{s}_{k+1}] \tag{6.3.7}$$

Clearly this expectation is a function of d, s_{k+1} and $k+p+1$ only. So if our assertion is true for p, then it is also true for $p+1$. The lemma is now proved. □

THEOREM 6.3.2.

Suppose an optimal strategy $\mathbf{d}^* = (d_1^*, d_2^*, \ldots)$ *exists for* P_1. *Then under assumptions (S1),(S2),(S3),(T1) and (T2), there exists an optimal strategy for which* d_{k+1}^* *is a function of past moves* $(\mathbf{m}_k,\mathbf{n}_k)$ *only through* s_{k+1}, $k = 1,2,\ldots$.

PROOF: It is sufficient to prove that for any strategy d, $\overline{U}(\mathbf{d}|T>k,\mathbf{m}_k,\mathbf{n}_k)$ is only a function of $(\mathbf{m}_k,\mathbf{n}_k)$ through s_{k+1}. Well

$$\overline{U}(\mathbf{d}|T>k,\mathbf{m}_k,\mathbf{n}_k)=\mathop{\mathsf{E}}\limits_{P}[\overline{U}(\mathbf{d}|T=k+P,\mathbf{m}_k,\mathbf{n}_k)] \tag{6.3.8}$$

where the random variable $P=(T|T>k)$ is independent of P_1's choice of d and his beliefs about N. Lemma 6.3.1 now gives us

$$\overline{U}(\mathbf{d}|T>k,\mathbf{m}_k,\mathbf{n}_k)=\mathop{\mathsf{E}}\limits_{P}[\overline{U}(\mathbf{d}|s_{k+1},T=k+P)] \tag{6.3.9}$$

which by (T1) and (T2) must be a function of $(\mathbf{m}_k,\mathbf{n}_k)$ only through s_{k+1}. P_1 can therefore choose an optimal decision d_{k+1}^* (if one exists), that need be a function of s_{k+1} only. The theorem is thus proved. □

This theorem implies that provided P_1's model of P_2's play satisfies the above assumptions, there exists a strategy of a relatively simple form which is optimal for P_1. We shall see later that this fact will enable us to identify optimal strategies in useful models of games. A relationship arising from this theorem between discrete dynamical systems and Bayes optimal play in a repeated game is given in Smith (1984).

Van der Wal (1981) shows that $\mathbf{d}^*$ can be determined by dynamic programming techniques for finitely repeated, two person, zero sum Markov games, i.e. games where the only permissible strategies are functions of the last move pair only. These games and respective strategies are obviously special cases of those considered by the theorem. However, his techniques are not appropriate for non–zero sum games such as PDGs, as equilibria are not necessarily unique and may be history dependent.

The next theorem characterises an optimal strategy in a repeated game with a dominating kth move m_k^*, $k = 1, 2, \ldots$, under the modelling assumption by P_1:

(F1) P_1's distribution of $\mathbf{N}^{(k)}|\mathbf{m}_k, \mathbf{n}_{k-1}$ does not depend upon $\mathbf{m}_k$ or $\mathbf{d}$.

There are two important situations in which (F1) is a good modelling assumption. The first is when P_1 believes that P_2 is essentially *unresponsive* to any moves P_1 might make. For example, this would be the case if P_1 was certain he was playing against an idiot.

The second situation arises when P_1 is completely *ignorant* about P_2's responses. His information is *so* vague that P_1 can learn nothing from P_2's past moves about P_2's future behaviour. For example, if P_1 sets equal probability to all of P_2's possible responses to each of P_1's possible sequences of past moves then assumption (F1) must hold. We now show that under (F1) and some of the regularity conditions given above, it is optimal for P_1 to play his dominating move at each stage of the game.

THEOREM 6.3.3.

If a dominating move m_k^* is available to P_1 at each stage, k, of the game ($k = 1, 2, \ldots$), then

(a) under assumptions (X1),(D1),(U1) and (F1) an optimal strategy for P_1 is to play the strategy $\mathbf{d}^* = (m_1^*, m_2^*, \ldots)$ of dominating moves. Furthermore,

(b) under assumptions (X1),(D2),(U2) and (F1) the strategy $\mathbf{d}^*$ is uniquely optimal, in the sense that if $\mathbf{d}'$ is also optimal for P_1 then $\mathbf{d}' = \mathbf{d}^*$ with (P_1's) probability one.

PROOF:

(a) Let $\mathbf{d} = (d_1, d_2, \ldots)$ be any (possibly randomised) strategy for P_1 and define

$$\mathbf{d}(k) = (d_1(k), d_2(k)\ldots) \qquad \text{where } d_r(k) = \begin{cases} d_r & \text{if } r < k \\ m_r^* & \text{if } r \geq k \end{cases} \tag{6.3.10}$$

Now

$$\begin{aligned} U(t,\mathbf{x}) &= U(x_1, x_2, \ldots, x_t), \text{ which by (X1)} \\ &= U[(m_1,n_1),(m_2(\mathbf{m}_1,\mathbf{n}_1),n_2(\mathbf{m}_1,\mathbf{n}_1)),\ldots,(m_k(\mathbf{m}_{k-1},\mathbf{n}_{k-1}),n_k(\mathbf{m}_{k-1},\mathbf{n}_{k-1})), \\ &\qquad \ldots,(m_t(\mathbf{m}_{t-1},\mathbf{n}_{t-1}),n_t(\mathbf{m}_{t-1},\mathbf{n}_{t-1}))] \\ &= U[(m_1,n_1),(m_2(\mathbf{m}_1,\mathbf{n}_1),n_2(\mathbf{n}_1)),\ldots,(m_k(\mathbf{m}_{k-1},\mathbf{n}_{k-1}),n_k(\mathbf{n}_{k-1})), \\ &\qquad \ldots,(m_t(\mathbf{m}_{t-1},\mathbf{n}_{t-1}),n_t(\mathbf{n}_{t-1}))] \quad \text{by (F1)} \end{aligned}$$

which by (X1),(D1) and (U1)

$$\leq U[(m_1,n_1),(m_2(\mathbf{m}_1,\mathbf{n}_1),n_2(\mathbf{n}_1)),\ldots(m_k^*,n_k(\mathbf{n}_{k-1})),\ldots,(m_t^*,n_t(\mathbf{n}_{t-1}))] \tag{6.3.11}$$

Thus from (F1) and the inequality (6.3.11), and taking expectations over the randomisation and $\mathbf{N}^{(k)}|\mathbf{m}_k, \mathbf{n}_{k-1}, \mathbf{d}$ gives

$$\overline{U}(\mathbf{d}|T = t, \mathbf{m}_{k-1}, \mathbf{n}_{k-1}) \leq \overline{U}(\mathbf{d}(k)|T = t, \mathbf{m}_{k-1}, \mathbf{n}_{k-1}) \tag{6.3.12}$$

where $\overline{U}(\mathbf{d}|T = t, \mathbf{m}_{k-1}, \mathbf{n}_{k-1})$ denotes the expected utility when P_1 uses decision rule $\mathbf{d}$ when he has observed $\mathbf{m}_{k-1}, \mathbf{n}_{k-1}$ and the game happens to end at *time* t.

Since by (T1) and (T2), T is independent of $\mathbf{d}$, we can conclude from (6.3.12) that

$$\overline{U}(\mathbf{d}|\mathbf{m}_{k-1}, \mathbf{n}_{k-1}) \leq \overline{U}(\mathbf{d}(k)|\mathbf{m}_{k-1}, \mathbf{n}_{k-1}) \tag{6.3.13}$$

So, whatever the past move sequence, it is optimal for P_1 to play his next move $d_k = m_k^*$. Consequently an optimal strategy for P_1 must be $\mathbf{d}^*$.

(b) If in addition the strict inequalities (D2) and (U2) exist, then (6.3.11) becomes a strict inequality. Inequality (6.3.12) will be strict when

$$k \leq t \tag{6.3.14}$$

$$\text{or} \quad \mathbf{d} \neq \mathbf{d}(k) \quad \text{with non–zero probability given } T = t \tag{6.3.15}$$

Inequality (6.3.13) will therefore be strict unless $d(k) = d$ with (P_1's) probability one. The same argument as before now tells us that a *unique* optimal strategy for P_1 must be d^*. □

In the context of experimental games, this theorem is sufficient to illustrate two points. Clearly, in practice, players playing games like the PDG do better than would be predicted by the modelling assumption (F1) (see, for example, Rapoport & Chammah, 1965). Therefore to be *realistic* it is necessary to assume:

(a) a degree of responsiveness on the part of your opponent; for example that he is, like you, intelligent,

(b) that players are not totally ignorant about each other's play, but respond (at least probabilistically) predictably to past patterns of play.

Once these assumptions are made it quickly becomes clear that reasonable models do exist which explain behaviour that has been observed in experimental games as rational behaviour.

In the next section we discuss models where such beliefs are held by P_1 without the additional assumption about the opponent's rationality. In section 6.5 we argue that good Bayesian models of experimental games will usually be based on the mutual belief of the rationality of one's opponent.

6.4 A Pragmatic Solution To Modelling Games.

A pragmatic solution to a repeated game is to use past information to determine how an opponent is likely to act in the future, in much the same way as we would in Bayesian models against nature. Kadane & Larkey (1983) suggest such an approach, which is used by Wilson (1986), and this is expanded upon in the next chapter.

In my opinion this solution has some merit when the ability of players to adapt the strategy they are employing is restricted in some way, such as in the practical settings given in the examples below. It does, however, have major difficulties if used to model experimental games. First we outline two problems where the pragmatic approach may work well.

6.4.1 An Advertising Model.

Here we give the simplest case of a Prisoner's Dilemma from marketing. Suppose 2 firms, X and Y, are in direct competition when selling a certain product to a specified market. Assume that only these firms produce such a product for this market and both products are equally desirable. Assume further that the total demand for this product is constant and at present a proportion ρ of the market that buys the product buys it from firm X, and proportion $(1-\rho)$ buys it from firm Y.

In order to increase sales, both firms consider at weekly intervals whether to advertise their product above the current level in the following week (decision d_2), or not (decision d_1). The pay–off matrix for firm X for each weekly decision is then given by:

$$\begin{array}{rc} & \text{Firm } Y \\ & \begin{array}{cc} d_1 & \quad d_2 \end{array} \\ \text{Firm } X \quad \begin{array}{c} d_1 \\ d_2 \end{array} & \begin{pmatrix} \rho & \rho-\delta \\ \rho+\delta-c & \rho-c \end{pmatrix} \end{array}$$

Figure 6.4.1

where $\delta \geq 0$ is the increase in the share of the market due to the increased advertising and $c \geq 0$ is the cost to the firm of the increased advertising. It is not unrealistic to assume that $c < \delta$ as otherwise the firm is unlikely to contemplate any increase in advertising. In this case we have the pay–off matrix of a Prisoner's Dilemma. The defecting move is to increase advertising in the following week, i.e. decision d_2, whereas the cooperative move is to leave advertising at its current level, decision d_1.

In such a Prisoner's Dilemma situation a pragmatic approach is feasible. For example, if it is *company policy* to react in a certain manner after an increased advertising campaign by the competitor, then it may be reasonable to assume that this policy will not change in the short term because of known constraints on the competitor's decision making processes. Also it is quite likely that it will not be possible to determine any future optimal strategy due to the complexity of the business and the number of external influences. Therefore a step–by–step approach to determine the optimal move, given the past history, seems plausible.

In this case, basing the prediction of the competitor's future behaviour on his past performance may provide a reasonable model. By assuming such consistency, a player may choose to model his opponent's behaviour by fixing a family of distributions for his own strategies without regard to whether, by playing these strategies, he would himself be acting rationally.

6.4.2 A Bidding Problem.

Another situation where it may well be reasonable to adopt a pragmatic approach is in a bidding problem. That is, when two firms, X and Y, are bidding for a job that has been put out to tender by a local authority or similar body. It can be seen that if we restrict the two firms to the unrealistic situation of only having two possible bids — high and low, then this problem can also be modelled by a PDG.

This is because a firm is always likely to obtain a higher utility from offering a lower price than offering a higher price, due to the fact that they are more likely to secure the job. On the other hand, the mutual high price outcome is preferred to the mutual low price outcome, thus defining a Prisoner's Dilemma.

Again, it may be known that the opponent will always react in the the same manner in a given situation for reasons of company policy etc. In this case, a pragmatic approach may well provide a good strategy that will use past bids made by the opponent, to forecast the next bid by the opponent, so that the highest bid that is likely to be able to secure the contract can be determined.

However, there are insidious problems with this approach when the opponent has the ability to change his policy quickly. If this is the case, then your belief that his past behaviour will determine future responses may well be fallacious. It is possible that he will suddenly start perceiving you correctly and react in a way inconsistent with his behaviour in the past, as soon as you do something provocative, like dramatically increase advertising, or make a very low bid. Of course, in experimental games the opponents' perceptions of each other can be observed to change rapidly especially early on in the game, so these problems are heightened, see Terhune (1974), Axelrod (1984), Harford & Soloman (1967).

In the approach of Grofman & Pool (1975), the opponent, P_2, is assumed to be playing a partial TFT strategy (i.e. mimicking P_1's previous move with fixed probability p) in order to determine the effectiveness of this class of strategies in eliciting a Cooperative move from P_1. These strategies were chosen because of their "simplicity and intuitive plausability as strategic choices in an iterated PD" (Grofman & Pool, 1975, p.191).

In the context of the experimental game where players are all drawn from the same population, a model of P_2's play of the type Grofman and Pool employ is very dubious in the light of Theorem 6.3.2. That such a model is poor is backed up by the findings of Axelrod (1980a, 1980b). In the computer competitions that he organised, the strategies that estimated the opponent's probability of cooperation from the most recent move pairs, and acted "optimally" given these probabilities, did not score very highly.

In Chapters 7 and 8 we shall consider games of the kind that Grofman & Pool (1975) considered. We will show that if P_1 believes P_2 to be playing a strategy from the class of partial TFT strategies, then the form of P_1's strategy can be determined. However, the optimal strategy for P_1 is never of the same form as the strategy that P_2 is assumed to be playing. We argue that this violates certain rationality criteria.

The type of argument presented in this section forces us to try to accommodate the idea into our model that our opponent is also a rational player. We show how to do this in the next section.

6.5 Rationality and Calibration in Symmetric Games.

It should be clear from the comments of section 6.4 that it is questionable, in an experimental game, whether any model that implies P_2 will play suboptimally will be much use in using any data set, D, to predict P_2's future behaviour. For example, if before the start of the game it should dawn on P_2 how to use his information to better achieve his objectives, then P_2 *will* play differently. We therefore choose to make the following assumption:

(A1) P_1 assumes that P_2 is Bayes rational.

That is, P_1's opponent, P_2, is maximising the expectation of his (P_2's) utility function U_2. Typically U_2 will be unknown to P_1. However in the context of this type of experimental game, we can assume that P_2's pay–offs at each stage of the game will be known to P_1.

Bayes rationality of the opponent is quite a common assumption to make (see for example, Harsanyi, 1967, Pearce, 1984 and Aumann, 1987) although it has been criticised (see Kadane & Larkey, 1982). This criticism is based on the distinction between 'subjective rationality' (rational given the beliefs of the player) and 'objective rationality' (rational given a third party's beliefs about the player). It is claimed that what is in fact subjectively rational is often taken not to be objectively rational.

However our emphasis here is unusual. With the exception of Harsanyi (1967) and Harsanyi & Selten (1972), most authors assume that P_1 knows P_2's utility on pay–off. This assumption we find very dubious in the context of experimental games where the identity of an opponent is typically kept secret from P_1. Indeed it is central to the ideas of this chapter that P_1 can be uncertain about at least *some* aspect of P_2's utility function. We therefore argue that P_1 should assume P_2 to be Bayes rational with respect to the utility function that P_1 believes P_2

to have, and then revise these beliefs as more information is received. This avoids the problem of the distinction mentioned above.

Unfortunately, as Aumann (1987) points out, Assumption (A1) is not strong enough to give much structure since we have as yet made no statements about P_2's beliefs about what P_1 will play. When the information available to players is symmetric, as in most experimental games, the next assumption is useful.

(A2) The distribution $\Pi(\mathbf{n}, U_2|\mathbf{d})$ that P_1 uses over P_2's strategies and utility is identical to the distribution that P_1 believes P_2 uses over P_1's strategies and utility.

Assumption (A2) is appropriate when one is modelling the types of experimental games that have been looked at by psychologists, when a selection of intelligent subjects (usually students) from a particular population play against each other. In general, suppose P_1 and P_2 have been drawn at random by an experimenter from a (possibly infinite) population G of players. Suppose further that a psychologist has previously made an extensive study of the game behaviour of players sharing the defining characteristics of G. He has found the distribution $\Pi(\mathbf{n}, U_2|\mathbf{d})$ over their strategies and utilities and all players in G have been informed of Π. Provided that experimental conditions ensure that neither P_1 or P_2 have any additional information about their opponent and both choose to believe the information that is given, then both players should use the distribution Π in their model.

So if all potential players have full probabilistic information about the group G but not the result of a randomisation which will choose the two players from G to play a game, each player should choose a strategy which maximises his expected utility under Π. In particular, if, as in well–constructed experimental games, all potential players have played many training games and had discussions, a reasonable modelling assumption (A2) is that all players' beliefs will converge to Π. For example, if the group of players has been drawn from a population of students, any player will know that his opponent is also a student and therefore is likely to have a very similar utility function to himself. Indeed, in such games it would seem to be reasonable to assume that your opponent believes about you exactly what you believe about him, unless you have any information to the contrary.

Note that (A2) is different from the Harsanyi Doctrine (Aumann, 1987). This states that an "objective" and agreed joint distribution exists between the two players about the prior values of any parameters. So this is saying that any differences in probability assessments by the players can be explained by different information that the players' have received. Assumption (A2) is stronger and is specifically designed to exploit the symmetry that exists in experimental games of the form described in this chapter.

Under Kadane & Larkey's (1982, 1983) methodology, all game theory is reduced to simply maximising the players expected utility with respect to the subjective probabilities that the player has over his opponent's future play. What we are arguing is that this is the correct way of tackling the problem, but the subjective probability distribution must be chosen under the assumption that P_2 is rational, and that P_2 assumes P_1 to be rational. Indeed, I contend that unless P_1 has information to the contrary, it would be dangerous not to make this assumption. Obviously, if P_1 does have information to the contrary, then this must be incorporated into the model. I believe that by making these rationality assumptions, and applying a Bayesian analysis, normative models can be built that are true to the way that people have been observed to play the types of experimental games that we are considering.

If P_1 satisfies the two assumptions (A1) and (A2), then the distribution $\Pi(\mathbf{n}, U_2|\mathbf{d})$ over P_2's responses and utility will be called a *Bayes–calibrated distribution*. So, if this is the case, P_1 believes P_2 to be Bayes rational and also believes that P_2 believes him (P_1) to be Bayes rational. After marginalising out U_2 from $\Pi(\mathbf{n}, U_2|\mathbf{d})$ we obtain $\Pi_1(\mathbf{n}|\mathbf{d})$ which is P_1's calibrated distribution over P_2's responses $\mathbf{n}$. Alternatively, by marginalising out $\mathbf{n}$ from $\Pi(\mathbf{n}, U_2|\mathbf{d})$, we obtain $\Pi_2(U_2)$, which is P_1's calibrated distribution over P_2's utility. It is clear from our assumptions that $\Pi_2(U_2)$ is independent of $\mathbf{d}$. We will discuss some aspects of these distributions and their derivation in the following examples.

6.5.1 Example 1 — Repeated games with known utilities.

In this example we shall make the assumption that within the set of players G, it is known that all players have the same utility function on their vector of pay–offs, and this utility function is any function that satisfies condition (U2). We shall also make assumptions (A1) and (A2). Our results are now analogous to those of Aumann (1987) and so outcomes can be seen to be correlated equilibria. Also, because of the calibration (symmetry) hypothesis, the results are comparable to those of Maynard–Smith (1982) on evolutionary stable strategies in evolutionary games. All P_1 now needs to do is to specify $\Pi_1(\mathbf{n}|\mathbf{d})$, his distribution of his opponent's responses $\mathbf{n}$ given that he (P_1) chooses a strategy $\mathbf{d}$. Assume $\mathbf{d}^*$, his utility maximising strategy, exists for any set of distributions $\{\Pi_1(\mathbf{n}|\mathbf{d})\}$ indexed by the strategy $\mathbf{d}$

he might use.

If $\mathbf{d}^*$ is unique, then as the group G is Bayes calibrated, all players in G are assumed to play $\mathbf{d}^*$. Therefore $\Pi_1(\mathbf{n}|\mathbf{d}^*)$ must be a degenerate distribution, which assigns probability one to P_2 also playing $\mathbf{d}^*$. On the other hand, if the set D^* of utility maximising strategies for P_1 contains more than one element, then for P_1's model to satisfy (A1) and (A2), $\Pi_1(\mathbf{n}|\mathbf{d})$ must assign probability one to P_2 also playing a strategy in the set D^*. Conversely if $\Pi_1(\mathbf{n}|\mathbf{d})$ does assign probability one to P_2 playing a strategy in the set D^*, then clearly $\Pi_1(\mathbf{n}|\mathbf{d})$ is a Bayes calibrated distribution for P_2's responses.

Note that, in a one–play zero–sum game with a maximin solution $\overline{\Pi}_1(\mathbf{n})$ and where utility is a linear function of pay–off for both players, the set $D^* = \{\mathbf{n} : \overline{\Pi}_1(\mathbf{n}) > 0\}$ gives a model satisfying $(A1)$ and $(A2)$ where for any strategy $\mathbf{d}^* \in D^*$,

$$\Pi_1(\mathbf{n}|\mathbf{d}^*) = \overline{\Pi}_1(\mathbf{n}) \tag{6.5.1}$$

since then

$$\overline{U}(\mathbf{d}_1) = \overline{U}(\mathbf{d}_2) \quad \text{for any} \quad \mathbf{d}_1, \mathbf{d}_2 \in D^*. \tag{6.5.2}$$

In this case D^* are called the "worthwhile strategies" (see Thomas, 1984). This result is, of course, consistent with P_1's choice of $\overline{\Pi}_1$, since P_2's utility must be identical to P_1's, and provides one vindication of maximin strategies in zero–sum games.

It is a well known fact that equilibrium strategies in a given repeated game are not in general unique, and so it follows directly that neither are calibrated distributions. The choice of an appropriate model therefore often needs to be governed by considerations external to game theoretic ones, and this will obviously need to be game specific. For example, consider the game which consists of repeated plays of the game E_1 whose pay–off matrix is given in Figure 6.5.1, when $a > b$. It is not difficult to check that any distribution which assigns probability one to an arbitary choice of strategy is Bayes calibrated under any utility of the form (U1). So at first sight it appears that no single Bayes calibrated distribution recommends itself. However by considering the specific game in hand, and provided P_1 is prepared to assume that his opponent is behaving in a reasonable way, it is apparent that P_2 will play move 1 at all stages of the game, in the belief that P_1 will play likewise. Thus a good model for P_1 is the model which has a calibrated distribution assigning probability one to P_2 playing move 1 at every stage of the game.

$$\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 1 & 2 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \end{array} \qquad\qquad \begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 1 & 2 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

Generating Game E_1 Generating Game E_2

Figure 6.5.1

Another compelling Bayes calibrated distribution that is based on common sense can be found for repeated plays of the game given by the pay–off matrix E_2 in Figure 6.5.1. Again, for any utility function of the form (U2), no single Bayes calibrated distribution is obviously optimal, and so it is not clear to P_1 how P_2 will play initially. However once a move pair (m, n) has been played which gives positive pay–off to both players there is no incentive for either to deviate from this move in the future. I contend that one very reasonable Bayes calibrated distribution for P_1 is to play move 1 with probability $\frac{1}{2}$ and move 2 with probability $\frac{1}{2}$ until a positive pay–off is obtained from a move pair (m, n), and then for P_1 to plan to play move m indefinitely, assuming that P_2 will likewise continue to play move n.

It is suggested that it is a game similar to E_2 that people appear to play regularly, when walking in a confined space towards someone else walking in the opposite direction. If both people become aware that they are on a collision course at the same time, then both are likely to take avoiding action, and therefore continue to be on a collision course. These moves will be repeated until one person moves and the other doesn't, and then the players are likely to stick to their current positions for the remainder of the game (i.e. until they pass). The alternative outcome is that a collision actually occurs, and this is assumed also to terminate the game. It is games like this that require us to have socially acceptable 'norms' or rules, to dictate how people should play games such as this in certain circumstances, e.g. for road traffic.

Returning to repeated plays of E_1, when $a = b$ it is not obvious what an appropriate model might be. One strong candidate is the distribution which assigns probability one to the strategy given above for E_2. However, if because of the structure of the experiment it is physically easier, or more psychologically compelling for each player to play move 1, then a better model might assign probability one to the opponent playing move 1 continually. Without this information about the environment in which the game is played it is impossible to state which of these models is more appropriate.

The point illustrated by these examples is that Bayesian models of a game invariably require an input of information about the environment in which the game is played. Just as for any good Bayesian model it is essential that the problem is not divorced from its context.

Now obviously the assumption commonly made by game theorists that players will have utilities on pay–off which are *known* to each other is extremely dubious in the context of these experimental games. It is essential that this assumption is relaxed if we are ever to approach a realistic model. In the two games E_1 and E_2, it is easy to check that if both players have a utility function satisfying condition (U2) but are *not* necessarily equal, then the same rationale justifies the "common sense" calibrated distributions given above.

Degenerate calibrated distributions can also be derived when dominant moves exist for each player. Suppose that both players of a symmetric repeated game, each have a utility function of the form given in (U1), and that a dominant move exists for each player at each stage of the game. Then, regardless of P_1's beliefs about P_2's utility function, a calibrated distribution for P_1 assigns probability one to P_2 continually playing his dominating move. This follows directly from Theorem 6.3.3.

Like in the examples above, there is no guarantee that a calibrated distribution is unique. When this is not the case, it is important to give a common sense rationale related to the context in which the game was played, before a good model for the game can be determined.

In the PDG defined by Figure 2.3.1 of section 2.3, Theorem 6.3.3 suggests that under certain conditions all players will continually defect. On the other hand, many authors (e.g. Maynard–Smith, 1982) have shown that if we know P_2's utility is linear in pay–off and the time of termination of the game, T, has a geometric distribution with probability of continuation greater than $\frac{1}{2}$, then a group where all players play TFT is also calibrated. Other calibrated move distributions over a different family of utilities are given below.

6.5.2 Example 2 — A Model for the Prisoner's Dilemma.

Consider the Prisoner's Dilemma game defined by the pay–off matrix given in Figure 6.5.2, and assume that all players have a utility on pay–off of the form

$$U_\theta(x) = \begin{cases} 0 \text{ if } x < \theta + 1 \\ 1 \text{ if } x \geq \theta + 1 \end{cases}, \text{ for } \theta = 1, 2, \ldots \tag{6.5.3}$$

$$\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 1 & 2 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix} \end{array}$$

Figure 6.5.2

Here each player's objective is to obtain at least £$(\theta+1)$. This might be because the players require £$(\theta+1)$ to purchase a particular item in short supply, or to recoup some initial charge. Suppose P_1 assumes that, with probability one, P_2 will choose a strategy s_ϕ of the form

$$s_\phi = \begin{cases} \text{TFT} & \text{before stage } \phi \\ \text{Defect} & \text{at and after stage } \phi \end{cases}, \quad \phi = 1, 2, \ldots \tag{6.5.4}$$

where an opponent's value of ϕ is unknown.

Now we make the claim that if $\Pi_1(\phi) = \Pi_2(\theta)$, then P_1 is Bayes calibrated.

To see this, first note that if the last move of the game occurs at time $t \leq \theta$, then regardless of the value of ϕ, any strategy for P_1 will give him zero utility. In this case all strategies are equally preferrable. Similarly if $\phi = 1$, P_1 cannot obtain positive utility. On the other hand, if $t > \theta$ and $\phi > 1$, then, given ϕ, $\mathbf{s}_{\phi-1}$ maximises P_1's pay–off which will then be ϕ with probability one. Hence, given $\phi > 1$, since U_θ is increasing in pay–off

$$\max_{\mathbf{s}_\phi} \overline{U}_\theta(\mathbf{s}_\phi|\phi) = \begin{cases} 0 & \text{if } \phi < \theta + 1 \\ \mathsf{P}(T > \theta) & \text{if } \phi \geq \theta + 1. \end{cases} \tag{6.5.5}$$

Now $\mathbf{s}_\theta$ also attains this maximum and so is an alternative Bayes strategy. Therefore $\mathbf{s}_\theta$ must be a Bayes strategy with expected utility

$$\overline{U}_\theta(\mathbf{s}_{\theta_1}) = \mathsf{P}(T > \theta_1)\Pi_1(\phi > \theta_1) \quad \text{when } \theta = \theta_1 \tag{6.5.6}$$

To ensure that the model is Bayes calibrated, we need only that the probability of $\theta = \theta_1$ agrees with the probability assigned to the optimal strategy $\mathbf{s}_{\theta_1}$. Hence all we need is that $\Pi_1(\phi) = \Pi_2(\theta)$ and the claim is proved. Call this model $\mathcal{M}_0$.

Of course other Bayes models exist under this utility function. Suppose, for a fixed value of k ($k = 1, 2, \ldots$) known to all players, everyone in a group G_k is known to play a strategy s^k_ϕ for some value of ϕ, where

$$s^k_\phi = \begin{cases} \text{Defect} & \text{before stage } k \\ \text{TFT} & \text{from stage } k \text{ to stage } \phi + k - 2 \\ \text{Defect} & \text{from stage } \phi + k - 1 \text{ to the end of the game.} \end{cases} \quad , \phi = 1, 2, \ldots \tag{6.5.7}$$

Here again ϕ is a parameter unknown to the other players. Call this model $\mathcal{M}_k$. Using exactly the same arguments it is easily checked that setting $\Pi_1(\phi) = \Pi_2(\theta)$ gives a Bayes calibrated distribution. In this case,

$$\overline{U}_\theta(s^k_{\theta_1}) = P(T > \theta_1 + k).\Pi_1(\phi > \theta_1) \tag{6.5.8}$$

Referring to the actual play of Pair A given in section 2.3 above, we can see that it is possible to explain P_1's behaviour in the following way. Both players treat the first two moves as a training period and each writes–off their gains and losses over this period (i.e. their utility functions are both constant in the first two arguments of pay–off). Thereafter they both believe their opponent is playing one of the class of strategies given above; here P_1 has a value of $\theta = 32$ and P_2 has some value of $\theta \geq 33$. In the actual game, P_1 achieves his objectives whereas P_2 does not — although he would have done had he been playing some other members of this group.

Although this model is unrealistically simple it does illustrate that, unlike classical game theory models, Bayes calibrated models exist consistent with the experimental behaviour we actually *observe* in successful players. So these models can be both descriptive and prescriptive. A similar type of model exists for the observed play corresponding to the game E_2 given in Figure 6.5.1.

Thus we have seen that a small adaptation of $\mathcal{M}_0$ can determine quite a realistic model of how people play and what objectives people have in experimental PDGs. Note that by using utilities of the form (U1) we can obtain an optimal strategy which is not continual defection *even* when the termination time of the game is known to all players. This would not be possible if we were to employ conventional ideas of game theory, such as extended rationality (see, for example, Hamburger, 1979).

An unusual feature of this example is that P_1's optimal strategy does not depend on his beliefs about the parameter θ in P_2's utility function. Usually $\Pi_2(\theta)$ will have a significant effect on P_1's choice of calibrated strategy as is illustrated in the next example.

6.5.3 Example 3.

Consider the one–play game given by the pay–off matrix in Figure 6.5.3,

$$\begin{array}{cc} & P_2 \\ & \begin{array}{cccc} & 1 & 2 & 3 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{pmatrix} 0 & 1 & -2 \\ -1 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix} \end{array}$$

Figure 6.5.3

and suppose P_1 believes that his opponent, P_2, has been drawn at random from a group G, and P_2 has a utility function $U_2(x)$ on his (P_2's) pay-off x of the form

$$U_2(x) = \begin{cases} x & \text{if } x \geq -1 \\ \theta(x+1) - 1 & \text{if } x < -1 \end{cases} \tag{6.5.9}$$

where $\theta > 0$ is unknown to P_1. Suppose further that P_1 has a utility function, U, of this form as well, with parameter θ_1. P_1 decides to use a family of distributions over his opponent's possible three moves, of the form

$$\Pi_1 = k(\phi)(3+\phi, 5+3\phi, 4) \qquad \text{for } \phi > 0 \tag{6.5.10}$$

where $k(\phi) = [4(3+\phi)]^{-1}$. Then it is easy to check that

$$(\overline{U}(1), \overline{U}(2), \overline{U}(3)) = k(\phi)(1+3\phi-4\theta_1, 1-\phi, 1-\phi). \tag{6.5.11}$$

We can conclude that if $\theta_1 < \phi$, P_1 should play move 1, and if $\theta_1 \geq \phi$ he should be indifferent between the strategy that continually plays move 2 and the strategy that continually plays move 3.

This is consistent with Π being Bayes calibrated if and only if ϕ is chosen to be a lower quartile of P_1's distribution over the parameter θ of $U_2(\theta)$. For then P_1 believes that $\frac{1}{4}$ ($= k(\phi)(3+\phi)$) of the players in G will play move 1, as stated in Π_1. So P_1 can use his beliefs about P_2's utility function to frame his beliefs on how he should play and what will happen in the game. These beliefs will be updated as the game progresses, and therefore as P_1 receives more and more information.

We now turn to the question—What characterises a Bayes calibrated society? The next result is useful in this regard.

THEOREM 6.5.1.

Suppose P_1 believes that P_2 will choose a strategy from a set $\{s_\phi : \phi \in \Phi\}$, where the set Φ is countable and the index ϕ has mass function $\Pi_1(\phi)$ such that $\Pi_1(\phi) > 0$ for all $\phi \in \Phi$. Let P_1 believe that everyone in the group G will have a utility function of the form $\mathbf{U} = \{U_\theta : \theta \in \Theta\}$. Then $\Pi_1(\phi)$ is a Bayes calibrated distribution over P_2's responses if and only if, for each value of $\phi \in \Phi$, there exists a value of θ such that s_ϕ is a Bayes strategy for P_1 under U_θ and $\Pi_1(\phi)$.

PROOF: If there exists a ϕ with $\Pi_1(\phi) > 0$ and no θ such that s_ϕ is optimal under the mass function $\Pi_1(\phi)$, then by (A2) those players playing ϕ must be acting suboptimally, in contradiction to (A1).

Conversely, if the condition given in the statement of the theorem holds, it follows that a mass function over the parameters of the utility function which is consistent with Π_1 sets

$$\mathsf{P}(\theta) = \sum_{\phi \in I(\theta)} \Pi_1(\phi) \tag{6.5.12}$$

where

$$I(\theta) \subseteq J(\theta),$$
$$J(\theta) = \{\phi \in \Phi : s_\phi \text{ is a Bayes strategy under } U_\theta \text{ and } \Pi_1(\phi)\}$$

and $I(\theta)$ are chosen to partition $\bigcup_{\theta \in \Theta} J(\theta)$. □

Analogous results exist when ϕ has a continuous distribution, although this is technically more difficult.

This theorem makes it fairly clear that if $\mathbf{U}$ is chosen after $\Pi_1(\phi)$, then $\mathbf{U}$ can be chosen to vindicate any choice of $\Pi_1(\phi)$. For example, this is the case if $\mathbf{U}$ is constant for all players with probability one, and all strategies are equally preferred by all players. Of course, this does not mean that any choice of $\Pi_1(\phi)$ is realistic, as P_1 may believe that some distributions over realistic choices of $\mathbf{U}$ would not correspond to $\Pi_1(\phi)$. On the other hand, when there are data which suggest a particular family for $\Pi_1(\phi)$, we will usually be able to explain this observed behaviour in terms of a Bayes calibrated model. The consequent inferences we can then make about the form of Π_2 are helpful in future play. However, all such restrictions that P_1 might place on the distribution $\Pi_1(\phi)$, are determined by external factors such as the social situation that the game is being played under and P_1's own psychological make–up, as opposed to any factors intrinsic to the game.

All this is in stark contrast to the situation where P_1 makes the unrealistic assumption that he knows his opponent's utility on pay–off. It also answers Kadane & Larkey (1982) in their criticism of models assuming rationality. Once utilities are assumed unknown *any* behaviour can be explained as rational.

Game theorists have tended to concentrate on modelling situations where P_1 has extremely weak information about P_2's patterns of responses, but is confident about the class of P_2's possible utility functions. Now it may well be appropriate for P_1 to construct his Bayes calibrated distribution (called *consistent* in a more general setting by Harsanyi, 1968b) with reference to a fixed distribution $\Pi_2(\phi)$ over the index θ of the chosen set of utilities $\mathbf{U}$. If this is the case, then finding the corresponding candidates for calibrated distributions over P_2's responses is more difficult, as is determining whether they do in fact exist.

Define τ to be a function which maps the mass function $\Pi_1(\phi)$ onto a mass function of the corresponding Bayes strategies under $U(\theta)$, where θ has mass function $\Pi_2(\phi)$. Then a Bayes calibrated distribution is just one for which Π_1 is a fixed point of τ. Example 4 provides a sketch of how a fixed point theorem might be used to find Bayes calibrated distributions.

6.5.4 Example 4.

Let $\mathbf{S} = (S_1, S_2, \ldots, S_i, \ldots)$ be sufficient statistics in P_1's model and suppose that the number, σ_i, of values that S_i can take satisfies

$$\sigma_i \leq c \qquad i = 1,2,3,\ldots \tag{6.5.14}$$

for some constant c. Then let Σ_k be the number of moves open to P_2 on the kth stage of the game, and assume that

$$\Sigma_k \leq A \qquad k = 1,2,3,\ldots \tag{6.5.15}$$

Then if the game is known to finish before time T^*, then the number of possible distinct strategies open to P_2, B, satisfies

$$B \leq A^{cT^*} \tag{6.5.16}$$

Now let the distribution Π_1 over P_2's responses $\mathbf{n}$ and utility function U_2 satisfy condition (U1), and also be consistent with $\mathbf{S}$ being sufficient. Then by Theorem 6.3.2, a Bayes decision exists which is one of the B strategies open to P_2 for any distribution Π_1. As B is finite, and provided τ is continuous, Brouwer's Theorem (Parthasarathy and Raghavan, 1971, p.27) states that a fixed point of r exists. It follows from this that a calibrated distribution consistent with Π_2 also exists.

Fixed points have been calculated for simple models (e.g. by Pearce, 1984). The problems that arise with this approach do not seem to lie in the *existence* of fixed points, but their *multiplicity*.

6.6 Diagnostics, Robustness and Beyond.

6.6.1 Diagnostics.

Because in practice, an initial choice of calibrated model $\mathcal{M}$ may be inappropriate, a player can use diagnostic tools to his advantage during the play of a repeated game. Indeed I believe that a player should always use such tools. Under the model $\mathcal{M}$, an opponent's responses will have a well–specified distribution, and his *observed* responses can be checked against them. This is in line with the recommendation in Kadane & Larkey (1983), that models should always be tested in accordance with their purpose.

In the Prisoner's Dilemma game an extreme case is the Bayes calibrated distribution which assigns probability one to all players continually defecting. Obviously, if, at any stage in repeated game P_2 cooperates, then P_1 must reject this model. On the other hand, suppose P_1's utility is of the form (6.5.3) of Example 6.5.2, and that the alternative model, named $\mathcal{M}_k$ in that example, is the one that P_1 believes to be correct. Then P_1 will never be able to diagnose his mis–specification, no matter how long the repeated game continues (or, for that matter, how many repeated games against other opponents he might play).

Game models are unusual in that the choice of a model influences the number of possible responses by the opponent, and hence affects the power of any diagnostic technique that might be used. For this reason, it might be desirable for a player to choose a Bayesian model whose optimal strategy encourages a variety of responses, rather than one specific response. In particular, if there exists an optimal strategy $\mathbf{s}^*$ which randomises over all of P_1's possible moves at each stage of the game, then this can be very helpful. For, in the long run, P_1 will observe P_2's responses to, for example, all finite sequences of moves he makes. He will therefore be able to check the appropriateness of his model against a very wide range of alternatives using the usual sorts of Bayesian methods (see, for example, Dawid, 1982, Smith, 1985, West, 1986).

6.6.2 The stability of calibrated models and the efficacy of A1 & A2.

We argued in the previous section that in the context of our models of incomplete (utility) information, assumption (A1) is essentially vacuous. On the other hand, the argument in section 6.5 justifying assumption (A2) may well not be realistic, because there may be a lack of symmetry of information amongst the players. This can be handled theoretically (see e.g. Harsanyi, 1967), but from a practical point of view the problem quickly becomes intractable as the parameters reflecting P_1's beliefs about different opponents (the "attributes") multiply. Such games can be modelled by hypergame models of the kind developed by Bennett (1977) and which are discussed in point 1 of chapter 9 of this thesis.

6.6.3 Dominated societies and stability for rational games.

Here we have argued that "good play" in a game can only be defined with reference to a group G of players, with a distribution over strategies. On the other hand, it is possible to compare the success of different groups of players. So, we shall say that group G_1 *dominates* group G_2 (written $G_1 \succ G_2$) given a distribution $\Pi_2(\boldsymbol{\theta})$ over a family of utility functions U indexed by $\boldsymbol{\theta}$, if and only if with probability one,

$$\mathsf{E}[U_{\boldsymbol{\theta}}|\Pi_1^1(\phi)] \geq \mathsf{E}[U_{\boldsymbol{\theta}}|\Pi_1^2(\phi)] \quad \text{for all } \boldsymbol{\theta} \tag{6.6.1}$$

with strict inequality for some $\boldsymbol{\theta} \in A$, where $\Pi_2(A) > 0$ and Π_1^i is the calibrated distribution of responses in G_i, $\quad i = 1, 2$.

For example, in Example 6.5.2, if G_i corresponds to the model $\mathcal{M}_i$, $\quad i = 0, 1, 2$, then by comparing equations (6.5.6) and (6.5.8) we see that $G_{i-1} \succ G_i$, $i = 1, 2, \ldots$, provided that $P(T = t) > 0, \quad t = 1, 2, 3, \ldots$.

So, whatever a player's utility function, it is at least as advantageous for him to be in group G_{i-1} as it is to be in group G_i. In a rational environment, groups G which are strongly dominated are unstable, in the sense that there is an incentive for *all* players to migrate from G to the dominating group. This migration may be physical, or may be achieved by passing laws that force members to act as if they were in the dominating group. In the example above, groups G_i, $i > 1$, are all *rationally unstable*, and so all rational players will be expected to leave these groups. Note in particular that a continual defection society in a PDG is unstable in this sense, as a society playing TFT will dominate it. However, this continual defection society is evolutionarily stable (Maynard-Smith, 1982) — i.e. it is stable in an irrational sense.

6.7 Conclusions.

By choosing a realistic definition of rationality appropriate to the actual experiments on, for example, repeated PDGs, we have shown that the apparent conflict between results of game theory and of experimental games can evaporate. Observed behaviour can be rational in a game theoretic sense, given a realistic class of utilities for the subjects and given a class of appropriate models of the behavioural relationships between subjects in the game. Having made this link it is now possible to use game theory not only to understand how people *should* play when completely ignorant about their opponents, but also to help make inferences about the relationships of perceptions of players, given that they have been observed to act in a certain way. In the latter activity I believe that game theory has a much stronger role to play in the social sciences than it has taken in the past.

I do not believe that this is adding to the confusion between "is" and "ought" in game theory that Kadane & Larkey (1983) refer to. Our theory is normative in the sense that we are determining how players should play these experimental games, but we are making the assumption that, until we have anything to tell us otherwise, the opponent is rational. I claim that this provides a basis for a good model of how people actually play these games, and can therefore be used as a positive theory. Obviously this must be used in conjunction with the player's subjective beliefs about his opponent to determine a first approximation to how the opponent will play, and this will be validated and revised as the game progresses.

Care must also be taken when applying the results from experimental games to real world applications. In most real world settings, there are far too many complex variables and interrelationships for any direct conclusions to be made. Therefore it is necessary to consider simpler experimental games, but it must be recognised that it would be incorrect to simply apply the results of these games to any real application, without thinking about it. As I have argued, solutions to such games under Bayesian models must be context dependent.

Some of the work in this chapter has previously been reported in Smith & Young (1987).

7. DERIVING OPTIMAL STRATEGIES

7.1 Introduction.

In previous chapters we have considered how Bayesian models of games can be constructed and some of the problems associated with this. In this chapter we shall continue to adopt the approach where the information flow is measured probabilistically, but with a view to determining a specific optimal strategy for a particular game. The game that we are concentrating on here is the PDG, but there is no reason why the approach could not be extended to any particular game. We have previously discussed various solution concepts, but in this chapter we are not concerned with these, as we wish to find the precise strategy that a player should adopt in a repeated game, given his prior beliefs about his opponent, and how his opponent has played in the past.

This pragmatic methodology is in line with that proposed by Kadane & Larkey (1982, 1983), which is a modelling approach to game theory, in an attempt to make game theory more practical. Wilson (1986) uses Kadane & Larkey's approach on the PDG and, by employing a form of backwards induction, provides an algorithm for determining optimal play. As in all Bayesian approaches, this algorithm allows the decision maker (P_1) to incorporate his subjective probabilities to determine how he expects his opponent (P_2) will play, and thus help to determine his (P_1's) optimal strategy. Wilson claimed that this Bayesian approach was "an intuitively attractive and viable alternative to more traditional methods" of solving decision making problems under uncertainty.

Wilson's algorithm is a good way of calculating an optimal next move at any stage of the game. However, there are at least two drawbacks to his approach to modelling repeated games. Firstly, Wilson's method only determines P_1's best next move. It cannot determine the *form* of the strategy for P_1 to adopt, because under Wilson's assumptions, he would then need to determine an infinite set of parameters. Thus Wilson's algorithm gives little insight into *why* the calculated solution is optimal.

We will show how to calculate the algebraic form of an optimal strategy in a particular Bayesian model, and thus discover the functions of past play that this strategy depends upon. Although not giving explicit instructions to P_1 about what to do on his next move, the optimal form of solution does give us an insight into how the Bayesian paradigm is determining P_1's moves. In particular, it shows how P_1 should plan his play over the whole of the remainder of the game. Through the discussion of a simple game, we illustrate how the form of an optimal solution can be combined with Wilson's numerical algorithm to obtain a graph from which P_1's optimal subsequent moves can be read as a function of the moves to date. This presents P_1's optimal strategy in a much more intuitively appealing and informative form. We also discuss how Wilson's algorithm can be speeded up by dramatically improving upper and lower bounds used in his calculations.

We use Wilson's example to illustrate the ideas of how his algorithm can be combined with the derived form of the solution to give an improved understanding of the implications of P_1's chosen model of P_2's play. Our method will apply to more complicated models and will be a helpful addition, provided there are assumptions regarding the probabilistic structure implicit in P_1's choice of model.

The second drawback of Wilson's approach is that the criterion by which P_1 chooses a model of P_2's play makes no reference to P_2's rationality. Rationality is a cornerstone of classical game theory and I believe mutual rationality, discussed in Harsanyi (1977), Aumann (1987) and also in the previous chapter of this thesis, is a vital ingredient of most sensible Bayesian models of games. In the example in his paper, Wilson makes the assumption that P_2 is playing a partial tit–for–tat strategy. By obtaining a greater understanding of the form of the optimal strategy, we are able to question the validity of the assumption made in the example used by Wilson, through addressing the implied lack of rationality.

7.2 Wilson's Algorithm.

First of all we define the problem that Wilson considered and state his assumptions. Consider the infinitely repeated game where at every stage, each of two players P_1 and P_2, can choose between two moves: C and D. The pay–off matrix that defines all stages of this game is given in Figure 7.2.1.

$$\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} C & \;D \end{array} \\ P_1 & \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1 & -\frac{1}{2} \\ 2 & 0 \end{pmatrix} \end{array}$$

Figure 7.2.1

As this game is symmetric, we have only given P_1's pay–offs in the pay–off matrix. The pay–offs to P_2 are simply the transpose of this matrix. We have divided the pay–off matrix that Wilson used in his example by 10. Because of the form of P_1's utility function, this does not affect the analysis, but makes the algebra neater. We obtain the same results, but for comparison, some of the results will obviously need multiplying by 10.

P_1's utility function is assumed to be the sum of discounted pay–offs, with discount factor λ. We refer to this as being a discounted linear utility function. Thus, P_1 is aiming to maximise

$$\sum_{k=0}^{\infty} \lambda^k x_k \tag{7.2.1}$$

where x_k is P_1's expected pay–off at stage k. In previous chapters we have argued that the use of utility functions of this form is not that compelling, and possibly more appropriate utility functions are discussed in the previous chapter of this thesis.

We shall assume that P_1 believes that P_2 is playing a partial tit–for–tat strategy with parameter p. This means that if P_1 played move C on the last stage of the game, then P_2's next move is expected to be

$$\begin{cases} C & \text{with probability } p \\ D & \text{with probability } (1-p) \end{cases}$$

and if P_1 played move D on the last stage of the game, then P_2's next move is expected to be

$$\begin{cases} C & \text{with probability } (1-p) \\ D & \text{with probability } p. \end{cases}$$

P_1 can thus express his beliefs about his opponent's next move via a distribution over p, which is continually updated by Bayes' rule as the game is repeated.

Also we shall assume that P_1's prior distribution over p, the probability that P_2 will mimic P_1's last move, is Beta with parameters $\alpha(0)$ and $\beta(0)$. Let $f(p)$ denote the density of this distribution. Explicitly,

$$f(p|\alpha(0),\beta(0)) = \begin{cases} \frac{1}{\mathrm{Be}(\alpha(0),\beta(0))} p^{\alpha(0)-1}(1-p)^{\beta(0)-1} & \text{if } 0 \le p \le 1 \\ 0 & \text{otherwise.} \end{cases} \tag{7.2.2}$$

where $\alpha(0),\beta(0) > 0$ and $\mathrm{Be}(\alpha(0),\beta(0)) = \int_0^1 p^{\alpha(0)-1}(1-p)^{\beta(0)-1}\,dp$. After observing each move pair the distribution is updated by Bayes' rule to another beta distribution, with parameters $\alpha(t)$ and $\beta(t)$ after stage t, where

$$\alpha(t) = \alpha(0) + s(t),$$

$$\beta(t) = \beta(0) - s(t) + t,$$

$$s(t) = \text{number of times } P_2 \text{ has mimicked } P_1 \text{ on the first } t \text{ moves.} \tag{7.2.3}$$

We define μ to be the current mean of P_1's distribution over p, so $\mu(t) = \frac{\alpha(t)}{\alpha(t)+\beta(t)}$. The parameter $\tau(t) = \alpha(t) + \beta(t)$ measures the number of observations included in the information. Larger values of $\tau(t)$ imply greater certainty in the expected value $\mu(t)$. Let $\phi(t)$ denote the state of the game and this will take the value C or D, depending on whether P_1 Cooperated (made move C) or Defected (made move D) at the previous stage of the game. Wilson considers the case when p has a non-degenerate Beta distribution, so that the parameter τ is finite.

Given P_1's Beta distribution, at any stage of the game it is simple to calculate these three parameters. Once these have been determined we can quite easily work out the expected utility to P_1 that would be obtained from any future move sequence (e.g. $(C, C, C, \dots)$, $(D, D, D, \dots)$, $(C, D, C, D, \dots)$, etc.).

Essentially, for any given fixed Beta prior distribution on p, Wilson's method calculates, for a given value of n, the maximum expected utility (defining utility to be discounted pay-off) from the next n moves starting with a Cooperation ($f_n(1)$), and the maximum expected utility from the next n moves starting with Defection ($f_n(2)$). By this we mean that $f_n(1)$ indicates the maximum expected utility from any sequence of n moves for P_1 where the first of these moves is Cooperation, and $f_n(2)$ is similarly defined for move sequences where the first move is Defection.

This difference in maximum expected utilities is calculated (using backwards induction) for increasing values of n until either

$$f_{n+1}(1) - f_{n+1}(2) \geq 2M\lambda^{n+2}(1-\lambda)^{-1}$$
$$\text{or } f_{n+1}(1) - f_{n+1}(2) \leq -2M\lambda^{n+2}(1-\lambda)^{-1} \tag{7.2.4}$$

where M is such that all entries in the pay-off matrix lie strictly in the range $[-M, M]$. In the game under consideration we can obviously take $M = 2$.

Therefore the maximum possible utility from all future stages (starting from stage $n+1$) is

$$M\lambda^{n+2} + M\lambda^{n+3} + M\lambda^{n+4} + \cdots = M\lambda^{n+2}(1-\lambda)^{-1} \tag{7.2.5}$$

So, we set S to be this maximum,

$$S = M\lambda^{n+2}(1-\lambda)^{-1} \tag{7.2.6}$$

and L to be the minimum utility that a move sequence could possibly obtain on all future stages of the game,

$$L = -M\lambda^{n+2}(1-\lambda)^{-1}. \tag{7.2.7}$$

The algorithm continues until the modulus of the difference between $f_{n+1}(1)$ and $f_{n+1}(2)$ is greater than the maximum possible difference in utility between any two strategies over all subsequent stages of the game, i.e. $|S - L|$, and therefore the smaller value cannot possibly 'catch–up' with the larger value. The optimal next move is then Cooperation if $f_{n+1}(1) > f_{n+1}(2)$ and Defection if $f_{n+1}(2) > f_{n+1}(1)$. This algorithm can be used for general 2 player games, but is illustrated by the Prisoner's Dilemma game.

7.3 Form of the Optimal Solution.

As stated above, at each stage of the game k, we have values of the parameters $\mu(k)$, $\tau(k)$ and $\phi(k)$. The results that we prove in this section will hold for all PDGs. For this reason we will consider the general PDG matrix given in Figure 7.3.1. In this matrix the variables b and c are strictly positive, and $b + c > 1$ for the matrix to define a PDG. Note that we obtain the matrix in Figure 7.2.1 by putting $b = 1$ and $c = \frac{1}{2}$.

$$\begin{array}{cc} & P_2 \\ P_1 & \begin{array}{c|cc} & C & D \\ \hline C & 1 & 1-b-c \\ D & 1+b & 0 \end{array} \end{array}$$

Figure 7.3.1

Consider the following result.

THEOREM 7.3.1.

Suppose an optimal strategy $\mathbf{d}^* = (d_1^*, d_2^*, \ldots)$ *exists for* P_1. *Then, given that* P_1*'s utility function is discounted linear with fixed discount factor* λ, *the optimal move at stage* $k+1$, d_{k+1}^*, *is a function of the past moves only through* P_1*'s current values of* μ, τ *and* ϕ, *for any value of* $k = 1, 2, \ldots$.

PROOF: At the kth stage of the game, it is clear from the above distributional assumptions, that P_1's expected utility for the next move is a function of the past moves only through $\mu(k)$, $\tau(k)$ and $\phi(k)$. Due to the form of P_1's utility function, the expected utility for the $(k+v)$th move $(v > 1)$ is simply a function of $\mu(k+v-1)$, $\tau(k+v-1)$ and $\phi(k+v-1)$. However, these last three quantities are functions of the move sequence up to stage k only through $\mu(k)$, $\tau(k)$ and $\phi(k)$. Thus P_1's expected utility for the next $n \geq 1$ moves is also only a function of the past moves through $\mu(k)$, $\tau(k)$ and $\phi(k)$.

Thus for any strategy $\mathbf{d}$ available to P_1, the expected utility is dependent on the past move sequence only through $\mu(k)$, $\tau(k)$ and $\phi(k)$. Thus an optimal decision $\mathbf{d}^*_{k+1}$ (which exists by assumption) can be made that is only a function of these three parameters. □

This is a special case of Theorem 6.3.2 presented in the previous chapter, which is itself a simple example of the types of results that can be obtained from stochastic control theory. More general results in this area are presented in Ross (1983) and Whittle (1983). Theorem 7.3.1 is therefore a simple corollary of Theorem 6.3.2 that is specific to this example. For any other given example, a similar result could be found.

From Theorem 7.3.1 we can see that at stage k of the game, P_1's optimal $(k+v)$th move depends on the states $(\mu(k), \tau(k), \phi(k))$. But, from the time homogeneous form of P_1's utility function, whenever

$$(\mu(k), \tau(k), \phi(k)) = (\mu, \tau, \phi) \tag{7.3.1}$$

the optimal move is the same. As a result of this it must be possible, for given values of ϕ and λ, to calculate the values of μ and τ where it is optimal to Cooperate, and those values where it is optimal to Defect. Thus we must be able to draw a graph of the regions in (μ, τ) space where Cooperation and Defection are optimal. If this is possible, then we will be able to define the optimal strategy in terms of a simple graph which players could easily refer to. The next theorem shows that in this example we are able to determine what these regions look like. For different examples we would prove similar insightful theorems, before resorting to the numerical calculation of optimal strategies. Before the theorem, we require a lemma.

First we need to define some more notation. At a given stage of the game, let σ be a move sequence that P_1 could employ from the next stage onwards. Also let $C\sigma$ denote the move sequence that is move C on the next move, and subsequent moves are defined by the move sequence σ. Define $D\sigma$, $CD\sigma$, etc. similarly. Then let $\overline{U}(\sigma|\phi, \mu)$ denote P_1's expected utility from playing the move sequence σ, when the mean of P_1's current distribution over P_2's play is μ and the current state of the game is ϕ. Define $\overline{U}(C\sigma|\phi, \mu)$, $\overline{U}(D\sigma|\phi, \mu)$, etc. similarly. Note that the parameter τ has been dropped from the arguments of the utility functions in the following proofs. This is because we are assuming that the priors for the move sequences being compared are equal, thus leading to the same value of τ. Therefore the specific value of τ does not affect the analysis.

LEMMA 7.3.2.

For a given state ϕ, and any strategy $\sigma = (\sigma_1, \sigma_2, \ldots)$, the difference

$$\overline{U}(D\sigma|\phi,\mu) - \overline{U}(C\sigma|\phi,\mu) \tag{7.3.2}$$

depends only on the first move σ_1 of the strategy σ.

PROOF: Let μ' denote the mean of P_1's distribution after he has observed one extra move pair, and μ'' after he has observed two extra move pairs. As P_2 is believed to be playing a partial TFT strategy, the probability of P_2 copying a C move by P_1 is the same as the probability of P_2 copying a D move by P_1. So, P_1's beliefs about μ'' given that his first two moves are (a_1, a_2) must be equal to P_1's beliefs about μ'' given that his first two moves are (b_1, b_2), for any $a_1, a_2, b_1, b_2 \in \{C, D\}$. Thus,

$$\begin{aligned}\overline{U}(D\sigma|\phi,\mu) - \overline{U}(C\sigma|\phi,\mu) &= \overline{U}(D|\phi,\mu) + \lambda\overline{U}(\sigma_1|D,\mu') + \lambda^2\overline{U}(\sigma_2,\sigma_3,\ldots|\sigma_1,\mu'') \\ &\quad - \overline{U}(C|\phi,\mu) - \lambda\overline{U}(\sigma_1|C,\mu') - \lambda^2\overline{U}(\sigma_2,\sigma_3,\ldots|\sigma_1,\mu'') \\ &= \overline{U}(D|\phi,\mu) + \lambda\overline{U}(\sigma_1|D,\mu') - \overline{U}(C|\phi,\mu) - \lambda\overline{U}(\sigma_1|C,\mu')\end{aligned} \tag{7.3.3}$$

which is dependent on σ only through its first move σ_1. □

The utilities from these first two moves can be simply calculated from the general pay-off matrix given in Figure 7.3.1, giving

$$\overline{U}(D\sigma|\phi,\mu) - \overline{U}(C\sigma|\phi,\mu) = \begin{cases} b+\mu(c-1)+\lambda(b+c)(1-2\mu') & \text{if } \phi = D,\ \sigma_1 = C, \\ b+\mu(c-1)+\lambda(b+1)(1-2\mu') & \text{if } \phi = D,\ \sigma_1 = D, \\ b+c-1+\mu(1-c)+\lambda(b+c)(1-2\mu') & \text{if } \phi = C,\ \sigma_1 = C, \\ b+c-1+\mu(1-c)+\lambda(b+1)(1-2\mu') & \text{if } \phi = C,\ \sigma_1 = D. \end{cases} \tag{7.3.4}$$

We show in Theorem 7.3.3 that in the two given situations, a move sequence starting with one of the two moves (i.e. C or D) dominates all move sequences starting with the other move. Because of P_1's beliefs about P_2, any move sequence for P_1 whose first move is the dominated move must obtain a lower expected utility than at least one strategy whose first move is the dominating move.

THEOREM 7.3.3. *If $\mu_A \geq \mu_B$ are the means of two beta distributions and $\tau_A = \tau_B$, in any repeated PDG where P_1 has the discounted linear utility given in equation (7.2.1), then*

(a) if the optimal move given μ_A is Defect, then the optimal move given μ_B is Defect, and

(b) if the optimal move given μ_B is Cooperate, then the optimal move given μ_A is Cooperate.

PROOF: (a) Define σ'' to be the move sequence such that $\overline{U}(D\sigma''|\phi,\mu_A) = \max_\sigma\{\overline{U}(D\sigma|\phi,\mu_A)\}$. Therefore

$$\overline{U}(D\sigma''|D,\mu_A) \geq \overline{U}(D\sigma|D,\mu_A) \geq \overline{U}(C\sigma|D,\mu_A) \tag{7.3.5}$$

and

$$\overline{U}(D\sigma''|C,\mu_A) \geq \overline{U}(D\sigma|C,\mu_A) \geq \overline{U}(C\sigma|C,\mu_A) \tag{7.3.6}$$

for all values of σ. Now, given that the optimal move given μ_A is Defect, irrespective of the state of the game, and then by taking expectations over the parameter μ'_A, we obtain

$$\overline{U}(DD\sigma''|\phi,\mu_A) \geq \overline{U}(DC\sigma''|\phi,\mu_A) \quad \text{and} \quad \overline{U}(CD\sigma''|\phi,\mu_A) \geq \overline{U}(CC\sigma''|\phi,\mu_A) \tag{7.3.7}$$

for any value of ϕ. We shall now prove the result by contradiction. Suppose that there exists a move sequence $\tilde{\sigma}$ such that

$$\overline{U}(C\tilde{\sigma}|\phi,\mu_B) > \overline{U}(D\sigma|\phi,\mu_B) \qquad \text{for all } \sigma. \tag{7.3.8}$$

Then we must have

$$\overline{U}(C\tilde{\sigma}|D,\mu_B) > \overline{U}(D\tilde{\sigma}|D,\mu_B) \quad \text{and} \quad \overline{U}(C\tilde{\sigma}|C,\mu_B) > \overline{U}(D\tilde{\sigma}|C,\mu_B). \tag{7.3.9}$$

Again by taking expectations over the parameter μ'_B, we obtain

$$\overline{U}(DC\tilde{\sigma}|\phi,\mu_B) > \overline{U}(DD\tilde{\sigma}|\phi,\mu_B) \quad \text{and} \quad \overline{U}(CC\tilde{\sigma}|\phi,\mu_B) > \overline{U}(CD\tilde{\sigma}|\phi,\mu_B) \tag{7.3.10}$$

for any value of ϕ. We consider the two cases where $0 < c < 1$ and $c \geq 1$ seperately.

(i) $0 < c < 1$

From Lemma 7.3.2 above we have

$$\overline{U}(DC\sigma|\phi,\mu) - \overline{U}(DD\sigma|\phi,\mu) = \begin{cases} \lambda[-b + \mu'(1-c) + \lambda(b+c)(2\mu''-1)] & \text{if } \sigma_1 = C, \\ \lambda[-b + \mu'(1-c) + \lambda(b+1)(2\mu''-1)] & \text{if } \sigma_1 = D \end{cases} \tag{7.3.11}$$

where μ' and μ'' are as defined in the proof of Lemma 7.3.2,

$$\geq \begin{cases} \lambda[\mu'(1-c+2\lambda(b+c)\frac{\tau}{\tau+1}) - b - \lambda(b+c)] & \text{if } \sigma_1 = C, \\ \lambda[\mu'(1-c+2\lambda(b+1)\frac{\tau}{\tau+1}) - b - \lambda(b+1)] & \text{if } \sigma_1 = D, \end{cases} \tag{7.3.12}$$

as $\mu'' = \{\mu' \frac{\tau}{\tau+1}$ or $\mu' \frac{\tau}{\tau+1} \frac{a+1}{a}\}$, where $\mu' = \frac{a}{\tau}$.

But $(1-c) + 2\lambda(b+1)\frac{\tau}{\tau+1} > (1-c) + 2\lambda(b-c)\frac{\tau}{\tau+1} > 0$, so

$$\overline{U}(DC\sigma|\phi,\mu) - \overline{U}(DD\sigma|\varphi,\mu) \text{ is increasing in } \mu. \tag{7.3.13}$$

Thus from the inequalities (7.3.7),

$$\overline{U}(DC\sigma''|\phi,\mu_A) - \overline{U}(DD\sigma''|\phi,\mu_A) \leq 0 \quad \Rightarrow \quad \overline{U}(DC\sigma|\phi,\mu_B) - \overline{U}(DD\sigma|\phi,\mu_B) \leq 0 \tag{7.3.14}$$

for all move sequences σ and for all $\mu_B \leq \mu_A$, which gives a contradiction to the supposition (7.3.10).

(ii)$c \geq 1$

From Lemma 7.3.2 above we have

$$\begin{aligned}\overline{U}(CC\sigma|\phi,\mu) - \overline{U}(CD\sigma|\phi,\mu) &= \begin{cases} \lambda[1-b-c+\mu'(c-1)+\lambda(b+c)(2\mu''-1)] & \text{if } \sigma_1 = C, \\ \lambda[1-b-c+\mu'(c-1)+\lambda(b+1)(2\mu''-1)] & \text{if } \sigma_1 = D, \end{cases} \\ &\geq \begin{cases} \lambda[\mu'(c-1+2\lambda(b+c)\frac{\tau}{\tau+1}) + 1-(b+c)(1+\lambda)] & \text{if } \sigma_1 = C, \\ \lambda[\mu'(c-1+2\lambda(b+1)\frac{\tau}{\tau+1}) + 1-b-b\lambda-c-\lambda] & \text{if } \sigma_1 = D, \end{cases}\end{aligned} \tag{7.3.15}$$

But $(c-1) + 2\lambda(b+c)\frac{\tau}{\tau+1} \geq (c-1) + 2\lambda(b+1)\frac{\tau}{\tau+1} > 0$, so

$$\overline{U}(CC\sigma|\phi,\mu) - \overline{U}(CD\sigma|\phi,\mu) \text{ is increasing in } \mu. \tag{7.3.16}$$

Thus from the inequalities (7.3.7),

$$\overline{U}(CC\sigma''|\phi,\mu_A) - \overline{U}(CD\sigma''|\phi,\mu_A) \leq 0 \quad \Rightarrow \quad \overline{U}(CC\sigma|\phi,\mu_B) - \overline{U}(CD\sigma|\phi,\mu_B) \leq 0 \tag{7.3.17}$$

for all move sequences σ and for all $\mu_B \leq \mu_A$, which gives a contradiction to the supposition (7.3.10).

Therefore, for any c there cannot exist a move sequence $\tilde{\sigma}$ such that

$$\overline{U}(C\tilde{\sigma}|\phi,\mu_B) > \overline{U}(D\sigma|\phi,\mu_B) \qquad \text{for all } \sigma. \tag{7.3.18}$$

(b) Define σ^* to be the move sequence such that $\overline{U}(D\sigma^*|\phi,\mu_B) = \max_{\sigma}\{\overline{U}(C\sigma|\phi,\mu_B)\}$. Therefore

$$\overline{U}(C\sigma^*|D,\mu_B) \geq \overline{U}(C\sigma|D,\mu_B) \geq \overline{U}(D\sigma|D,\mu_B) \tag{7.3.19}$$

and

$$\overline{U}(C\sigma^*|C,\mu_B) \geq \overline{U}(C\sigma|C,\mu_B) \geq \overline{U}(D\sigma|C,\mu_B) \tag{7.3.20}$$

for any value of σ. Now, given that the optimal move given μ_B is Cooperate, irrespective of the state of the game, and by taking expectations over the parameter μ'_B, we obtain

$$\overline{U}(DC\sigma^*|\phi,\mu_B) \geq \overline{U}(DD\sigma^*|\phi,\mu_B) \quad \text{and} \quad \overline{U}(CC\sigma^*|\phi,\mu_B) \geq \overline{U}(CD\sigma^*|\phi,\mu_B) \tag{7.3.21}$$

for any value of ϕ. We shall again prove the result by contradiction. Suppose that there exists a move sequence $\tilde{\sigma}$ such that

$$\overline{U}(D\tilde{\sigma}|\phi,\mu_A) > \overline{U}(C\sigma|\phi,\mu_A) \qquad \text{for all } \sigma. \tag{7.3.22}$$

Then we must have

$$\overline{U}(D\tilde{\sigma}|D,\mu_A) > \overline{U}(C\tilde{\sigma}|D,\mu_A) \quad \text{and} \quad \overline{U}(D\tilde{\sigma}|C,\mu_A) > \overline{U}(C\tilde{\sigma}|C,\mu_A). \tag{7.3.23}$$

Again, by taking expectations over the parameter μ'_A, we obtain

$$\overline{U}(DD\tilde{\sigma}|\phi,\mu_A) > \overline{U}(DC\tilde{\sigma}|\phi,\mu_A) \quad \text{and} \quad \overline{U}(CD\tilde{\sigma}|\phi,\mu_A) > \overline{U}(CC\tilde{\sigma}|\phi,\mu_A) \tag{7.3.24}$$

for any value of ϕ. Again we consider the two cases where $0 < c < 1$ and $c \geq 1$ seperately.

(i)$0 < c < 1$

From Lemma 7.3.2 above we have

$$\begin{aligned}\overline{U}(DD\sigma|\phi,\mu) - \overline{U}(DC\sigma|\phi,\mu) &= \begin{cases} \lambda[b + \mu'(c-1) + \lambda(b+c)(1-2\mu'')] & \text{if } \sigma_1 = C, \\ \lambda[b + \mu'(c-1) + \lambda(b+1)(1-2\mu'')] & \text{if } \sigma_1 = D, \end{cases} \\ &\geq \begin{cases} \lambda[\mu'(c-1-2\lambda(b+c)\frac{\tau}{\tau+1}) + b + \lambda(b+c)] & \text{if } \sigma_1 = C, \\ \lambda[\mu'(c-1-2\lambda(b+1)\frac{\tau}{\tau+1}) + b + \lambda(b+1)] & \text{if } \sigma_1 = D, \end{cases}\end{aligned} \tag{7.3.25}$$

But $(c-1) - 2\lambda(b+1)\dfrac{\tau}{\tau+1} < (c-1) - 2\lambda(b+c)\dfrac{\tau}{\tau+1} < 0$, so

$$\overline{U}(DD\sigma|\phi,\mu) - \overline{U}(DC\sigma|\phi,\mu) \text{ is decreasing in } \mu. \tag{7.3.26}$$

Thus from the inequalities (7.3.21),

$$\overline{U}(DD\sigma^*|\phi,\mu_B) - \overline{U}(DC\sigma^*|\phi,\mu_B) \leq 0 \quad \Rightarrow \quad \overline{U}(DD\sigma|\phi,\mu_A) - \overline{U}(DC\sigma|\phi,\mu_A) \leq 0 \tag{7.3.27}$$

for all move sequences σ and for all $\mu_A \geq \mu_B$, which gives a contradiction to the supposition (7.3.24).

(ii)$c \geq 1$

From Lemma 7.3.2 above we have

$$\begin{aligned}\overline{U}(CD\sigma|\phi,\mu) - \overline{U}(CC\sigma|\phi,\mu) &= \begin{cases} \lambda[b+c-1+\mu'(1-c)+\lambda(b+c)(1-2\mu'')] & \text{if } \sigma_1 = C, \\ \lambda[b+c-1+\mu'(1-c)+\lambda(b+1)(1-2\mu'')] & \text{if } \sigma_1 = D, \end{cases} \\ &\geq \begin{cases} \lambda[\mu'(1-c-2\lambda(b+c)\frac{\tau}{\tau+1}) - 1+ (b+c)(1+\lambda)] & \text{if } \sigma_1 = C, \\ \lambda[\mu'(1-c-2\lambda(b+1)\frac{\tau}{\tau+1}) - 1+ b+ b\lambda+ c+\lambda] & \text{if } \sigma_1 = D, \end{cases}\end{aligned} \tag{7.3.28}$$

But $(1-c) - 2\lambda(b+c)\dfrac{\tau}{\tau+1} \leq (1-c) - 2\lambda(b+1)\dfrac{\tau}{\tau+1} < 0$, so

$$\overline{U}(CD\sigma|\phi,\mu) - \overline{U}(CC\sigma|\phi,\mu) \text{ is decreasing in } \mu. \tag{7.3.29}$$

Thus from the inequalities (7.3.21),

$$\overline{U}(CD\sigma^*|\phi,\mu_B) - \overline{U}(CC\sigma^*|\phi,\mu_B) \leq 0 \quad \Rightarrow \quad \overline{U}(CD\sigma|\phi,\mu_A) - \overline{U}(CC\sigma|\phi,\mu_A) \leq 0 \tag{7.3.30}$$

for all move sequences σ and for all $\mu_A \geq \mu_B$, which gives a contradiction to the supposition (7.3.24).

Therefore, for any c there cannot exist a move sequence $\tilde{\sigma}$ such that

$$\overline{U}(D\tilde{\sigma}|\phi,\mu_A) > \overline{U}(C\sigma|\phi,\mu_A) \qquad \text{for all } \sigma. \tag{7.3.31}$$ □

From this it is clear that the area of (μ,λ) space where it is optimal to Cooperate is distinct from the area of (μ,λ) space where it is optimal to Defect, for either value of ϕ and for any fixed value of τ. Thus for a given value of τ we can determine a point, μ_1, such that for any mean $\mu \leq \mu_1$ the optimal move is to Defect, and another point, μ_2, such that for any mean $\mu \geq \mu_2$, the optimal move is to Cooperate.

7.3.1 Case when p is known.

First let us consider the special case when the probability p is assumed to be known and therefore $\mu = p$ with probability one at all stages of the game. In this case the optimal move can easily be determined for the PDG in question, and it can be seen from Theorem 7.3.1 that the optimal strategy is going to be one of : continual Cooperation ($\mathbf{S}_C = (C, C, \ldots)$), continual Defection ($\mathbf{S}_D = (D, D, \ldots)$) or Alternation ($(C, D, C, D, \ldots)$ or $(D, C, D, C, \ldots)$).

We can determine the expected utility obtained from each of these and thus determine which of these is the optimal strategy for a particular value of p (known). Graphically, the areas in which these strategies are optimal partition the (μ, λ) space. So we just need to calculate where the edges of these areas are, i.e. the values of μ_1 and μ_2 for varying λ and for both values of ϕ.

It is easy to calculate the expected utility from each of these strategies. For instance, the expected utility from strategy S_C in the example in question when the state is $\phi = C$, is simply

$$\begin{aligned}\overline{U}(\mathrm{S}_C) &= [\mu + (1-\mu)(-\frac{1}{2})](1 + \lambda + \lambda^2 + \lambda^3 + \ldots) \\ &= \frac{1}{2}(1-\lambda)^{-1}(3\mu - 1). \qquad (7.3.32)\end{aligned}$$

Comparing these expected utilities we can determine the region in (μ, λ) space where each strategy is optimal. We can see from this that $\mu_1 = \mu_2$ for all values of r and ϕ. Note that the Alternating strategy is optimal because the value of μ_1 is different for different values of ϕ. If $\phi = C$ then μ_1 has a lower value than if $\phi = D$. Therefore, if μ lies between these two values it will be utility maximising to play an alternating strategy.

Let $\overline{U}(\mathrm{S}_A)$ denote the expected utility from the Alternating strategy starting with the opposite move to the current state. The other Alternating strategy, i.e. that starting with the move equal to the current state, is always dominated by one of the other strategies. Then if p is known we obtain, irrespective of ϕ,

$$\overline{U}(\mathrm{S}_D) > \overline{U}(\mathrm{S}_A) \text{ when } \mu < \frac{4\lambda + 2}{8\lambda + 1} \qquad (7.3.33)$$

$$\overline{U}(\mathrm{S}_C) > \overline{U}(\mathrm{S}_A) \text{ when } \mu > \frac{3\lambda + 1}{6\lambda - 1} \qquad (7.3.34)$$

which are sufficient to find the optimal regions. These regions can be seen more clearly graphically, as shown in Figure 7.3.2. A similar figure can be obtained for any specified pair of b and c, as is stated in section 5 of this chapter.

Throughout the rest of this section we shall assume the discount factor (λ) to be fixed at the level of 0.9 in line with the example given in Wilson's paper. It seems reasonable to assume that a player will believe one discount factor to be appropriate for the whole game. It is

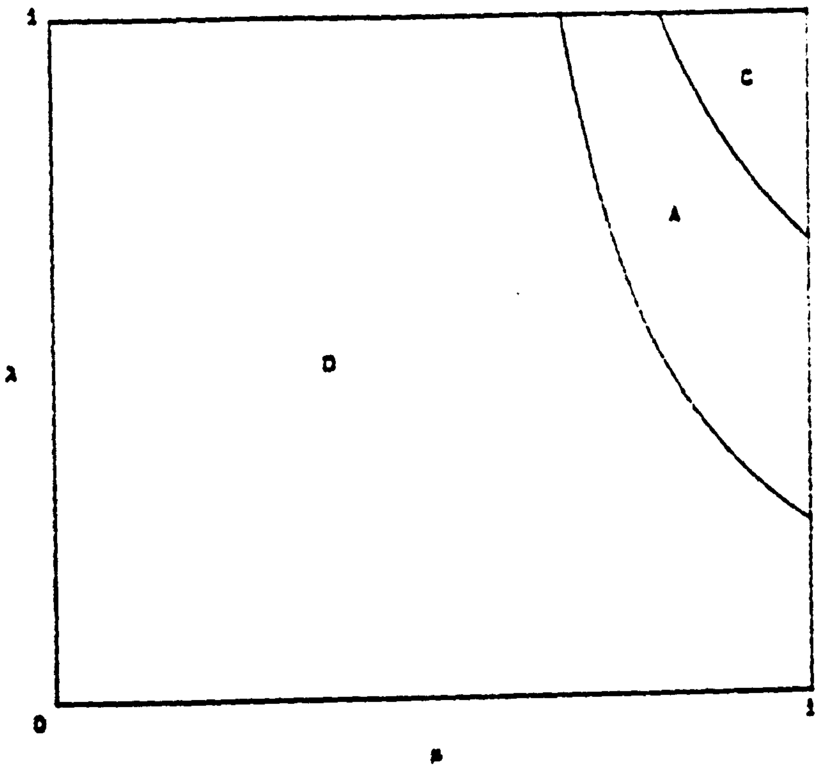

Figure 7.3.2

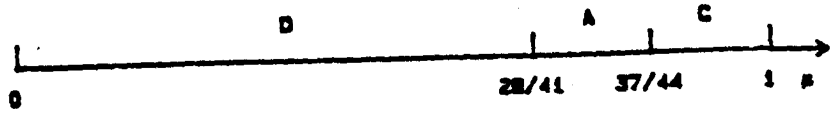

Figure 7.3.3

possible to vary the value of λ as the game progresses, although why any player would wish to do this is not immediately obvious. Given this information we need only consider the cross–section of Figure 7.3.2 given in Figure 7.3.3.

7.3.2 Case when p is unknown.

Now, obviously in practice P_1 will be uncertain about the probability p. However, the regions above correspond to the limiting optimal strategy (where $r = \infty$), as the number of moves played increases and the distribution of p degenerates.

The explicit algebraic solution to this has not been found in closed form, but we can capture all the dynamics of the problem by considering Theorem 7.3.1 and Theorem 7.3.3. Wilson's method will determine the optimal move for any given situation as discussed earlier. However, we show in the next section that, by considering the nature of the solution we can find strategies that are much easier to calculate than Wilson's and are very close to the optimal strategy.

7.4 Determining Optimal Play.

In section 7.2 we gave the algorithm that Wilson (1986) uses to calculate the optimal next move. This involves finding an upper and a lower bound (S and L) for the expected utility of any strategy at all subsequent stages of the game. In fact these bounds are extremely loose and tighter bounds can easily be determined. If we were to replace the bounds given above with tighter ones, we would decrease the amount that one maximum expected utility ($f_n(1)$ or $f_n(2)$) needs to be greater than the other, for the former to determine the optimal next move. Therefore the algorithm can calculate the optimal move much earlier, and so the method has been speeded up.

For instance, an improved upper bound, S, is simply the expected value of perfect information (EVPI). That is, suppose P_1 is told what the true value of p is. He can then play the optimal move given this information for the rest of the game, in the same way that he would in the case where the probability p is known. So this is saying that the highest utility P_1 is likely to achieve with any strategy is the same as if he knows the probability p with probability one. This depends upon the value of p used in the EVPI strategy, which must be such that P_1's expected utility is maximised. This is at most the utility that P_1 would obtain from a pay–off of M from every future move, and can be less than the original upper bound. Therefore the utility from the EVPI strategy must provide a tighter bound than S given in the previous section. It must also be an upper bound, as P_1 could never expect to obtain a higher pay–off than when he knows the value of p with probability one.

Tighter lower bounds, L, can also be found. One strategy that gives such a bound is where P_1 determines his move from a figure like Figure 7.3.3, but uses the mean μ of his distribution at each stage of the game, *as if* it were equal to p with probability one. Call this strategy $\hat{\sigma}$. This is a strategy that P_1 could employ. However there are instances when this strategy does not give the optimal next move. Therefore it is not a utility maximising strategy and so forms a lower bound. The expected utility for this strategy can be calculated and then, when comparing maximum expected utilities, we know that any move sequence that P_1 is considering must achieve at least this amount on the subsequent stages of the game. This amount is obviously greater than the utility that P_1 would obtain from a pay–off of $-M$ on all future moves. So the lower bound could be increased to this value.

Another lower bound is the strategy where P_1 uses his mean μ at the present stage of the game, *as if* it were equal to p with probability one, to determine whether to play S_C, S_D, or S_A for the whole of the rest of the game. Therefore, at a given stage, P_1 constrains himself to only play one particular strategy (S_C, S_D, or S_A) on *all* subsequent moves. Again this is a strategy that P_1 could adopt, and obviously will not always be optimal, so forms another lower bound. This latter strategy is easier to calculate than $\hat{\sigma}$, but is however a looser bound.

Now we have shown from the theorems in the previous section that the optimal solution for any PDG can be defined simply in terms of the regions where it is optimal to Defect, Alternate or Cooperate for any values of μ and r. Obviously here, to 'Alternate' is to simply make the opposite move to the current state, at each stage of the game. Therefore, in stead of calculating the optimal next move for all values of μ and r, we just need to find where the boundaries of these regions lie in (μ, r) space. Using Wilson's algorithm with the improved bounds just suggested to determine these regions, we quickly obtained Figure 7.4.1 which compares the optimal regions (continuous line) of the parameter space (μ, r, ϕ) for each move, with the regions for the limiting strategy (dotted line), in the example under consideration. From Theorem 7.3.1, we can see that the form of the solution, and therefore the optimal strategy, can be found for any PDG.

Therefore, to drawing accuracy at least, it is simple to find the optimal strategy. We have now overcome the problem that, as Wilson comments "it is impossible to tell [P_1] the moves

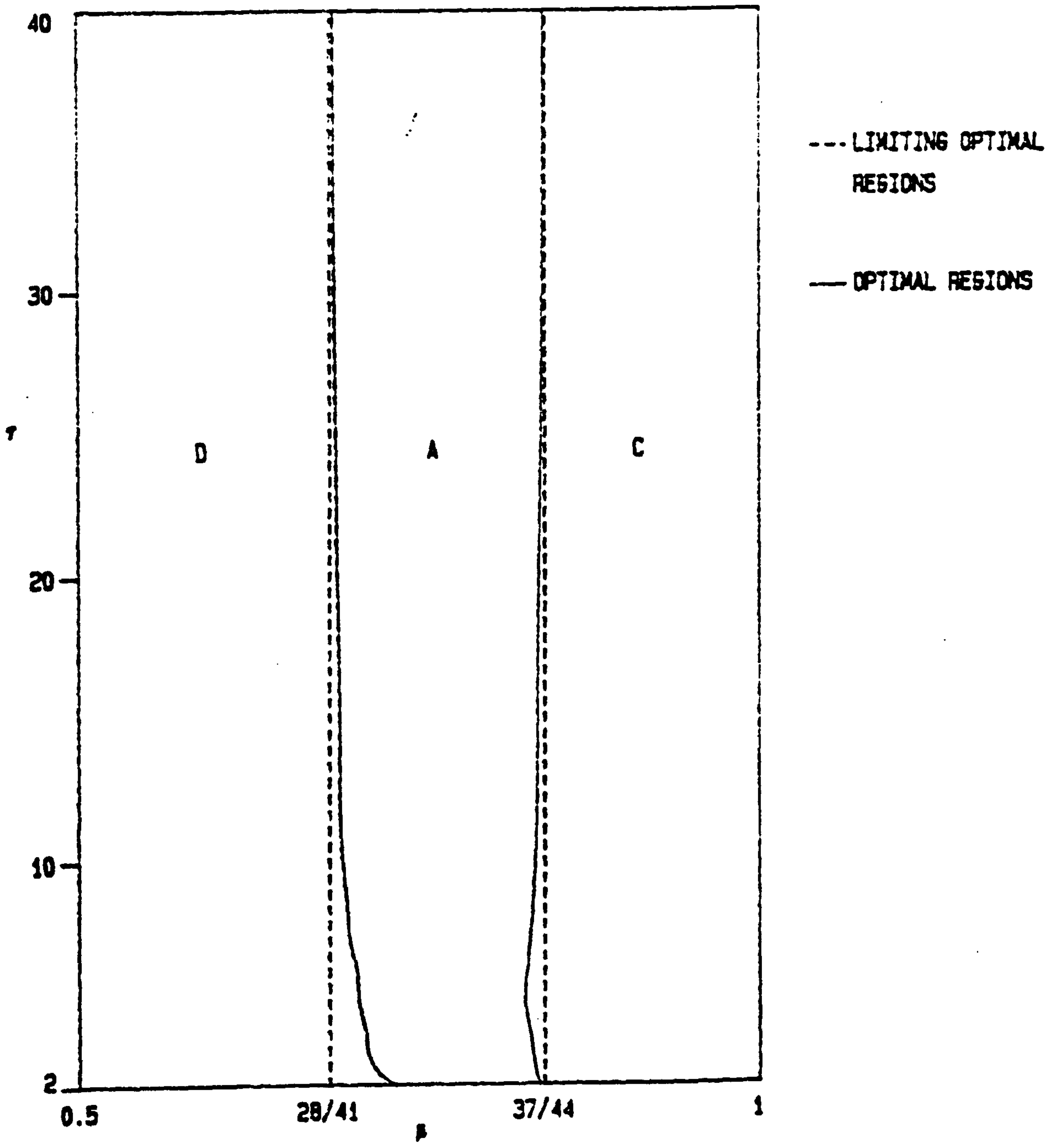
40
30
20
10
2
γ
0.5
28/41
37/44
1
β
D
A
C
---LIMITING OPTIMAL REGIONS
—OPTIMAL REGIONS

Figure 7.4.1

he should make for every eventuality as the set of [parameters] is infinite." Although our 'solution' is still in terms of a procedure for determining the optimal move, a simple graph can be drawn to determine optimal moves for the whole game. This is an obvious improvement on the Wilson methodology used on its own, which requires us to calculate the optimal move for all possible combinations of the parameters μ, τ and ϕ.

P_1 will continue to play this strategy, i.e. the move corresponding to $\mu(k+v)$ and $\tau(k+v)$ at stage $k+v+1$. The parameter $\tau(k+v)$ is known and, although $\mu(k+v)$ is currently unknown, its distribution can be calculated from $\mu(k)$. So it is possible for P_1 to determine his strategy for all subsequent stages of the game, and the expected utility associated with it. This strategy may or may not change as P_1 receives more information about the value of the parameter p.

Note that the strategy $\hat{\sigma}$, mentioned above as a lower bound, is very close to being optimal. In the PDG that Wilson considered in his example (Figure 7.2.1), the value of μ has to be so close to the boundary for $\hat{\sigma}$ to give a suboptimal next move, that it does so for only 13 distributions where $\tau \leq 100$, in the case where $\alpha(t)$ and $\beta(t)$ take integer values for all t.

Also, comparing strategy $\hat{\sigma}$ with the optimal results given in Table 1 of Wilson (1986, page 52) it only gives one suboptimal move out of ten, and in table 2 all ten are correct. The only discrepancy occurs because this point lies very close to one of the boundaries in Figure 7.4.1. It may appear that $\hat{\sigma}$ gives two suboptimal moves in Table 1: History $(1,1,1)$ and History $(2,1,2)$. However, in checking these results we find History $(2,1,2)$ should read 1 *not* 2. The loss in expected pay-off from using the limiting strategy for all values of τ is therefore nearly always negligible.

So, by considering the form of the optimal solution we have found a way of speeding up a rather slow method of determining the expected utility maximising next move, and a way of representing such moves on an easy to read graph. In the process of doing this we have found a strategy, namely $\hat{\sigma}$, that is extremely close to the optimal strategy, especially for large values of τ. The advantage of this strategy is that it is very easy for P_1 to use. At any stage of the game, he just needs to calculate the mean of his distribution over how he believes P_2 will play, and then determine which area it lies in on a simple graph to work out which move to play.

Now it might appear that the method used here is only applicable if P_1 believes P_2 to be playing a partial TFT strategy. However, Wilson's algorithm depends upon probabilities dependent on the past history of the game being specified. These probabilities can be determined as artifacts of a particular probabilistic model, or they can be specified individually. Once specified, the form of the optimal solution can then be determined from the structure of these probabilities, in a similar way to that shown for the partial TFT case. So, the only requirements of this model are that at all stages of the game an optimal move can be determined, and that a probabilistic structure exists. The more structure that exists within the model, the more helpful the derived form of the solution will be.

7.5 Extensions.

7.5.1 Discount Factor.

As can be seen from Figure 7.3.2 in section 7.3, any value of the discount factor $\lambda \in [0,1)$ can be accommodated. We stated earlier that in most games a constant discount factor would appear most realistic. However, as any value of λ can be used, the model is capable of handling a dynamic discount factor. Also any value of λ is permissable if a constant discount factor is required. The values for the boundaries of the regions for different values of λ can be simply read from this graph.

The accuracy of strategy $\hat{\sigma}$ (i.e. how far the boundaries for this strategy are from that of the optimal strategy) is virtually unaffected by a change in λ. For example, the change in accuracy from varying λ from 0.9 to 0.99 in the above example is less than 0.0025 for all values of r, decreasing as r increases, and is virtually non–existent for $r \geq 30$. So for different values of λ we have different limiting strategies and therefore different optimal strategies. The strategy $\hat{\sigma}$ for each value of λ is approximately the same distance away from the optimal strategy for a given value of r.

For small values of λ the only optimal move at any stage of the game is D, and thus $\hat{\sigma}$ determines the exact optimal strategy for all values of r. The special case where $\lambda = 0$ effectively makes the game a one–move game, and in this example the only utility maximising move for all values of μ, r and ϕ is move D. Note that at the other end of the range, where $\lambda = 1$, the utility function is not discounted, and therefore the optimal move cannot be calculated by Wilson's method. This is because the pay–offs at all subsequent stages could be infinite. The limiting strategy and $\hat{\sigma}$ can, however, be determined. Also, from the form of the solution we can see that after observing P_2's play for a reasonable length of time ($r > 30$, say), P_1 will become fairly confident about the value of p, and can therefore work out how to play the remaining stages of the game. It should also be noted that despite the accuracy of $\hat{\sigma}$ decreasing as $\lambda \to 1$, the difference is small.

7.5.2 Other Games.

Consider the general pay-off matrix given in Figure 7.3.1. For any values of b and c, subject to the constraints for the game to be a PDG ($b, c > 0$ and $b + c > 1$), it is easy to calculate the boundaries for the relevant limiting optimal regions (i.e. when $\tau = \infty$), which must exist due to Theorem 7.3.3. It is interesting to note that the Alternating region in the limiting case decreases as c increases, and does not exist for $c \geq 1$. The general inequalities for $c \leq 1$ are, irrespective of ϕ,

$$\overline{U}(\mathbf{S}_D) > \overline{U}(\mathbf{S}_A) \text{ when } \mu < \frac{\lambda(1+b)+b}{2\lambda(1+b)+1-c} \tag{7.5.1}$$

$$\overline{U}(\mathbf{S}_C) > \overline{U}(\mathbf{S}_A) \text{ when } \mu > \frac{\lambda(b+c)+b+c-1}{2\lambda(b+c)+c-1} \tag{7.5.2}$$

where $\mathbf{S}_A$ again denotes the Alternating strategy, starting with the opposite move to the current state.

When $c > 1$ we still have three possible strategies despite there being no Alternating strategy. The three strategies in these games are $\mathbf{S}_C$, $\mathbf{S}_D$ and $\mathbf{S}_R$ where

$$\mathbf{S}_R = \begin{cases} \text{move } C & \text{if } \phi = C \\ \text{move } D & \text{if } \phi = D \end{cases} \tag{7.5.3}$$

This occurs because the dividing line between C being optimal and D being optimal is at a lower value of μ for $\phi = C$, than for $\phi = D$. The Alternating region occurs for $c < 1$ precisely because this dividing line occurs at a higher value of μ for $\phi = C$ than for $\phi = D$. This is shown diagramatically in Figure 7.5.1 for the case where $b = 1$ and $c = 2$. The general inequalities for these limiting regions are

$$\overline{U}(\mathbf{S}_C) > \overline{U}(\mathbf{S}_D) \text{ if } \mu > \frac{\lambda(b+1)+b+c-1}{2\lambda(b+1)+c-1} \text{ for } \phi = C \tag{7.5.4}$$

$$\overline{U}(\mathbf{S}_C) > \overline{U}(\mathbf{S}_D) \text{ if } \mu > \frac{\lambda(b+c)+b}{2\lambda(b+c)-(c-1)} \text{ for } \phi = D \tag{7.5.5}$$

Graphs of the boundaries of the limiting case (and therefore of strategy $\hat{\sigma}$) where $\lambda = 0.9$, $b = 1$, and c varies from 0 to 5 is given in Figure 7.5.2.

When $c = 1$ the $\hat{\sigma}$ strategy appears to be virtually indistinguishable from the optimal strategy, for all values of b and τ. At any stage we can compare the expected utility from

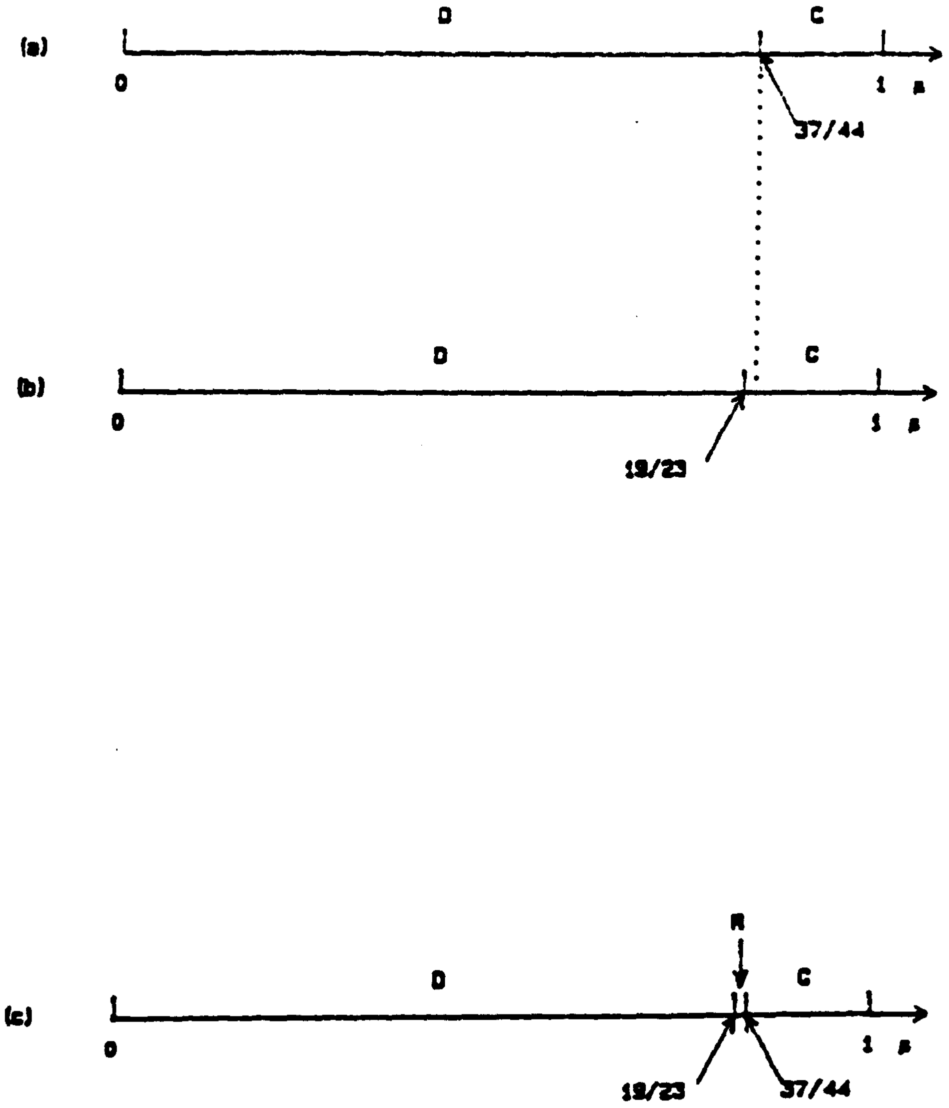

Figure 7.5.1

(a) $\phi = D$

(b) $\phi = C$

(c) summary for all ϕ

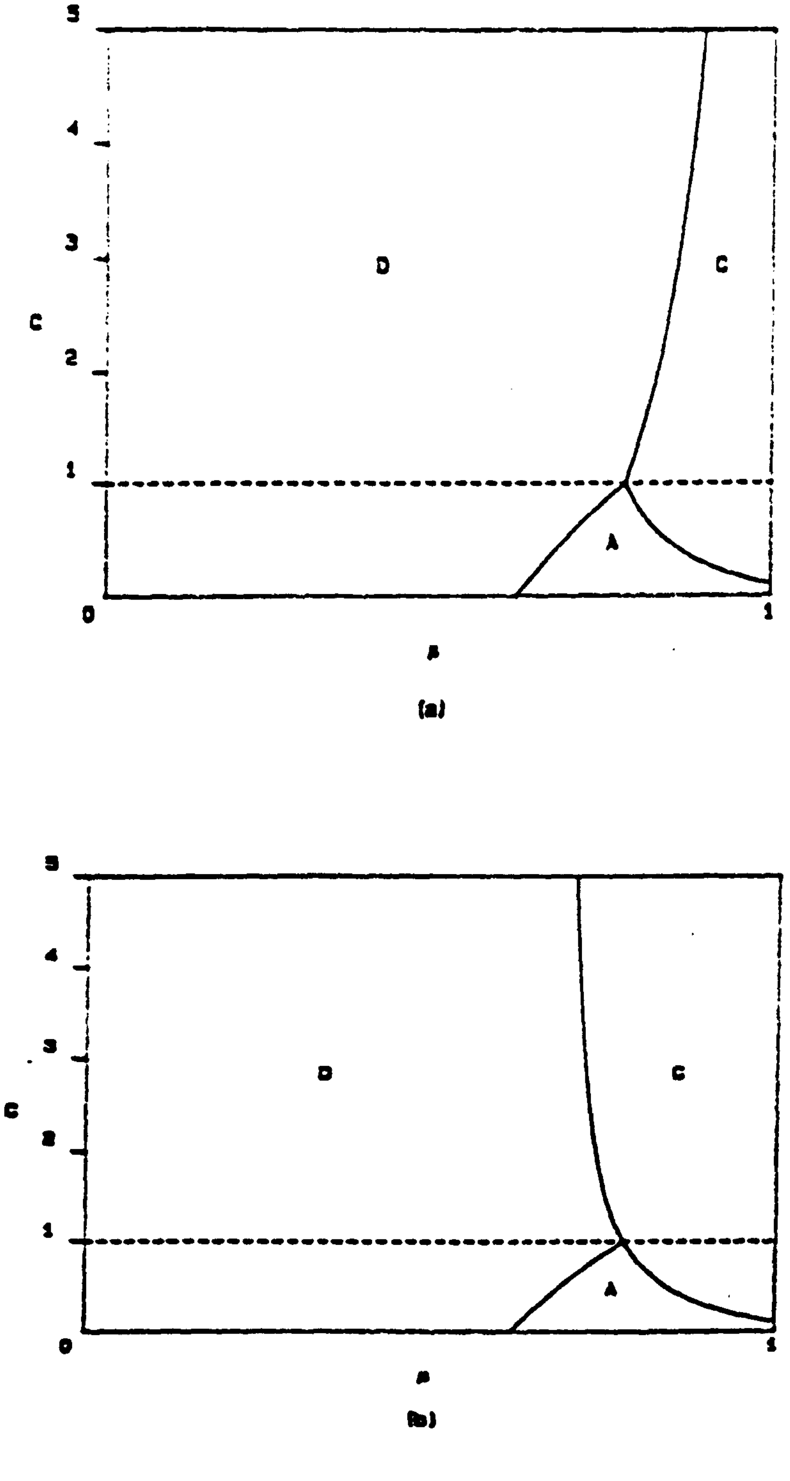

Figure 7.5.2

(a) $\phi = D$

(b) $\phi = C$

making move C and making move D. From equations (7.3.4) we find that, for any move sequence σ and either state,

$$\overline{U}(D\sigma|\mu) - \overline{U}(C\sigma|\mu) = b + \lambda(b+1)(1-2\mu') \tag{7.5.6}$$

where μ' is as defined in section 7.3, and ϕ and τ are omitted, as in all comparitive terms they must be equal. So the optimal move is C if

$$\mu' > \frac{b+\lambda(b+1)}{2\lambda(b+1)}, \tag{7.5.7}$$

and is D if

$$\mu' < \frac{b+\lambda(b+1)}{2\lambda(b+1)}. \tag{7.5.8}$$

Also we can check that if the mean $\mu > \frac{b+\lambda(b+1)}{2\lambda(b+1)}$ then the optimal move in the limiting case is C, and if $\mu < \frac{b+\lambda(b+1)}{2\lambda(b+1)}$ then the optimal move in the limiting case is D. Therefore, as the expected value of μ' is μ, the unique optimal move for any value of τ is the same as the unique optimal move for the limiting case, because the discrepancy in expected utilities remains constant. So, the strategy $\hat{\sigma}$ determines the optimal move for all values of μ and τ. Further, the discrepancy of this strategy $\hat{\sigma}$ from the optimal is small for all $0 < c < 1$, and despite increasing for $c > 1$, it is never large. We consider an example of a game where $c > 1$ in the next section.

If we are considering a general pay–off matrix as opposed to a PDG pay–off matrix, then there is no reason, in general, why we should not employ the same methodology as that given above. Theorem 7.3.3 will not hold in general for non–PDG games, but a result similar to Theorem 7.3.1 can be found for all 2 player games. Therefore we can construct regions for the optimal moves, but these regions will not necessarily be as well behaved as those calculated for PDGs.

Also, it is possible to use distributions other than the Beta for a prior distribution in this problem. The argument of Theorem 7.3.1 still goes through, but because of the lack of conjugacy, the prior parameters have also to be considered in the vector of states, thus leading to a much more complicated model.

7.6 A Further Example.

To illustrate that this method can be extended to other PDGs, we shall now consider the game that is generated by the pay–off matrix given in Figure 7.6.1. We shall calculate the limiting optimal regions (and hence the strategy $\hat{\sigma}$) as well as the optimal regions for this new game.

$$\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} C & D \end{array} \\ P_1 & \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1 & -2 \\ 2 & 0 \end{pmatrix} \end{array}$$

Figure 7.6.1

Referring to the general PDG matrix given in Figure 7.3.1, we note that, in the game that we are considering here, $b = 1$ and $c = 2$. Hence we have the situation where $c > 1$. As was discussed in the previous section, the region where the alternating strategy is optimal does not exist for this game, but instead we have another optimal strategy, $\mathbf{S}_R$, which is defined in the equations (7.5.3). So, once again there are three strategies that are optimal for different regions of (μ, τ) space.

Now, as the game is a PDG, the analysis of section 7.3 still applies, and we can work out the limiting optimal regions in a similar manner. Also, Wilson's algorithm will obviously work for this game, and so we can find the optimal regions and hence the utility maximising strategy for all values of μ, τ and ϕ.

From the inequalities (7.5.4) and (7.5.5) we can calculate that, in the limiting case, Cooperation is preferred to Defection at the next stage of the game if

$$\mu > \frac{2\lambda + 2}{4\lambda + 1} \quad \text{if } \phi = C \tag{7.6.1}$$

$$\text{and} \quad \mu > \frac{3\lambda + 1}{6\lambda - 1} \quad \text{if } \phi = D. \tag{7.6.2}$$

These regions are shown graphically in Figure 7.6.2. As in the previous example, we shall fix upon one particular value of the discount factor λ, and again we shall choose the value $\lambda = 0.9$. This gives the inequalities such that a Cooperation move is optimal,

$$\mu > \frac{38}{46} \quad \text{if } \phi = C \tag{7.6.3}$$

$$\text{and} \quad \mu > \frac{37}{44} \quad \text{if } \phi = D. \tag{7.6.4}$$

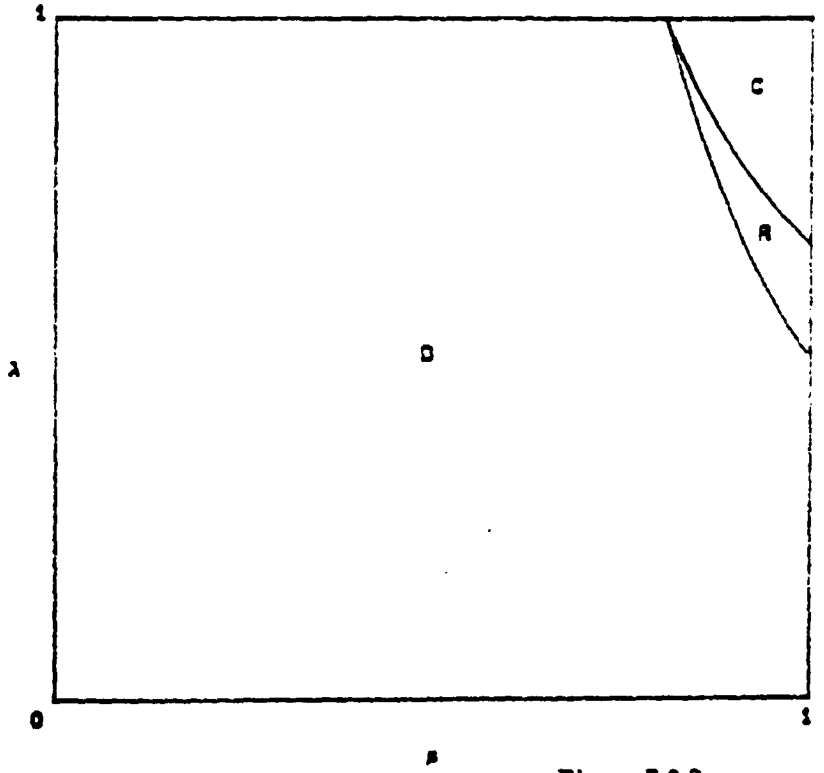

Figure 7.6.2

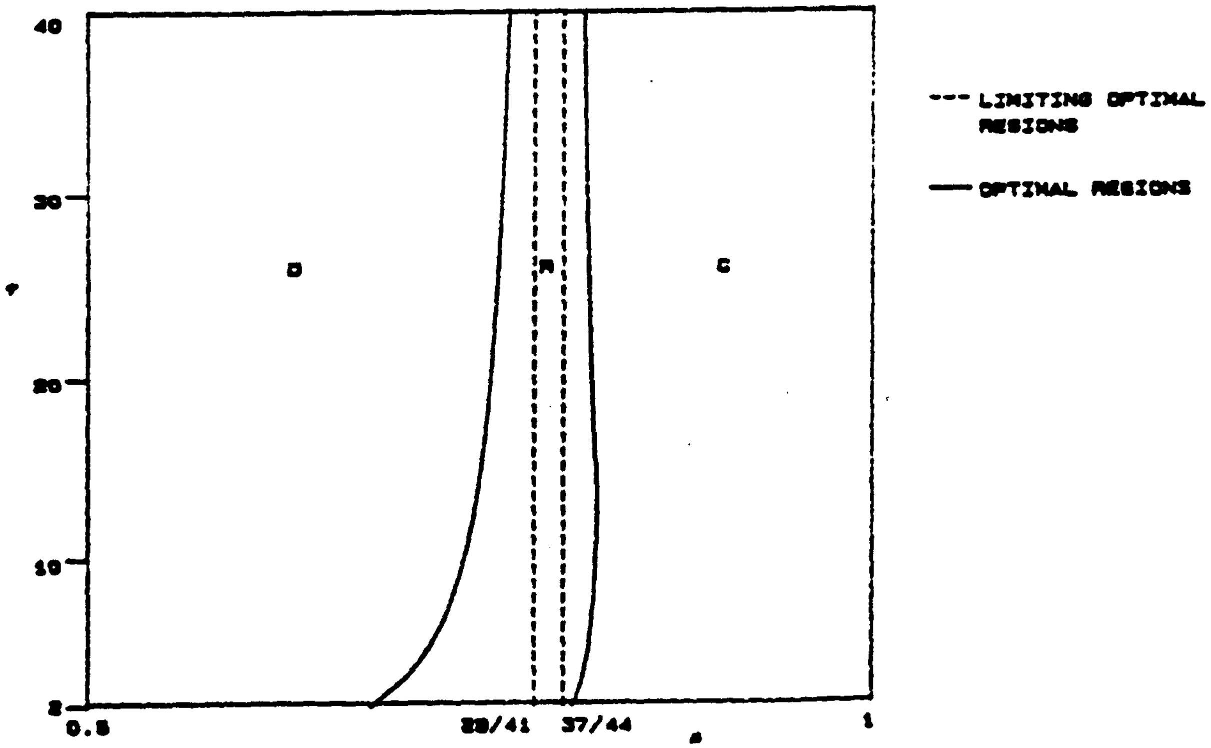

Figure 7.6.3

These obviously determine the limiting regions where the three strategies $\mathbf{S}_C$, $\mathbf{S}_D$ and $\mathbf{S}_R$ are optimal. We can then apply Wilson's algorithm, with the tighter bounds suggested in section 7.4, to determine the precise regions where these strategies are optimal, for any values of μ, τ and ϕ. The regions for the τ finite, and the limiting ($\tau = \infty$) cases are presented in Figure 7.6.3.

As was discussed in the previous section, the accuracy of the strategy $\hat{\sigma}$ is less for the case where $c > 1$, but it is still good. So, again it as easy to read off from the graph that has been produced (Figure 7.6.3) the move which is optimal for particular values of the parameters. Also, the effect of a change in the values of these parameters can be calculated (e.g. by assuming different prior beliefs about P_2's probability of mimicking). It is therefore a simple task to modify the analysis to another PDG.

7.7 Testing the Appropriateness of the Model.

The model used above assumes P_2 to be playing a partial tit–for–tat strategy. Grofman & Pool (1975) base their analysis of optimal play in PDGs on this assumption, with the parameter p known. It is shown that if P_2 is assumed to be playing such a strategy in a specific class of PDGs, then the optimal 1–step back strategy for P_1 is always either continual Cooperation or continual Defection. When the parameter p is unknown, we have shown that in any PDG, P_1 should choose between Alternation, continual Cooperation and continual Defection at all stages, but the choice depends on all of the past move sequence. Thus to play optimally against P_2, P_1 chooses from a different class of strategies. The question now is — why did P_2 choose the partial tit–for–tat strategy in the first place? Even if he assumes P_1 does not know p but only how to estimate it on–line, as above, then P_2 will play very differently to P_1.

One possibility is that P_2 has assumed P_1 to have a utility function different to the one he uses himself. However, we show in the next chapter that, if P_1 believes P_2 to be playing any unknown *n–step back strategy*, then under weak regularity conditions that restrict P_1's utility function to be a sensible function of the pay–offs in the game only, any Bayes strategy for P_1 is not an m–step back strategy, for any m.

In the above, the class of n–step back strategies is a wide class of strategies, that includes partial tit–for–tat strategies, where a player's move at any stage of the game is dependent on the past move sequence only through the last n move pairs. Thus, in the game we are considering at present, for P_1 to believe P_2 to be rational, P_1 must believe P_2 to have a very strange utility function which is very different from his own.

Of course, depending on the problem, P_1 may well believe that P_2 may not be acting rationally. This could be because the opponent is constrained in some way to play a particular type of strategy, by some law, social norm or company policy. Some authors have considered the effects of relaxing the rationality assumptions (e.g. Selten, 1975 and Simon, 1957).

However, by considering the form of the optimal solution we can directly address the issues arising from considering the rationality of an opponent. These issues cannot be addressed by deriving the optimal next move alone. By using the form of the optimal solution and Bayesian inference, we can question the appropriateness of the model assumed in the particular circumstances that the game is being played under, and the implications of such a model. Bayesian models which *do* have the property of mutual rationality can be constructed, and some are given in chapter 6 above.

7.8 Conclusions.

Wilson's backward induction method for calculating optimal solutions for Bayesian models of games provides a valuable algorithm, but it is vastly improved if used in conjunction with the analytic form of an optimal solution. The method used by Wilson requires a probabilistic structure in order to determine the optimal solutions, and the richer this structure is, the more we will be able to deduce from the form of the solution. By using this extra information we can gain insight into which moves P_1 should make on *all* subsequent stages of the game, and these can be determined from a simple graph. We can also use the form of the solution to criticise and adapt P_1's model of how he expects his opponent to play.

Some of the work in this chapter has previously been reported in Young & Smith (1988b).

8. SUBOPTIMALITY OF M–STEP BACK STRATEGIES

8.1 Introduction.

In the previous chapter we discussed a Bayesian decision theoretic approach as an alternative to the *rational* approach of determining how to play a specified game, which was described in the chapter before that. To limit the complexity of the algorithm it was necessary to assume that the opponent in this 2 player game, P_2, was playing a strategy which at the time of the t^{th} move took account of only a limited history of the past move sequence, for $t = 1, 2, \ldots$. As a central example of this method, Wilson (1986) chose a model that assumes P_2 to be playing a strategy that belongs to the class of strategies that, at any stage of the game, only consider the previous n move pairs (an *n–step back strategy*).

However, the optimal strategy for P_1 that we determined for this game was not an m–step back strategy, for any value of m. This strategy depended on all of the previous moves to date. Therefore P_1 assumed the strategy that P_2 was using to be of a type that he would not use himself in response to P_2's strategy.

In this chapter we consider any game where P_1 assumes P_2 to be using a strategy that is a function of all previous move pairs only through the last n move pairs, for some value of n. Initially, we also consider P_1 to be playing a strategy of the same type himself. We shall show that, under mild regularity conditions, no m–step back strategy for P_1, $m = 1, 2, \ldots$, can ever be optimal for a rational player, when that player believes his opponent is playing an n–step back strategy, for some n. That is, there is always another strategy that is not of this form (i.e. can depend on all of the past move sequence) which will obtain a higher expected utility than the utility maximising m–step back strategy, for any m.

This should be very disturbing for P_1 in the context of, for example, experimental games when he is using a particular probability model of how P_2 would play in every possible state of the game. Suppose he believes that his opponent has a probability model similar to his own and has any *sensible* utility function (for a definition of the term sensible here, see the regularity condition (8.2) of section 8.2.2). Then P_1 has implicitly assigned probability one to the event that P_2 chooses a strategy that would be suboptimal for P_1 to use, if the roles were reversed.

Thus, by using this probability model, P_1 implicitly assumes that if P_2 is rational he is employing a model of P_1's behaviour which is quite different from P_1's model of P_2's behaviour. Indeed, models of P_2's beliefs about P_1 that are consistent with P_2 playing an m–step back strategy are most peculiar and, in my opinion, not realistic in the context of symmetric two player games, played by players drawn from a homogeneous population. On the other hand, models which imply very reasonable belief systems for P_2 abound for most games (as we argued in chapter 6). I contend that models which allow both P_1 and P_2 to be rational should, in general, be preferred to classes of models like that used in Wilson's example, which do not.

We shall also discuss more general games, and show that analagous results hold when P_1 and P_2 have very different prior opinions about the class of models that the other player might choose to employ. We also consider an example that does not violate the mutual calibration concept that both players assume each other to be utility maximising.

8.2 Notation and Assumptions.

We shall begin by only considering a *binary repeated game* in which the 2 players, P_1 and P_2, each have the choice of playing one of move 0 or move 1 at every stage of the game. Throughout this chapter we shall assume that P_2 is playing an n–step back strategy. Also we shall make an initial assumption that P_1 is playing an m–step back strategy.

Definition. An *n–step back strategy* is defined to be a decision rule that always plays move 0 with probability $\rho^{(i)}$ if the game is in state $s^{(i)}$, where $s^{(i)}$ is a given history of the 2 players' previous n moves (i.e. n move pairs).

Note that a player commits himself to a particular n–step back decision rule once he has played a certain move when the game is in state $s^{(i)}$, and he must subsequently employ the same decision rule whenever the game is in state $s^{(i)}$ again. In particular, his choice of strategy when the game is in state $s^{(i)}$ can only be based on his prior information and any information gained from any move sequences before the first occurence of $s^{(i)}$. When the game is in state $s^{(i)}$ for the second, third, etc. time, the player cannot use additional information that he has received since the first occurence of state $s^{(i)}$ — he is committed to the same decision rule. Clearly, an n–step back strategy is also a z–step back strategy for any $z \geq n$. This is because if a move is uniquely determined by only a player's prior information and the last n move pairs, it must also be determined by prior information and the last z move pairs.

In line with the rest of this thesis, we shall be restricting the repeated games that we are considering to be ones which are non–cooperative, i.e. no enforceable agreements can be made, and the moves are made simultaneously.

We shall now set up some notation and define our assumptions.

8.2.1 Notation.

Firstly we define S to be the set of all possible z–step back states, $S = (s^{(1)}, s^{(2)}, \ldots, s^{(y)})$, where $y = 2^{2z}$ and $s^{(i)}$ is defined as the binary expansion of $(i-1)$, indicating the last z moves for both players, and $z = \max\{m, n\}$. As noted above, n–step back strategies and m–step back strategies are both simply special cases of z–step back strategies, and so we can concentrate on these z–step back strategies for conformity. For example, if both players are assumed to be playing 1–step back strategies, then obviously $z = 1$ and $S = \{(0,0), (0,1), (1,0), (1,1)\}$.

Then ρ is defined to be the vector of probabilities determined by P_2's z–step back decision rule, where ρ can be expressed as $(\rho^{(1)}, \ldots, \rho^{(y)})$ and $y = 2^{2z}$. Each $\rho^{(i)}$ corresponds to the probability that P_2 will play move 0 when the game is in state $s^{(i)}$, $i = 1, \ldots, 2^{2z}$. In the case where $n < z$, certain states will always have the same decision rule associated with them as those associated with other states. That is, any state, $s^{(j)}$, that differs from state $s^{(k)}$ only on move sequences after z steps back, but before n steps back in the move sequence will have an associated probability $\rho^{(j)} = \rho^{(k)}$. Also we shall denote by ρ_0 the true value of ρ, which is assumed to be unknown to P_1, although he does have beliefs about it.

The utility maximising m–step back decision rule for P_1 will be defined by $\mathbf{d}^*_m$. As it is m–step back, $\mathbf{d}^*_m$ must determine how P_1 plays given the last m move pairs, and therefore how P_1 plays given the last z move pairs, where $z = \max\{m, n\}$, i.e. for any z–step back move sequence i, $i = 1, \ldots, 2^{2z}$. This must hold for every occurence of the move sequence i throughout the game, and so must be determined by the time that move sequence i first occurs. $\mathbf{d}^*_m$ therefore plays the move that maximises the expected utility given P_1's distribution on ρ at the time that move sequence i first occurs. P_1's distribution on ρ may change as P_1 observes further move sequences, but $\mathbf{d}^*_m$ is committed to play in a certain way whenever the game is in state $s^{(i)}$, which is determined before $\rho^{(i)}$ has even been observed once.

Then we define S to be the set of states belonging to S that are positive recurrent with probability one under the action of $\mathbf{d}^*_m$. Note that as both m and n are finite, the set S is finite, and thus S is non–empty. This is because at least one state in S must be recurrent and none of the states in S can be null–recurrent. In addition to this, T will denote the termination time of the game.

We shall then denote by τ_0 the stage of the game when some fixed state $s_0 \in S$ is achieved for the first time, under the action of d^*_m. As the game progresses, we determine τ_k, the stage of the game when s_0 is achieved for the first time, under the action of d^*_m, *after* every state in S have been observed at least k times after stage τ_0, $k = 1, 2, \ldots$. Hence, by stage τ_1, all states in S have been observed at least once, and at least one state, s_0, has been observed more than once. Therefore, by stage τ_1 the decision rule d^*_m must be determined. Also we let Π_k denote P_1's distribution over ρ at stage τ_k.

Then we let the decision rule $d_k(\rho_0)$ be where P_1 plays d^*_m up to stage τ_k. After stage τ_k, this decision rule determines the utility maximising move at all future stages of the game, that P_1 would play if he knew ρ_0 with probability one, if the present state is contained in S. If the present state is not contained in S then this decision rule after stage τ_k is simply d^*_m, and it is defined for all values of $k = 1, 2, \ldots$. As a special case of this, we denote by $d(\rho_0)$ the decision rule $d_k(\rho_0)$ at $k = 0$.

The decision rule $d_k(\tilde{\rho}_k)$ also plays d^*_m up to stage τ_k. After stage τ_k it is the utility maximising decision rule, given that P_1 believes the true value of ρ to be equal to a fixed sequence $\{\tilde{\rho}_k, k = 1, 2, \ldots\}$ (with probability one) when the present state is contained in S, and playing d^*_m when the present state is not contained in S, for all values of $k = 1, 2, \ldots$. Then we also define another decision rule, $d^*_k(\Pi_1)$, to be the decision rule that again plays d^*_m up to stage τ_k. After stage τ_k it plays the utility maximising m–step back decision rule based on the distribution Π_1 determined at stage τ_1, but *not* on the move pairs after stage τ_1 and before stage τ_k.

We also define $\overline{U}(d)$ to be P_1's expected utility from a decision rule d, the expectation being taken across possible randomisations in d and over the termination time of the game, T. Having defined these we define one final decision rule, $d_k(0)$. $d_k(0)$ is defined to be a decision rule that is equal to d^*_m up to stage τ_k. After stage τ_k it is such that $\overline{U}(d_k(0)|T > \tau_j) \leq \overline{U}(d_k|T > \tau_j)$ for all values of j, and all possible decision rules d_k open to P_1 that are also equal to d^*_m up to stage τ_k.

8.2.2 Assumptions.

In order to prove the results in the next section we require a couple of assumptions. The first is to say that a player can always gain a higher utility from the future stages of the game if he knows ρ_0 with probability one, and the second requires P_1's utility function to satisfy a given regularity condition. Specifically these are

(8.1) At any stage of the game, τ_k, $k = 1, 2, \ldots$, we shall assume that there exists a value K^* and a vector ρ_0 such that

$$I(\mathbf{s}_0) = \min_{k:\tau_k > K^*} \{I_k(\mathbf{s}_0)\} > 0, \tag{8.2.1}$$

and $\mathsf{P}[T > \tau_k] > 0$ for all values of k such that $\tau_k > K^*$, where we define

$$I_k(\mathbf{s}_0) = \frac{\overline{U}(\mathbf{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathbf{d}_k^*(\Pi_1)|\rho = \rho_0, T > \tau_k)}{\lambda(\tau_k)} \tag{8.2.2}$$

given that we are in state $\mathbf{s}_0$ at stage τ_k, $k = 1, 2, \ldots$, where

$$\lambda(\tau_k) = \overline{U}(\mathbf{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathbf{d}_k(0)|\rho = \rho_0, T > \tau_k), \tag{8.2.3}$$

and where $\overline{U}(\mathbf{d})$, $\mathbf{d}_k(\rho_0)$, $\mathbf{d}_k^*(\Pi_1)$ and $\mathbf{d}_k(0)$ are all defined in subsection 8.2.1 above.

Note that this implies that for all k such that $\tau_k > K^*$, $\lambda(\tau_k) > 0$ as

$$\overline{U}(\mathbf{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) \geq \overline{U}(\mathbf{d}_k^*(\Pi_1)|\rho = \rho_0, T > \tau_k) \tag{8.2.4}$$

by the definitions of $\mathbf{d}_k(\rho_0)$ and $\mathbf{d}_k^*(\Pi_1)$.

Thus we are assuming that the maximum possible gain in utility from stage τ_k onwards (where $\tau_k > K^*$) from using the perfect information decision rule is strictly greater than the gain in utility from using the decision rule $\mathbf{d}_k^*(\Pi_1)$. That is to say that after stage K^*, there is no Bayes decision that does not depend on either ρ or P_1's distribution over the termination time, T.

(8.2) P_1's utility function is such that for any fixed sequence $\{\tilde{\rho}_k, k = 1, 2, \ldots\}$, and any $\epsilon > 0$, there exists a $\delta > 0$, and also a value, K, such that $\mathsf{P}[T > K] > 0$, then for any k where $\tau_k > K$,

$$\overline{U}(\mathbf{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathbf{d}_k(\tilde{\rho}_k)|\rho = \rho_0, T > \tau_k) \leq \lambda(\tau_k)\epsilon$$

$$\text{whenever} \quad \max_i |\rho_0^{(i)} - \tilde{\rho}_k^{(i)}| \leq \delta. \tag{8.2.5}$$

where $\overline{U}(\mathbf{d})$ and $\lambda(\tau_k)$ are defined above. Thus we are assuming P_1's utility function to be of an equi–continuous form, but only around the true value of ρ, i.e. we are not restricting P_1's utility function for values of ρ outside a δ neighbourhood of ρ_0.

These two assmptions preclude the following situations :

(i) if P_1 does not receive any information about P_2's previous moves, the problem degenerates into one of a simple maximisation over known values and hence an optimal strategy can be determined before the start of the game. This breaks Assumption (8.1) as, in this case, $\lambda(\tau_k) = 0$ for all values of k.

(ii) if P_1 will play a strategy belonging to a given set S_0 with probability one, where one strategy $\mathbf{s}^* \in S_0$ dominates all other strategies belonging to S_0 irrespective of the actual value of ρ, i.e.

$$\overline{U}(\mathbf{s}^*) \geq \overline{U}(\mathbf{s}) \text{ for all } \mathbf{s} \in S_0.$$

For example, if the pay–off matrix for P_1 were that given in Figure 8.2.1.

$$\begin{array}{cccc} & & \multicolumn{2}{c}{P_2} \\ & & 0 & 1 \\ P_1 & 0 & 1 & 1 \\ & 1 & 0 & 0 \end{array}$$

Figure 8.2.1

then obviously a utility maximising P_1 would choose move 0 at all stages of the game, for any utility function that is increasing in pay–off, whatever his beliefs about P_2 are. This breaks Assumption (8.1) as again $\lambda(\tau_k) = 0$ for all values of k.

(iii) if P_1 assumes that P_2's future behaviour for all time will be known to P_1, once P_1 has observed the first τ move pairs for some fixed value of τ. This breaks Assumption (8.1) as $I(\mathbf{s}_0) = 0$ for all values of k.

(iv) if P_1 knows that the game will terminate by a time τ, then the set of m–step back strategies for P_1, where $m = \tau$, obviously contains all possible strategies. This also breaks Assumption (8.1) as for any $\tau_k > \tau$, $I(\mathbf{s}_0) = 0$.

(v) if P_1's utility function is such that he obtains positive utility from the maximum possible pay–off at every stage of the game, and zero for any other pay–off. This violates Assumption (8.2) as for any $\delta > 0$, there is no K such that the difference in expected utility from the decision rules $\mathbf{d}_k(\rho_0)$ and $\mathbf{d}_k(\tilde{\rho}_k)$ is less than $\lambda(\tau_k)\epsilon$, for any $\epsilon > 0$.

8.2.3 Example 1.

For the vast majority of possible games, most models of P_2's behaviour that P_1 might choose will satisfy Assumptions (8.1) and (8.2). One of the simplest is a well known model of the

Prisoner's Dilemma game. A typical pay–off matrix for P_1 of a Prisoner's Dilemma game is given in Figure 8.2.2. Since the game is symmetric, the pay–off matrix for P_2 is just the transpose of this matrix.

$$\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 0 & 1 \end{array} \\ P_1 & \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 5 & 1 \\ 6 & 2 \end{pmatrix} \end{array}$$

Figure 8.2.2

We shall assume that at all stages of the game, P_1 knows all of the previous move sequence. Suppose that P_1 knows that P_2 is playing a 1–step back strategy such that P_2 mimics P_1's last move with probability p, but does not know this probability p, i.e. a partial Tit–For–Tat strategy with the value of p unknown to P_1, as in the previous chapter.

We shall show that this game, along with any utility function that is discounted linear (with discount factor λ, say) after a stage K^*, where $\mathsf{P}[T > K^*] > 0$, satisfies Assumptions (8.1) and (8.2).

Now P_1 cannot determine p with probability one at any stage of the game. Thus P_1's beliefs about P_2's probability of making any particular move at stage τ_1, are not precisely equal to p with probability one. Due to this uncertainty, a value of ρ_0 can be found such that the expected pay–off to $\mathrm{d}_k^*(\Pi_1)$ is less than the expected pay–off to $\mathrm{d}_k(\rho_0)$, for any value of $k = 1, 2, \ldots$. Therefore, for any stage $\tau_k > K^*$,

$$\overline{U}(\mathrm{d}(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathrm{d}_k^*(\Pi_1)|\rho = \rho_0, T > \tau_k) > 0 \qquad (8.2.6)$$

as the utility function is discounted linear after stage K^*. Therefore

$$\begin{aligned} \lambda(\tau_k) &= \overline{U}(\mathrm{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathrm{d}_k(0)|\rho = \rho_0, T > \tau_k) \\ &\geq \overline{U}(\mathrm{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathrm{d}_k^*(\Pi_1)|\rho = \rho_0, T > \tau_k) \\ &> 0 \end{aligned} \qquad (8.2.7)$$

Therefore, from equations (8.2.6) and (8.2.7), we have that

$$I(\mathbf{s}_0) = \min_{k:\tau_k > K^*} \{I_k(\mathbf{s}_0)\} > 0 \qquad (8.2.8)$$

satisfying Assumption (8.1).

Now, as the decision rules $d_k(\rho_0)$ and $d_k(0)$ are equal up to stage τ_k, for any k such that $\tau_k > K^*$, we can set

$$\begin{aligned}\lambda(\tau_k) &= c\lambda^{\tau_k} + c\lambda^{\tau_k+1} + \cdots + c\lambda^T \\ &< c\frac{\lambda^{\tau_k}}{1-\lambda}\end{aligned} \tag{8.2.9}$$

for some constant c.

Also, assume that there is a fixed sequence $\{\tilde{\rho}_k, k = 1,2,\ldots\}$ such that after some stage $\tau_k > K^*$,

$$|\rho_0 - \tilde{\rho}_k| \leq \delta \tag{8.2.10}$$

for some $\delta > 0$.

Now after stage K^*, the utility function is discounted linear and thus the maximum possible utility is found by calculating the maximum pay–off at all individual future stages. The maximum expected pay–off at stage v ($v > K^*$) is

$$\begin{aligned}E_v^1 &= m_v(5p_v + 1(1-p_v)) + (1-m_v)(6p_v + 2(1-p_v)) \\ &= (4p_v + 2) - m_v\end{aligned} \tag{8.2.11}$$

where m_v is P_1's pay–off maximising probability of playing move 0 given his beliefs about p_v, i.e. the probability of P_2 playing move 0 at stage v. Now if P_1 knows for certain that P_2 will play move 0 with probability p_0 at stage v, then the maximum expected pay–off at stage v ($v > K^*$) is

$$\begin{aligned}E_v^0 &= (4p_0 + 2) - m_v \\ \Rightarrow |E_v^0 - E_v^1| &= 4|p_0 - p_v| \\ &\leq 4\delta \text{ by } (8.2.10)\end{aligned} \tag{8.2.12}$$

for some $\delta > 0$. So we can find an $\epsilon' > 0$ such that $4\delta < \epsilon'$ and thus

$$\begin{aligned}&\overline{U}(d_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(d_k(\tilde{\rho}_k)|\rho = \rho_0, T > \tau_k) \\ &< \lambda^{\tau_k}\epsilon' + \lambda^{\tau_k+1}\epsilon' + \cdots + \lambda^T\epsilon' \\ &< \frac{\lambda^{\tau_k}}{1-\lambda}\epsilon' \\ &= \lambda(\tau_k)\epsilon\end{aligned} \tag{8.2.13}$$

for $\epsilon = c\epsilon'$ where c is a constant defined in inequality (8.2.9) above. Therefore Assumption (8.2) holds.

It is straightforward to check that the two assumptions also hold for more elaborate models of the Prisoner's Dilemma game when the utility function has any reasonable non–linear form, and where P_1's model of P_2 assumes that P_2 is employing an n–step back strategy. Also many games other than PDGs can be seen to satisfy Assumptions (8.1) and (8.2), for a wide variety of different forms of utility functions, but we concentrate on the PDG here as it provides an interesting example.

8.3 Results.

We now prove the main result of this chapter. This is that if P_1 makes the assumption that P_2 is playing some unknown n–step back strategy, then as a utility maximising player, P_1 should not play any m–step back strategy, for any value of m. We prove this by means of three lemmas.

LEMMA 8.3.1. *Let $\hat{\rho}_k$ be the sample proportions of moves that P_2 has been observed to make after each sequence $\mathbf{s} \in \mathcal{S}$, before stage τ_k. Then, for any $\delta > 0$,*

$$\mathsf{P}\left[\max_i |\rho_0^{(i)} - \hat{\rho}_k^{(i)}| > \delta\right] \to 0 \text{ as } k \to \infty. \tag{8.3.1}$$

PROOF: From the definition of $\mathcal{S}$ (in subsection 8.2.1), we need only consider states belonging to $\mathcal{S}$. As $\mathcal{S}$ is a finite set, we can denote the number of states in $\mathcal{S}$ by $f \leq 2^{2s}$. Let $n_k^{(i)}$ be the number of observations P_1 has made of $\rho^{(i)}$ up to stage τ_k.

Now, by definition, $\hat{\rho}_k^{(i)}$ is the sample proportion from a Binomial experiment with mean $\rho_0^{(i)}$ and variance $\dfrac{\rho_0^{(i)}(1-\rho_0^{(i)})}{n_k^{(i)}}$.

Hence, for all $\delta > 0$,

$$\begin{aligned} \mathsf{P}\left[\max_i |\rho_0^{(i)} - \hat{\rho}_k^{(i)}| > \delta\right] &\equiv \mathsf{P}\left[\bigcup_{i=1}^{f} \left\{|\rho_0^{(i)} - \hat{\rho}_k^{(i)}| > \delta\right\}\right], \\ &\leq \sum_{i=1}^{f} \mathsf{P}\left[|\rho_0^{(i)} - \hat{\rho}_k^{(i)}| > \delta\right], \end{aligned}$$

which by the Chebyshev inequality,

$$< \sum_{i=1}^{f} \frac{f}{\delta^2} \cdot \frac{\rho_0^{(i)}(1-\rho_0^{(i)})}{n_k^{(i)}},$$

which as $q(1-q) \leq \frac{1}{4}$ if $q \in [0,1]$,

$$\leq \frac{f}{4\delta^2} \sum_{i=1}^{f} \frac{1}{n_k^{(i)}}. \tag{8.3.2}$$

However, $n_k^{(i)} \to \infty$ as $k \to \infty$ for all i, by the definition of S,

$$\Rightarrow \frac{f}{4\delta^2} \sum_{i=1}^{f} \frac{1}{n_k^{(i)}} \to 0 \text{ as } k \to \infty. \tag{8.3.3}$$

which gives the result. □

So from this we can see that if P_1 observes P_2's play for long enough, the sample proportion of times that P_2 is observed to play move 0, will converge to the true probability ρ_0.

LEMMA 8.3.2. *If, for some fixed sequence $\{\tilde{\rho}_k, k = 1, 2, \ldots\}$, for all $\delta > 0$ there exists a $\delta' > 0$ and a K' such that for all $k > K'$,*

$$\mathsf{P}\left[\max_i |\rho_0^{(i)} - \tilde{\rho}_k^{(i)}| > \delta\right] < \delta' \tag{8.3.4}$$

then, under Assumptions (8.1) and (8.2), for any $\epsilon > 0$, there exists a value, K, such that $\mathsf{P}[T > K] > 0$, and for all k where $\tau_k > K$,

$$\overline{U}(\mathrm{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathrm{d}_k(\tilde{\rho}_k)|\rho = \rho_0, T > \tau_k) < \lambda(\tau_k)\epsilon. \tag{8.3.5}$$

PROOF: At a stage of the game τ_k, define the event E_k to be

$$\max_i |\rho_0^{(i)} - \tilde{\rho}_k^{(i)}| \leq \delta. \tag{8.3.6}$$

Also, let D_k be the events that $(\rho = \rho_0, T > \tau_k)$. Now, by assumption, $\tilde{\rho}_k$ converges in probability to ρ_0 irrespective of the value of ρ_0, and depends on the game not already having finished, so we can define $\mathsf{P}[E_k|D_k] = \eta_k$. Then,

$$\overline{U}(\mathrm{d}_k(\rho_0)|D_k) - \overline{U}(\mathrm{d}_k(\tilde{\rho}_k)|D_k) = \eta_k A + (1 - \eta_k)B \tag{8.3.7}$$

where

$$\begin{aligned} A &= \overline{U}(\mathrm{d}_k(\rho_0)|D_k, E_k) - \overline{U}(\mathrm{d}_k(\tilde{\rho}_k)|D_k, E_k) \\ B &= \overline{U}(\mathrm{d}_k(\rho_0)|D_k, \overline{E}_k) - \overline{U}(\mathrm{d}_k(\tilde{\rho}_k)|D_k, \overline{E}_k) \end{aligned} \tag{8.3.8}$$

and $\overline{E}_k$ is the compliment of the event E_k.

Now, by Assumption (8.2), for every $\epsilon' > 0$ there exists a value, K, such that $\mathsf{P}[T > K] > 0$, and also that for all k where $\tau_k > K$, there exists a $\delta > 0$ such that given that the event E_k occurs,

$$\begin{aligned}\left|\overline{U}(\mathrm{d}_k(\rho_0)|D_k) - \overline{U}(\mathrm{d}_k(\tilde{\rho}_k)|D_k)\right| &< \lambda(\tau_k)\epsilon' \\ \Rightarrow A &< \lambda(\tau_k)\epsilon' \end{aligned} \tag{8.3.9}$$

Also, by Assumption (8.1), and the definition of $\mathrm{d}_k(0)$,

$$\begin{aligned} B &\le \overline{U}(\mathrm{d}_k(\rho_0)|D_k, E_k) - \overline{U}(\mathrm{d}_k(0)|D_k, E_k) \\ &= \lambda(\tau_k). \end{aligned} \tag{8.3.10}$$

Therefore, from equation (8.3.7) we have that

$$\begin{aligned}\overline{U}(\mathrm{d}_k(\rho_0)|D_k) - \overline{U}(\mathrm{d}_k(\tilde{\rho}_k)|D_k) &= \eta_k A + (1-\eta_k)B \\ &< \eta_k\lambda(\tau_k)\epsilon' + (1-\eta_k)\lambda(\tau_k) \\ &= \lambda(\tau_k)[\eta_k\epsilon' + (1-\eta_k)]. \end{aligned} \tag{8.3.11}$$

However, for any $\delta > 0$, $\eta_k \to 1$ as $k \to \infty$ for any value of ρ_0, by our assumption. So we can always find a value of ϵ' such that for any $\epsilon > 0$, there is a K such that $\mathsf{P}[T > K] > 0$, and also that for all k where $\tau_k > K$,

$$\eta_k\epsilon' + (1-\eta_k) < \epsilon. \tag{8.3.12}$$ □

Therefore we can deduce from the two assumptions in the previous section that when the sequence $\tilde{\rho}_k$ converges in probability to the true value ρ_0, the difference in expected utility from using a decision rule based on $\tilde{\rho}_k$ rather than ρ_0 is bounded above.

LEMMA 8.3.3. *Under Assumption (8.1), for a state $s_0 \in S$, there exists a K^* and a ρ_0 such that for any value of k where $\tau_k > K^*$,*

$$\overline{U}(\mathrm{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathrm{d}^*_m|\rho = \rho_0, T > \tau_k) \ge \lambda(\tau_k)I(s_0) > 0 \tag{8.3.13}$$

where K^*, $\lambda(\tau_k)$ and $I(s_0)$ are defined in Assumption (8.1).

PROOF: We have, by the definition of $\mathrm{d}^*_{k+1}(\Pi_1)$, for any $k > 0$,

$$\begin{aligned}\overline{U}(\mathrm{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathrm{d}^*_{k+1}(\Pi_1)|\rho = \rho_0, T > \tau_k) \ge \\ \overline{U}(\mathrm{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathrm{d}^*_k(\Pi_1)|\rho = \rho_0, T > \tau_k). \end{aligned} \tag{8.3.14}$$

So, dividing inequality (8.3.14) by $\lambda(\tau_k)$, and using Assumption (8.1), we have that for any k such that $\tau_k > K^*$,

$$\frac{\overline{U}(\mathbf{d}_k(\rho_0)|\rho=\rho_0,T>\tau_k)-\overline{U}(\mathbf{d}^*_{k+1}(\Pi_1)|\rho=\rho_0,T>\tau_k)}{\lambda(\tau_k)} \geq I_k(\mathbf{s}_0)$$

$$\begin{aligned}\Rightarrow \overline{U}(\mathbf{d}_k(\rho_0)|\rho=\rho_0,T>\tau_k)-\overline{U}(\mathbf{d}^*_{k+1}(\Pi_1)|\rho=\rho_0,T>\tau_k) &\geq \lambda(\tau_k)I_k(\mathbf{s}_0)\\ &\geq \lambda(\tau_k)I(\mathbf{s}_0).\end{aligned} \tag{8.3.15}$$

Now let $\{\mathbf{d}_{k+1}(\Pi_1)\}$ denote the set of decision rules, such that any decision rule belonging to $\{\mathbf{d}_{k+1}(\Pi_1)\}$ is equal to $\mathbf{d}^*_m$ up to stage τ_{k+1} and after which is an m–step back decision rule based only upon P_1's prior information about P_2's future moves, *and* information gained up to stage τ_1.

Clearly $\mathbf{d}^*_m \in \{\mathbf{d}_{k+1}(\Pi_1)\}$ and so for any value of k,

$$\overline{U}(\mathbf{d}^*_m|\rho=\rho_0,T>\tau_k) \leq \overline{U}(\mathbf{d}^*_{k+1}(\Pi_1)|\rho=\rho_0,T>\tau_k) \tag{8.3.16}$$

as by the definition of the decision rule, $\mathbf{d}^*_{k+1}(\Pi_1)$ is the utility maximising decision rule belonging to $\{\mathbf{d}_{k+1}(\Pi_1)\}$. Hence,

$$\begin{aligned}&\overline{U}(\mathbf{d}_k(\rho_0)|\rho=\rho_0,T>\tau_k)-\overline{U}(\mathbf{d}^*_m|\rho=\rho_0,T>\tau_k) \geq\\ &\qquad \overline{U}(\mathbf{d}_k(\rho_0)|\rho=\rho_0,T>\tau_k)-\overline{U}(\mathbf{d}^*_{k+1}(\Pi_1)|\rho=\rho_0,T>\tau_k)\end{aligned} \tag{8.3.17}$$

and so, by (8.3.15), for any k such that $\tau_k > K^*$,

$$\Rightarrow \overline{U}(\mathbf{d}_k(\rho_0)|\rho=\rho_0,T>\tau_k)-\overline{U}(\mathbf{d}^*_m|\rho=\rho_0,T>\tau_k) \geq \lambda(\tau_k)I(\mathbf{s}_0) > 0. \tag{8.3.18}$$ □

So it is clear that there is always a stage in the game, K^*, such that adopting the utility maximising decision rule at any stage after K^* will attain a utility higher than simply maintaining the utility maximising m–step back strategy $\mathbf{d}^*_m$. Now we have all that we require to prove the main theorem.

THEOREM 8.3.4. *Suppose that P_1 knows that P_2 is playing an n–step back strategy, for any n, but the strategy is unknown. Then, under Assumptions (8.1) and (8.2) above, any Bayes strategy for P_1 is not an m–step back strategy, for any m.*

PROOF: We shall prove this by contradiction. We shall assume that there is a Bayes strategy that is an m–step back strategy, for some m. Then we must have

$$\overline{U}(\mathbf{d}_m^*) \geq \overline{U}(\mathbf{d}) \tag{8.3.19}$$

for all strategies $\mathbf{d}$ that are open to P_1, as $\mathbf{d}_m^*$ is the utility maximising m–step back strategy.

Now, from Lemma 8.3.1 we have that for any $\delta > 0$,

$$\mathsf{P}\left[\max_i |\rho_0^{(i)} - \hat{\rho}_k^{(i)}| > \delta\right] \to 0 \text{ as } k \to \infty. \tag{8.3.20}$$

Then, Lemma 8.3.2 gives us that for any $\epsilon > 0$ there exists a K such that for all k where $\tau_k > K$,

$$\overline{U}(\mathbf{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathbf{d}_k(\hat{\rho}_k)|\rho = \rho_0, T > \tau_k) < \lambda(\tau_k)\epsilon. \tag{8.3.21}$$

So, in particular, we can find a K such that for any k where $\tau_k > K$,

$$\overline{U}(\mathbf{d}_k(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathbf{d}_k(\hat{\rho}_k)|\rho = \rho_0, T > \tau_k) < \lambda(\tau_k)\frac{I(\mathbf{s}_0)}{2} \tag{8.3.22}$$

where $I(\mathbf{s}_0)$ is as defined in Assumption (8.1).

Also, by Lemma 8.3.3, we have that for a state $\mathbf{s}_0 \in S$, there exists a K^* and a ρ_0 such that for any k where $\tau_k > K^*$,

$$\overline{U}(\mathbf{d}(\rho_0)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathbf{d}_m^*|\rho = \rho_0, T > \tau_k) \geq \lambda(\tau_k)I(\mathbf{s}_0) > 0 \tag{8.3.23}$$

and so by (8.3.22), for any k such that $\tau_k > \max\{K, K^*\}$,

$$\overline{U}(\mathbf{d}_k(\hat{\rho}_k)|\rho = \rho_0, T > \tau_k) - \overline{U}(\mathbf{d}_m^*|\rho = \rho_0, T > \tau_k) > \lambda(\tau_k)\frac{I(\mathbf{s}_0)}{2} > 0. \tag{8.3.24}$$

Now, by Assumption (8.1), $\mathsf{P}[T > K^*] > 0$. Also, by Assumption (8.2), $\mathsf{P}[T > K] > 0$. Therefore, let

$$\mathsf{P}[T > \max\{K, K^*\}] = \alpha > 0. \tag{8.3.25}$$

Now,

$$\overline{U}(\mathbf{d}_k(\hat{\rho}_k)|\rho = \rho_0, T \leq \max\{K, K^*\}) - \overline{U}(\mathbf{d}_m^*|\rho = \rho_0, T \leq \max\{K, K^*\}) = 0 \tag{8.3.26}$$

as the decision rules $\mathrm{d}_k(\hat{\rho}_k)$ and d_m^* are equal up to stage τ_k by definition. Also,

$$\overline{U}(\mathrm{d}_k(\hat{\rho}_k)|\rho = \rho_0, T > \max\{K, K^*\}) - \overline{U}(\mathrm{d}_m^*|\rho = \rho_0, T > \max\{K, K^*\}) > \lambda(\tau_k)\frac{I(\mathbf{s}_0)}{2} \tag{8.3.27}$$

from equation (8.3.24). Therefore, from equations (8.3.26) and (8.3.27) we have that for any value of k, such that $\tau_k > \max\{K, K^*\}$,

$$\begin{aligned}\overline{U}(\mathrm{d}_k(\hat{\rho}_k)|\rho = \rho_0) - \overline{U}(\mathrm{d}_m^*|\rho = \rho_0) &> (1-\alpha)0 + \alpha\lambda(\tau_k)\frac{I(\mathbf{s}_0)}{2}\\ &= \alpha\lambda(\tau_k)\frac{I(\mathbf{s}_0)}{2},\end{aligned} \tag{8.3.28}$$

for some constant $\alpha > 0$. This is contrary to our assumption that

$$\overline{U}(\mathrm{d}_m^*) \geq \overline{U}(\mathrm{d}) \tag{8.3.29}$$

for any strategy d open to P_1, giving a contradiction. Hence d_m^* is not a Bayes decision rule, and therefore no other m–step back decision rule is a Bayes decision rule. So it is suboptimal for P_1 to play any m–step back decision rule. □

This proves that it is always possible for P_1 to construct a decision rule that achieves a higher utility than would have been achieved from any m–step back decision rule, given that P_2 is playing an n–step back decision rule. In the next section we see that this result generalises to other types of games, and we consider the effect of making an assumption about the opponent's rationality.

8.4 Extensions and Implications.

8.4.1 Extensions of the theorem.

We begin this section with two corollaries to the theorem of section 8.3. First we drop the assumption that the game is a binary one, i.e. at any stage of the game, either player has w moves available to him, for some finite $w = 2, 3, \ldots$.

COROLLARY 8.4.1. *Under the conditions of Theorem 8.3.4, for any 2 player, non–cooperative sequential game, any Bayes strategy for P_1 is not an m–step back strategy, for any m.*

PROOF: In Lemma 8.3.1 we need to define

$$\rho^{(i)} = \left(\rho^{(i,1)}, \rho^{(i,2)}, \ldots, \rho^{(i,w)}\right) \tag{8.4.1}$$

where $\rho^{(i,j)}$ is the probability that P_2 will play move j ($j = 2, 3, \ldots, w$) when the game is in state i ($i = 1, 2, \ldots, w^{2s}$). Also, let $\hat{\rho}_k^{(i)}$ be the sample proportion of moves that P_2 has been observed to make when in state i, before stage τ_k. Then $\hat{\rho}_k^{(i)}$ is the sample proportion from a Multinomial experiment. By the same argument as that of Lemma 8.3.1, we obtain that $\hat{\rho}_k^{(i)}$ converges to the true vector $\rho_0^{(i)}$, i.e. for some $\delta > 0$,

$$\mathsf{P}[\max_i\{\max_j |\rho_0^{(i,j)} - \hat{\rho}_k^{(i,j)}|\} > \delta] \to 0 \quad \text{as} \quad k \to \infty. \tag{8.4.2}$$

It is easily seen that Lemmas 8.3.2 and 8.3.3 hold in this situation, and thus the theorem holds for non-binary, 2 player sequential games. □

So we can see from this that in any 2 player games such that the Assumptions (8.1) and (8.2) are satisfied, Theorem 8.3.4 will hold. Now we show that we do not require P_1 to believe P_2 to be playing an n-step back strategy throughout the game.

COROLLARY 8.4.2. *Suppose there is some stage of the game, K', where* $\mathsf{P}[T > K'] \neq 0$, *and after which P_1 believes that P_2 will play an n-step back strategy with probability one, but the exact strategy is unknown. Then, under Assumptions* (8.1) *and* (8.2) *above, any Bayes strategy is not an m-step back strategy, for any m.*

PROOF: Assume that there is a Bayes strategy that is an m-step back strategy for some m. Now $\mathsf{P}[T > K'] \neq 0$, so we can define

$$\mathsf{P}[T > K, K^*, K'] = \beta > 0 \tag{8.4.3}$$

where K and K^* are defined in the proof of the theorem. To do this we need to define τ_0 to be the first stage of the game when $\mathbf{s}_0$ is achieved for the first time after stage K', under the action of d_m^*. It should be noted that this change in definition makes Assumption (8.1) a stronger assumption.

We can now conclude that for any k such that $\tau_k > \max\{K, K^*, K'\}$,

$$\overline{U}(\mathrm{d}_k(\hat{\rho}_k)|\rho = \rho_0, T \leq \max\{K, K^*, K'\}) - \overline{U}(\mathrm{d}_m^*|T \leq \max\{K, K^*, K'\}) = 0 \tag{8.4.4}$$

and also

$$\overline{U}(\mathrm{d}_k(\hat{\rho}_k)|\rho = \rho_0, T > \max\{K, K^*, K'\}) - \overline{U}(\mathrm{d}_m^*|T > \max\{K, K^*, K'\}) > \beta\lambda(\tau_k)\frac{I(\mathbf{s}_0)}{2} \tag{8.4.5}$$

by the same argument as the one given in the proof of the theorem. This gives a contradiction to the assumption that any m–step back strategy is a Bayes decision rule, and the corollary is proved. □

Therefore, provided P_1 believes P_2 will employ an n–step back strategy at some stage in the game, it will be suboptimal for P_1 to employ any m–step back strategy himself.

In chapter 6 we considered Bayes–calibrated games. As we showed in the theorem above, P_1 is assuming that P_2 is employing a strategy that P_1 would not consider playing if the roles were reversed. Under the assumption of mutual calibration, P_1's model would require P_2 to believe that P_1 is playing an m–step back strategy, thus implying by the theorem above that P_2 would not play an n–step back strategy himself. So, if P_1 has a model that assumes P_2 to be playing an n–step back strategy, the game is not Bayes–calibrated. The question then arises as to why a player should adopt such a model. If he has beliefs consistent with the lack of calibration, then this may be reasonable. However, in the context of symmetric experimental games this would seem a very dubious assumption, especially when the players are drawn from a homogeneous population.

On the other hand, models do exist which admit mutual calibration between two utility maximising players — for instance, consider the following example.

8.4.2 Example 2.

As we have shown that it is not possible to construct a mutually calibrated model on the basis of m–step back strategies, in this example we consider a game where both players believe each other to be using a strategy from a particular set of strategies that are not m–step back. P_1 can then use his beliefs about P_2's strategy to deduce his utility maximising strategy from this set.

Consider the symmetric game where the pay–off to P_1 is given by the pay–off matrix given in Figure 8.4.1.

$$\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 0 & \;\;1 \end{array} \\ P_1 & \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix} \end{array}$$

Figure 8.4.1

This is another Prisoner's Dilemma game with move 0 being the cooperation move, and move 1 being the defection move. We shall assume that both players are using a decision rule of the form

$$\delta_q = \begin{cases} \text{TFT} & \text{up to, and including, stage } q \\ \text{Continual defection} & \text{after stage q} \end{cases}$$

for some value of q, with the proviso that the decision rule begins to play the continual defection strategy as soon as the opponent adopts the continual defection strategy (i.e. on the following stage of the game). Both players also believe that their opponent is playing a strategy of the form δ_q, but the value of q is unknown.

Suppose that P_1's utility function is

$$U_\theta(x) = 1 - e^{-\theta x} \qquad \text{for some } \theta > 0 \tag{8.4.6}$$

where x is the aggregated pay-off from the whole game to P_1. Suppose further that P_1 believes P_2 to be playing strategy δ_q with probability

$$\Pi(\delta_q) = \frac{e^{-\lambda}\lambda^q}{q!} \quad \text{for some } \lambda > 0. \tag{8.4.7}$$

We wish to show that for any values of θ and λ there is a value r^*, such that for any $r < r^*$ P_1 should prefer δ_r to δ_{r-1} (i.e. should continue playing TFT), and for any $r \geq r^*$, P_1 should prefer δ_r to δ_{r+1} (i.e. should defect). So we require a unique value of r^* where it is optimal to change from the TFT strategy to the continual defection strategy.

At stage $r-1$, we can assume that P_2 has played TFT on every move so far, otherwise δ_r and δ_{r-1} are the same and continual defection will remain the utility maximising strategy for the rest of the game. Let $\overline{U}(\delta_r)$ be the expected utility to P_1 from employing decision rule δ_r, and define

$$D = \overline{U}(\delta_r) - \overline{U}(\delta_{r-1}) \tag{8.4.8}$$

Let $\Pi^*(\delta_{r+1}) = \sum_{i=r+1}^{\infty} \Pi(\delta_i)$, then by expanding the right hand side of equation (8.4.8) we obtain

$$\begin{aligned} D &= [U_\theta(r-2)\Pi(\delta_{r-1}) + U_\theta(r)\Pi(\delta_r) + U_\theta(r+2)\Pi^*(\delta_{r+1})] \\ &\quad - [U_\theta(r-1)\Pi(\delta_{r-1}) + U_\theta(r+1)\Pi(\delta_r) + U_\theta(r+1)\Pi^*(\delta_{r+1})] \\ &= \Pi^*(\delta_{r+1}) - e^\theta\Pi(\delta_r) - e^{3\theta}\Pi(\delta_{r-1}). \end{aligned} \tag{8.4.9}$$

Therefore P_1 should prefer δ_r to δ_{r-1} if and only if

$$\Pi^*(\delta_{r+1}) > e^{\theta}\Pi(\delta_r) - e^{3\theta}\Pi(\delta_{r-1}) \tag{8.4.10}$$

and prefer δ_{r-1} to δ_r if and only if

$$\Pi^*(\delta_{r+1}) < e^{\theta}\Pi(\delta_r) - e^{3\theta}\Pi(\delta_{r-1}) \tag{8.4.11}$$

Now $\Pi(\delta_r) = \dfrac{e^{-\lambda}\lambda^r}{r!}$, $\Pi(\delta_{r-1}) = \dfrac{e^{-\lambda}\lambda^{r-1}}{(r-1)!}$, and

$$\Pi^*(\delta_{r+1}) = \frac{e^{-\lambda}\lambda^{r+1}}{(r+1)!}\left[1 + \frac{\lambda}{r+2} + \frac{\lambda^2}{(r+2)(r+3)} + \ldots\right]. \tag{8.4.12}$$

From these we can see that inequality (8.4.10) holds if and only if

$$\lambda^2 + \frac{\lambda^3}{(r+2)} + \frac{\lambda^4}{(r+2)(r+3)} + \cdots > (r+1)e^{\theta}\lambda + r(r+1)e^{3\theta}. \tag{8.4.13}$$

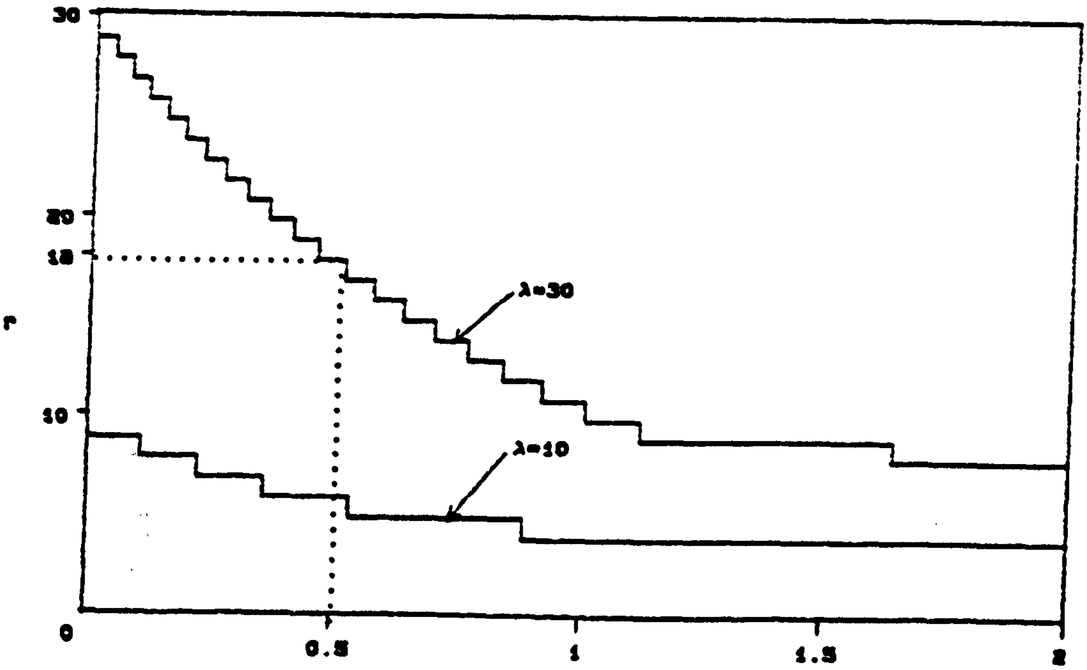

Figure 8.4.2

For λ large enough and θ small enough there will be values of r such that this inequality will hold and it will therefore be optimal for P_1 to continue playing TFT until this inequality is violated. Also, inequality (8.4.11) holds if and only if

$$\lambda^2 + \frac{\lambda^3}{(r+2)} + \frac{\lambda^4}{(r+2)(r+3)} + \cdots < (r+1)e^{\theta}\lambda + r(r+1)e^{3\theta}. \qquad (8.4.14)$$

For all values of λ and θ there will always be a value of r such that inequality (8.4.14) will hold. In particular, if λ and θ are such that there is a set of values of r such that inequality (8.4.10) holds, then for these same values of λ and θ, inequality (8.4.11) will hold, and a utility maximising value r^* can be found.

Figure 8.4.2 shows the values of r^* for varying θ, in the particular cases where $\lambda = 10$ and $\lambda = 30$. In this figure, all points that lie below the respective line indicate a value of r where P_1 should continue playing TFT, and if r is on or above the line, P_1 should defect. For example, if $\lambda = 30$ and $\theta = 0.5$, then $r^* = 18$, i.e. if $r < 18$ then P_1 should continue playing TFT, and if $r \geq 18$ then P_1 should defect.

If P_1 is calibrated he will believe P_2 to have a similar model of his (P_1's) strategies and utility, as he (P_1) has about P_2's. If this is so, then P_2 will adopt a very similar strategy to P_1. In the case where P_2 has exactly the same beliefs about the parameters as P_1, both players will start to defect at the same stage. Therefore, P_1 believes that P_2 will play the same as P_1 will. So, if P_2 actually thinks about the game in the same way as P_1 does, then a utility maximising outcome will result.

Grofman & Pool (1975) consider optimal play in a similar Prisoner's Dilemma game when P_2 is known to be playing a partial Tit–For–Tat strategy (i.e. mimicking P_1's previous move with probability p), and the utility function is linear on pay–offs. In the two computer tournaments that Axelrod (1984) ran in order to determine an effective strategy for such Prisoner's Dilemma games, one participant in the first tournament and two participants in the second tournament employed strategies of the type Grofman and Pool considered. These strategies were based on the dubious assumption that every opponent was playing a 1–step back strategy, i.e. playing suboptimally under the model assumptions given above. It should not be surprising to learn in the light of this chapter, that all three finished in the bottom half of the participants in their respective tournaments.

The winner of both of these tournaments was TFT (i.e. partial TFT with probability $p = 1$). The reason that this rule could be so successful was because it is a degenerate and well–known strategy from the class of 1–step back strategies. It can be shown that plausible models for P_1 which allow P_2 to play TFT exist, with the property that TFT is the optimal strategy for P_1, because in these models the assumptions of Section 8.2.2 are broken.

8.4.3 Implications of dropping the calibration assumption.

There are obviously going to be many situations where the calibration hypothesis above is not appropriate. However, the next result suggests that a player should not consider non–degenerate, m–step back strategies even if he does not believe calibration, as defined above, to be a sensible modelling assumption.

COROLLARY 8.4.3. *Let ρ'_k be P_1's probability of P_2 playing an n–step back strategy at stage k of the game. Suppose that $\rho'_k > 0$ for all stages k, and that under P_1's model of P_2, ρ'_k converges to 1 with probability $\rho' > 0$, and ρ'_k converges to 0 with probability $(1 - \rho')$ as $k \to \infty$. Then under Assumptions* (8.1) *and* (8.2) *above, any Bayes strategy for P_1 is not an m–step back strategy, for any m.*

PROOF: Again we assume that $\mathbf{d}^*_m$ is a utility maximising strategy for P_1, and then find a contradiction. Now, by the assumption of the corollary, for any $\epsilon > 0$, there exists a K' such that for any $k > K'$

$$\rho'_k > 1 - \epsilon \qquad \text{or} \qquad \rho'_k < \epsilon. \tag{8.4.15}$$

Define a strategy $\mathbf{d}_\epsilon(\rho')$ to be equal to $\mathbf{d}^*_m$ up to stage K' for some value of ϵ. Then for $k > K'$, if $\rho_k < \epsilon$ define $\mathbf{d}_\epsilon(\rho')$ to be equal to $\mathbf{d}^*_m$ throughout the game. If, for $k > K'$, $\rho_k > 1-\epsilon$ then define $\mathbf{d}_\epsilon(\rho')$ to be equal to $\mathbf{d}_k(\hat{\rho}_k)$ from stage K' onwards. Thus, from Theorem 8.3.4,

$$\begin{aligned}\overline{U}(\mathbf{d}_\epsilon(\rho')|T > \tau_k) - \overline{U}(\mathbf{d}^*_m|T > \tau_k) &> \rho'\left[(1-\epsilon)\alpha\lambda(\tau_k)\frac{I(\mathbf{s}_0)}{2} - \epsilon\lambda(\tau_k)\right] + (1-\rho')\,[0] \\ &= \rho'\lambda(\tau_k)\left[\frac{\alpha I(\mathbf{s}_0)}{2}(1-\epsilon) - \epsilon\right]\end{aligned} \tag{8.4.16}$$

for α defined in the proof of the theorem.

So, as we can always find an ϵ such that

$$\rho'\lambda(\tau_k)\left[\frac{\alpha I(\mathbf{s}_0)}{2}(1-\epsilon) - \epsilon\right] > 0, \tag{8.4.17}$$

there exists a strategy that obtains a higher expected utility than the utility maximising m–step back decision rule. Hence, we have found a contradiction, and so any Bayes strategy for P_1 is not an m–step back strategy. □

A simple example of when this corollary holds is when, under P_1's model, P_1 can determine whether P_2 is employing an n–step back strategy or not, before a given time point. On the other hand, the corollary does not hold if, for example, it is assumed that P_2 believes that if he deviates from an n–step back strategy at any stage of the game, P_1 will be provoked to play a strategy that is unfavourable to P_2. However, this would be an unlikely and paranoic model in most situations.

Some of the work in this chapter has previously been reported in Young & Smith (1988a).

9. FURTHER RESEARCH

In this chapter we shall discuss some areas that I believe are worthy of some further research. The points stem from the work in the previous chapters, and are extensions to the theory presented there. The areas are given as individual directions down which further research may be fruitful, and can be considered independently, although some areas are related to others.

1. Repeated Asymmetric Games.

The models and methods considered in the previous chapters of this thesis have been concerned mainly with symmetric games, or at least games where the players know all moves that are available to all players. Research into games where the symmetry does not hold and players might have 'unthought of' actions available may well produce interesting results. Bennett (1977) and Bennett & Huxham (1982) develop a theory that is designed to allow for the possibility of the players having differing perceptions of the game situation, and called this a theory of *hypergames.*

Bennett (1977) defined a hypergame to be a system consisting of (a) the players, (b) the strategies available to player p, as perceived by player q, and (c) the ordering of the outcomes to the player p, as perceived by player q. By developing this theory on each player's perception about the game being played, the effects of differing perceptions can be analysed. The theory requires a more complicated representation, as each player has his own perceived pay-off matrix, and these pay-off matrices are joined by 'link' functions that describe the perceived association of moves. Bennett & Huxham (1982) present this theory of hypergames as an aid to understanding a particular problem, rather than as a solution to it. This is facilitated by a preliminary problem structuring phase, which then leads to a formal model building and analysis phase. Bennett (1985) shows how the hypergame methodology relates to different approaches in decision analysis. After developing the theory, several case studies were considered to demonstrate its use, e.g. Bennett, Dando & Sharp (1980), Bennett & Dando (1979, 1982) and Bennett, Huxham & Dando (1981).

However, several problems arise from the form of the hypergames presented in the above papers. First of all there is an implicit assumption that the opponent orders the outcomes of the game in a given manner. The player (P_1) is then assumed to play the game, believing that his opponent (P_2) will play in a manner consistent with this ordering *with probability one*. The methodology does not permit P_1 to incorporate any degree of uncertainty into the model, and no updating of beliefs is permitted. A second problem is that only the relative orderings of the pay–offs are considered, as opposed to a realistic pay–off structure. Therefore only weak stability conditions can be determined for the game, rather than specific optimal strategies. This can be seen to be a problem by considering Figure 9.1.

$$
\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 1 & 2 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 10 & -1 \\ 2 & 8 \end{pmatrix} \end{array}
\qquad\qquad
\begin{array}{ccc} & & P_2 \\ & & \begin{array}{cc} 1 & 2 \end{array} \\ P_1 & \begin{array}{c} 1 \\ 2 \end{array} & \begin{pmatrix} 10 & -1000000 \\ 2 & 8 \end{pmatrix} \end{array}
$$

Figure 9.1(a) Figure 9.1(b)

Whilst the pay–offs in the two games in Figure 9.1 have the same ordering, it would seem reasonable to assume that P_1 might approach the games differently.

A third problem is that the methodology applies to single one–off situations, but the case studies (for example Bennett, Dando & Sharp, 1980) considered are actually repeated situations, and part of an on–going process. If the theory is adapted to permit repeated plays of the game, then the previous problems are made much worse, and more considerations must be taken into account to allow for the dynamic nature of the game. If only one–off situations are considered, then only how to play in particular circumstances can be considered, not an overall strategy. A fourth problem is the timing of actions in the model. The models permit actions occuring at different times, but the games are modelled as single play games and therefore important details are being overlooked. This problem would be overcome if the games were modelled as multi–stage games.

I believe that a Bayesian model of asymmetric games could be developed as a generalisation of the Bennett hypergame methodology, that overcomes the above problems. As this model will be applicable to repeated games as well as one–play games, we will require it to be able to react to any changes quickly. These changes could be in the the game being played, changes in the actions of the opponents, or changes in the setting of the game (and therefore affecting the player himself). In line with the models that we have considered in earlier chapters, we shall wish to determine normative inferences from this model, as opposed to positive. However we would like the model to learn from how people *do* play the game in order to produce these normative inferences.

Two further essential requirements of this model are that it is easy to use and that it is easy to interpret. I would hope that such a model could be used by people without any game theoretic training, or be incorporated into a set of computer programs to provide quick and easy to understand strategies for the game in question. Also, one other desirable feature of this model would be such that the stages of the game could defined on a single matrix, for at least the case where there are two players.

A potential basis for such a model for a two player game, is a system comprising:

(i) a set $P = \{P_1, P_2\}$, which is the set of players of the game,

(ii) for each $P_i \in P$, a non–empty finite set $S_k^{i,j}$, which is the set of moves available to $P_j \in P$ as perceived by P_i, at stage k of the game, $k = 1, 2, \ldots$,

(iii) for each $P_i, P_j \in P$, a function $L_k^{i,j}$, the link function, that maps the set $S_k^{i,j}$ onto $S_k^{j,j}$ at each stage k of the game, $k = 1, 2, \ldots$,

(iv) for each $P_i, P_j \in P$ a function $R_k^{i,j}$ that maps the set of all possible outcomes onto the real line R. This denotes the pay–off that P_i believes P_j will receive from each outcome, at each stage k of the game, $k = 1, 2, \ldots$,

(v) for each $P_i, P_j \in P$, a function $U_k^{i,j}$, P_j's utility function as perceived by P_i at stage k of the game, $k = 1, 2, \ldots$, and

(vi) for each $P_i, P_j \in P$, a probability density function $f_k^{i,j}$, giving P_i's subjective probabilities over the move that P_j $(j \neq i)$ will make at stage k of the game, $k = 1, 2, \ldots$

Given that we can define all of the above factors and distributions, a player can then determine his subjective probabilities of what his opponent will do at every future stage of the game. After observing a further play of the game, a player can update his subjective probabilities of the future play of his opponent, after he has updated the various parameters in the above system.

This new model could then be used to obtain not only a better understanding of the games considered by Bennett, but also how a player of such a game could determine his optimal strategy. For instance it would be possible to determine an optimal strategy for the soccer hooliganism example of Bennett, Dando & Sharp (1980) or the Arms limitation example of Bennett & Dando (1982). Care must however be taken in 'games' such as these, in determining what is meant by the utility function of a group, and the updating of the beliefs of a group.

A considerable amount of research has been performed on the combining of opinions (see Genest & Zidek, 1986), some of which could be helpful here. Also other confrontations, such as a dispute between a trade–union and an employer could also be considered by this new approach.

2. Development of an Effective Strategy for Axelrod's Tournaments.

The tournaments run and analysed by Axelrod were reported on in chapter 3 of this thesis. I feel that it would be instructive to determine a general strategy that would do well in an experimental game setting, such as that set by Axelrod. Such a strategy would be aiming to maximise the pay–off obtained, given that the opponent is thought to be playing a strategy from a particular set (i.e. the other strategies submitted). It is likely that an optimal strategy in such a competition will encourage cooperation (and will therefore be forgiving), and will punish defection (and will therefore be provocative). Also the ability to be 'nice' would appear to be a desireable attribute of such a strategy. As well as these three attributes that Axelrod discusses, there is a fourth that would appear to be advisable: reciprocity. By reciprocity I mean the ability to recognise and reciprocate a forgiving move by the opponent.

It would seem unlikely that these attributes would be optimal if they were simply hard–and–fast rules, and therefore more flexible rules should be developed. These rules should depend upon the move sequence to date, and in particular, the responsiveness of P_2 to earlier applications of these rules. A model could be determined for the four attributes, such as in a PDG:

(a) Niceness: $\mathsf{P}\left[m_1(t+a)=1 \mid \mathbf{m}(t-p,t)=(1,1)\right]=\alpha_p$ for $p=0,\ldots,t-1$ and $a=1,2,\ldots,$

(b) Provocability: $\mathsf{P}\left[m_1(t+b)=1 \mid \mathbf{m}(t-q,t-1)=(1,1),\mathbf{m}(t)=(1,2)\right]=\beta_q$ for $q=0,\ldots,t-1$ and $b=1,2,\ldots,$

(c) Reciprocity: $\mathsf{P}\left[m_1(t+c)=1 \mid \mathbf{m}(t-r,t-1)=(2,2),\mathbf{m}(t)=(2,1)\right]=\gamma_r$ for $r=0,\ldots,t-1$ and $c=1,2,\ldots,$

(d) Forgiveness: $\mathsf{P}\left[m_1(t+d)=1 \mid \mathbf{m}(t-s,t)=(2,2)\right]=\delta_s$ for $s=0,\ldots,t-1$ and $d=1,2,\ldots,$

where

$$t = \text{ present time period}$$

$$1 = \text{ Cooperation,} \quad 2 = \text{ Defection} \qquad \text{(and bold type indicates a vector)}$$

$$m_i(k) = \text{ move made by } P_i \text{ at time } k$$

$$\mathbf{m}(k) = (m_1(k), m_2(k))$$

$$\mathbf{m}(k-v,k) = (\mathbf{m}(k-v), \mathbf{m}(k-v+1), \ldots, \mathbf{m}(k))$$

TFT is modelled quite simply by setting $\alpha_1 = 1$ and $\gamma_1 = 1$ and all other parameters to zero. Obviously the values of α_p, β_q, γ_r and δ_s will be dependent upon the game parameters and the player's prior beliefs, and can be updated as the game progresses. Strategies can then be determined to attempt to break out of runs of mutual defections, or have other desireable features. Also it may be desireable to incorporate discount factors into the model to discount the effect of the four attributes in future periods, and how much past play affects the next decision. By altering the parameters of these attributes (and discount factors), various different strategies can be determined. It would be interesting to calculate the effect of altering these attributes, so that an optimal strategy can be determined for various game situations and differing types of opponents.

3. An Advertising Example.

One real world application of the type of games that have been considered above would be an advertising example. Consider a market where there are only two manufacturers of a particular product (firm 1 and firm 2). Both companies invest in advertising for their own products on a regular basis. Now obviously an increase (or decrease) in advertising for Firm 1 will potentially affect the sales of the product for both firm 1 and firm 2. This is similar to the PDG model discussed in subsection 6.4.1 above. Now companies may well be interested in the most effective amount to spend on advertising. Or they may be interested in the possibility of a change in the pack of their product, or in a major relaunch of their product. In this case, they will wish to determine the likely response of the competing firm after such a change or relaunch.

A time series study could be made of the reactions and responses of firm 2 to various strategies by firm 1. Having determined how firm 2 is likely to respond to any given strategy by firm 1, firm 1 can determine the optimal strategy to adopt, in order to maximise the effect of advertising per cost of advertising, or any other desired effect. Other effects such as the seasonality of the product (if any), the rate of dimishing returns of increased advertising or the decay rate of past advertising could also be considered by such a study. Obviously a model such as this could then be extended to markets where there are more than two competitors, or to a completely different effect in a market, or to a completely different kind of market altogether.

4. Calibrated Strategies.

Harsanyi (1977) introduced the notion of *unprofitable games* which are games where all equilibrium points yield at most the maximin pay–off to each player. It is argued that in such games, one should always play one's maximin strategy as this is more stable than any equilibrium available. Indeed, it can be shown that the maximin strategy will obtain at least as much utility as any other strategy when playing against a rational player. From this it can be seen that the only Bayes calibrated strategy for any player is the maximin strategy. By a Bayes calibrated strategy (or simply calibrated strategy) we mean a strategy that maximises expected utility given a player's beliefs that the other players of the game are likewise maximising their expected utilities. This leads us to question the form of the calibrated strategies for more general games. Also, under what conditions will such a calibrated strategy be unique?.

We can then determine calibrated societies, i.e. groups of players that are all calibrated. To do this we specify a density $f(d^*)$ of the predictive distribution of the players over the set of moves, and a distribution $\Pi(\theta)$ over utilities, corresponding to the parameters of the game θ. For each set of values of the parameters θ we can find the set of Bayes decisions $d^*(\theta)$. The pair (f, Π) that define the society are then calibrated if the density function of $d^*(\theta)$ corresponding to $\Pi(\theta)$ is equal to $f(d^*)$. Note here that if $d^*(\theta)$ is not unique, we can take a distribution g_θ over the choice of $d^*(\theta)$ for given values of the parameters θ.

So I feel that it would be instructive to find the calibrated strategies and societies for particular games and utility structures. This should lead to a better understanding of how to play these games, or groups of games. Note that this is very similar to the concept of evolutionary stable strategies as was discussed by Maynard–Smith (see chapter 3 above), where unsuccessful strategies are replaced by successful strategies as the game continues, leaving only the stable strategies. A direct comparison between calibrated strategies and evolutionary stable strategies may also prove useful, as well as a study to determine the conditions that ensure that players form a calibrated unit during the game.

5. Continuous Games.

An extension to the games that we have considered that might give some insightful results is obtained by considering continuously repeated games. By a continuously repeated game, we mean a game that is being continuously played by a number of players, and each player is always playing a particular strategy. A player can choose to change his strategy at any time–point in the game, and will play this new strategy up to the time that he decides to change it again. The game is defined at any time–point by a pay–off matrix, which may be held the same throughout time, or may change with time. The derivation of optimal strategies for these continuous games may prove useful in our understanding of optimal strategies for the more usual discrete games.

6. Optimal Control Approach.

By adopting an optimal control approach (see, for example, Ross, 1983 or Whittle, 1983) to the games that we have considered above, we would obtain more sophisticated techniques to find the optimal strategies. For instance, in the PDG example in chapter 7 above, an optimal control approach could be used to calculate a more precise formulation of the solution. This could be carried over to many other types of games where the actions of a player at any one stage of the game have effects apparent for a number of future stages. From the type of results that we would obtain from such an approach, we would be able to determine a lot more about the structure of the problem, and the form of the optimal strategies.

7. Comparison with Stochastic Games.

As was mentioned in chapter 5 above, several authors (e.g. Mertens & Neyman, 1981) have considered stochastic games — games where the players' strategies not only determine the pay–off, but also control the transition probabilities that determine the pay–off matrix for the next stage of the game. Now there are obviously strong links between optimal strategies in these stochastic games, and optimal strategies in the repeated games with incomplete information that have been considered above. A formal comparison of the form of the solutions for the two types of games, and the conditions necessary for equality on the form of the stochastic game pay–off matrices, would hopefully produce some interesting results. We can see that, for comparable pay–off matrices, the limiting stochastic game (as the transition probabilities degenerate) will correspond to the limiting repeated game (as the incomplete information diminishes).

8. Rationalizable Strategies and the Infinite Regress.

In chapter 5 we discussed the concept of rationalizable strategies, as developed by Pearce (1984) and Bernheim (1984). This concept can be seen to be similar to the stable strategies that are determined for the infinite regress in chapter 4. However the two approaches work in opposite directions, as the rationalizable strategies are calculated by a decreasing iterative procedure, whereas the stable strategies are calculated by considering higher and higher levels of the regress. It would be interesting to compare the two methods, and to calculate when they determine the same solutions, e.g. what assumptions we must place upon the players' beliefs about the utility functions of their opponents.

Also I believe that it would be fruitful to consider the effects on the infinite regress of other beliefs about the utility functions, i.e. under what conditions can we truncate the regress. Simple beliefs were considered in chapter 4, but it must be possible to determine other sets of beliefs that lead to the regress being curtailed. Also in chapter 4 we discussed the possibility of placing a distribution over the levels of the regress that an opponent is believed to consider. This would seem to be a natural extension to the work presented before. A model could be developed with such a feature to determine the effect of differing distributions of beliefs about the opponent.

9. Optimal Summaries for Games.

Smale (1980) showed that equilibria could be found for games when the players only retained some average or summary of the previous outcomes of the game. Now this raises the question of, for any given game, which is the most efficient summary of the previous outcomes or interactions for a player to retain, in order to determine his optimal (utility maximising) strategy? That is, for any particular game, what set of statistics is sufficient for the calculation of optimal strategies. Smale (1980) considered a simple averaging summary for a PDG in order to determine an equilibrium point. However, a more complicated summary may be required when considering utility maximising strategies as opposed to equilibria, depending on the utility framework assumed. I believe that in many, if not all games, such a set of sufficient statistics can be determined that would eliminate the necessity for the players to remember *all* previous interactions in the game.

10. Influence Diagrams.

As discussed above, one representation of a game is in terms of its extensive form, i.e. a game tree. A much more compact and efficient representation than a game tree is an ***influence diagram***, which is a schematic representation showing the relationships between the component decision variables and random vectors. For a full description of influence diagrams and their applications, see Smith (1987) and Smith (1988). The theory that exists for these influence diagrams should be able to give us some insight into the relationships and dependencies that are pertinent in the game under consideration, and possible short–cuts that could be taken in the analysis of a game. These influence diagrams would also enable us, given the structure of the game, to analyse the form of the optimal solution for different players.

For example, in the industrial example in point 3 of this chapter, we could use influence diagrams to determine the interdependencies of various factors such as the rate of inflation, the population's affluence and the sales of the product. Also, in the asymmetric games discussed in point 1 above, influence diagrams could be used to investigate the relationship between the perceived moves avavilable to the players.

11. Effect of a Training Period

One area that might be interesting to consider is the effect on the interactions in a game, of an initial 'training period'. If the players have the opportunity to play a small number of stages of the game, where the utility gained from these stages is uniform over all outcomes, how will this affect the moves that they choose during the actual game? Players may use such a period to 'agree' upon a mutually beneficial outcome (e.g. mutual Cooperation in a PDG), or they may use it to inform an opponent of their intentions (e.g. always to play move 2 in a game of 'chicken'). I believe that an experimental study of the different effects of such a training period on different games may well provide some interesting results.

12. Multiple Objectives.

A further study that could be made is an experimental study to determine the types of utility functions that people have in games such as those discussed above. We would wish to find out whether players are maximising their utility over only their own pay–offs, or whether they have utilities over the pay–offs to the other player(s) of the game. If they do have utilities over the pay–offs to an opponent, does a player in general gain utility from his opponent obtaining a high pay–off, or does he gain more utility from maximising the difference between his pay–off and his opponent's. Also, how the results of such a study affect the results obtained in the previous chapters should be considered.

10. CONCLUSIONS

In this thesis we have developed Bayesian models of non–cooperative games. Bayesian game theory extends from traditional game theory, but is concerned with a player trying to achieve his maximum expected utility, given his subjective beliefs about the game, rather than concentrating on equilibria. This area has been considered by a large number of authors, and we have extended earlier work as well as developing new ideas. Like traditional game theory, the literature on Bayesian game theory has been developed in a variety of disciplines and areas, that often do not communicate with each other, and so the literature is widely dispersed. We have considered the major strands of Bayesian game theoretic research, and tried to show how they fit in with each other. We have concentrated mainly on the modelling aspects and assumptions of these games, rather than explicitly determining precise models for particular games.

We have developed a framework for considering the infinite regress that arises in games with incomplete information. From this we can see which assumptions are required to limit this regress to a finite number of levels, and when it is necessary to use other finite approximations. From this framework we can also see how previous work relates to other work in this area. The framework is developed in a natural way, as it considers the increasing levels of thoughts that players can think about, until a stable solution is reached. From this we could go on to consider other aspects, such as a player's beliefs about the limit of the number of levels that any opponent will consider. This infinite regress is an important concept, as it is implicitly incorporated in many Bayesian models of games.

We then considered the dichotomy between theoretical results and experimental results, which is prevalent in the literature. We developed Bayesian models for the types of experimental games that have been used. We argued that these models must incorporate considerations of the rationality of the opponents. By assuming a realistic class of utility functions, we can determine appropriate models for the behavioural relationships between the players of the game, and from this we can show observed behaviour to be rational in a game theoretic sense. From this we can determine a normative theory of how a player ought to play, given his beliefs about how his opponents will play, and we can also draw inferences about the players from their observed moves. Therefore it should be possible to determine appropriate models for all game situations that players face, whether experimental or not.

In chapter 7 we focused on one particular model for determining optimal moves in a game. We showed how we can adapt this model by incorporating the form of the optimal solution, to improve the efficiency, speed and applicability of the model. We demonstrated this explicitly for one particular example (a Prisoner's Dilemma game), although we *showed that the model* can be applied to all of the types of games that we are considering. The improvement achieved by incorporating the form of the solution depends upon the amount of probabilistic structure that is assumed by the model. The more structure that is assumed, the more the model will be improved by using *the form of the solution*. From these improvements we can test the appropriateness of the model, and also adapt the player's beliefs about his opponents' future play. So algorithms can be used to determine optimal strategies for a player of a game, and by considering the mathematical implications of the assumed model, the algorithm can be extremely useful. The appropriateness of the model can be tested by considering the rationality arguments mentioned above, and only when reasonable assumptions are made will the model be at all realistic.

Then we considered a large and reasonably widely used class of strategies. We showed that, under the assumption that the opponent was playing a strategy *from this class*, and some weak regularity conditions, it was not optimal for the player to use such a strategy. This is an important result if we consider some of the rationality arguments above, as players might be assumed to have similar beliefs to each other. We argued that players must be careful about assuming an opponent to be playing a strategy such as this when the opponent is from essentially the same population as the player, as in for example, experimental games. One strategy from this class (TFT) has proved to be exceptionally good in a variety of game settings. This is due to its degenerate form, and also its transparency. It is possible, however, that strategies that achieve better results than TFT could be *found for most of the game settings*. So, by comparing the effectiveness of various strategies, and considering the rationality of the players, we can determine the optimality of a given set of strategies. From these results we can see that strategies that do not utilise all the available information are, in given circumstances, suboptimal.

Despite there being a wide and well spread literature, there are still many areas left untouched, and much left to do. In the previous chapter we discussed twelve areas that I believe will lead to interesting results and the development of the subject. Therefore, we can see from the above chapters, that Bayesian game theoretic models can provide good prescriptions for behaviour. These models must, however, be guided by ideas of rationality, and therefore appropriate to the game in question. Models can be enhanced by incorporating these rationality concepts, and by considering the mathematical forms of the strategies and solutions. By incorporating all of these features, realistic models of games can be found.

REFERENCES

Abakuks, A. (1980), *Conditions for evolutionary stable strategies*, Journal of Applied Probability 17 : 559 – 562.

Aumann, R.J. (1974), *Subjectivity and correlation in randomised strategies*, Journal of Mathematical Economics 1 : 67 – 96.

Aumann, R.J. (1976), *Agreeing to disagree*, Annals of Statistics 4 : 1236 – 1239.

Aumann, R.J. (1981), *Survey of repeated games*, in Essays in Game Theory and Mathematical Economics in Honour of Oskar Morgenstern, Bibliographisches Institut, 11 – 42.

Aumann, R.J. (1987), *Correlated equilibrium as an expression of Bayesian rationality*, Econometrica 55 : 1 – 18.

Aumann, R.J., Katznelson, Y., Radner, R., Rosenthal, R., Weiss, W. (1983), *Approximate purification of mixed strategies*, Mathematics of Operational Research 8 : 327 – 341.

Axelrod, R. (1980a), *Effective choice in the Prisoner's dilemma*, Journal of Conflict Resolution 24 : 3 – 25.

Axelrod, R. (1980b), *More effective choice in the Prisoner's dilemma*, Journal of Conflict Resolution 24 : 379 – 403.

Axelrod, R. (1984), *The evolution of cooperation*, Basic Books, New York.

Behr, R.L. (1981), *Nice guys finish last — sometimes*, Journal of Conflict Resolution 25 : 289 – 300.

Bennett, P.G. (1977), *Toward a theory of hypergames*, Omega 5 : 749 – 751.

Bennett, P.G. (1985), *On linking approaches to decision–aiding*, Journal of the Operational Research Society 36 : 659 – 669.

Bennett, P.G. (1987), *Analysing conflict and its resolution*, Oxford University Press.

Bennett, P.G. and Dando, M.R. (1979), *Complex strategic analysis: a hypergame study on the fall of France*, Journal of the Operational Research Society 30 : 23 – 32.

Bennett, P.G. and Dando, M.R. (1982), *The Arms race as a hypergame: a study of routes toward a safer world*, Futures 14 : 293 – 306.

Bennett, P.G., Dando, M.R. and Sharp, R.G. (1980), *Using hypergames to model difficult social issues: an approach to the case of soccer hooliganism*, Journal of the Operational Research Society 31 : 621 – 635.

Bennett, P.G. and Huxham, C.S. (1982), *Hypergames and what they do: a 'soft O.R.' approach*, Journal of the Operational Research Society 33 : 41 – 50.

Bennett, P.G., Huxham, C.S. and Dando, M.R. (1981), *Shipping in crisis: a trial run for live application of the hypergame approach*, Omega 9 : 579 – 594.

Bernheim, D. (1984), *Rationalizable strategic behaviour*, Econometrica 52 : 1007 – 1028.

Bishop, D.T. and Cannings, C. (1976), *Models of animal conflict*, Advances in Applied Probability 8 : 616 – 621.

Bishop, D.T. and Cannings, C. (1978), *A generalised war of attrition*, Journal of Theoretical Biology 70 : 85 – 124.

Blad, M.C. (1986), *A dynamic analysis of the repeated Prisoner's dilemma game*, International Journal of Game Theory 15 : 83 – 99.

Colman, A.M. (1982), *Game theory and experimental games*, Pergamon Press, Oxford.

Dawid, A.P. (1982), *The well calibrated Bayesian*, Journal of the American Statistical Association 77 : 605 – 613.

DeGroot, M.H. (1970), *Optimal statistical decisions*, McGraw–Hill.

Flood, M.M. (1954a), *Game learning theory and some decision–making experiments*, in Decision Processes (eds. Thrall, Coombs and Davis), Wiley.

Flood, M.M. (1954b), *Environmental non–stationarity in a sequential decision–making experiment*, in Decision Processes (eds. Thrall, Coombs and Davis), Wiley.

Flood, M.M. (1958), *Some experimental games*, Management Science 5 : 5 – 26.

Genest, C. and Zidek, J. (1986), *Combining probability distributions: a critique and an annotated bibliography*, Statistical Science 1 : 114 – 148.

Grofman, B. and Pool, J. (1975), *Bayesian models for iterated Prisoner's dilemma games*, General Systems 20 : 185 – 194.

Haigh, J. (1975), *Game theory and evolution*, Advances in Probability 7 : 8 – 11.

Hamburger, H. (1979), *Games as models of social phenomena*, W.H. Freeman & Co.

Harford, T. and Solomon, L. (1967), *"Reformed Sinner" and "Lapsed Saint" strategies in the Prisoner's dilemma game*, Journal of Conflict Resolution 11 : 104 – 109.

Harsanyi, J.C. (1967), *Games with incomplete information played by 'Bayesian' players. Part I*, Management Science 14 : 159 – 182.

Harsanyi, J.C. (1968a), *Games with incomplete information played by 'Bayesian' players.*

Part II, Management Science 14 : 320 – 334.

Harsanyi, J.C. (1968b), *Games with incomplete information played by 'Bayesian' players. Part III*, Management Science 14 : 486 – 502.

Harsanyi, J.C. (1977), *Rational behaviour and bargaining equilibrium in games and social situations*, Cambridge University Press.

Harsanyi, J.C. (1982), *Subjective probability and the theory of games : comments on Kadane and Larkey's paper*, Management Science 28 : 120 – 124.

Harsanyi, J.C. and Selten, R. (1972), *A generalised Nash solution for two-person bargaining games with incomplete information*, Management Science 18 : P80 – P106.

Howard N. (1966a), *The theory of metagames*, General Systems 11 : 167 – 186.

Howard N. (1966b), *The mathematics of metagames*, General Systems 11 : 187 – 200.

Howard N. (1970), *Some developments in the theory and applications of metagames*, General Systems 15 : 205 – 231.

Howard N. (1971), *Paradoxes of rationality*, MIT Press, Cambridge, Massachusetts.

Jeffrey, R.C. (1983), *The logic of decision*, 2nd edition, Chicago University Press.

Kadane, J.B. and Larkey, P.D. (1982), *Subjective probability and the theory of games*, Management Science 28 : 113 – 120.

Kadane, J.B. and Larkey, P.D. (1983), *The confusion of is and ought in game theoretic contexts*, Management Science 29 : 1365 – 1379.

Kelley, H.H. and Stahelski, A.J. (1970), *Social interactions basis of Cooperators' and Competitors' beliefs about others*, Journal of Personality and Social Psychology 16 : 66 – 91.

Kreps, D. and Wilson, R. (1982), *Sequential equilibria*, Econometrica 50 : 863 – 894.

Laskey, K.B. (1985), *Bayesian models of strategic interaction*, Ph.D. thesis, Carnegie–Mellon University, Pittsburg.

Lindley, D.V. (1985), *Making decisions*, Wiley, New York.

Luce, D. and Raiffa, H. (1957), *Games and decisions*, Wiley, New York.

Maynard-Smith, J. (1982), *Evolution and the theory of games*, Cambridge University Press.

Mertens, J.F. and Neyman, A. (1981), *Stochastic games*, International Journal of Game Theory 10 : 53 – 66.

Mertens, J.F. and Zamir, S. (1971), *Value of two person repeated games with lack of information on both sides*, International Journal of Game Theory 1 : 39 – 64.

Mertens, J.F. and Zamir, S. (1985), *Formulation of Bayesian analysis for games with incomplete information*, International Journal of Game Theory 14 : 1 – 29.

Myerson, R.B. (1986), *Acceptable and predominant correlated equilibria*, International Journal of Game Theory 15 : 133 –154.

Nash, J.F. (1951), *Non–cooperative games*, Annals of Mathematics 54 : 286 – 295.

von Neumann, J. and Morgenstern, O. (1947), *Theory of games and economic behaviour*, Princeton University Press.

Osgood, C.E. (1980), *The GRIT strategy*, Bulletin of the Atomic Scientist, May, 58 – 60.

Owen, G. (1982), *Game Theory*, Academic Press, New York.

Parthasarathy, T. and Raghavan, T.E.S. (1971), *Some topics in two-person games*, American Elsevier Publishing Company, Inc., New York.

Pearce, D.G. (1984), *Rationalizable strategic behaviour and the problem of perfection*, Econometrica 52 : 1029 – 1050.

Pruitt, D.G. and Kimmel, M.J. (1977), *Twenty years of experimental gaming : critique, synthesis and suggestions for the future*, Annual Review of Psychology 28 : 363 – 392.

Rapoport, A. (1967), *Exploiter, Leader, Hero, and Martyr: the four archetypes of the 2×2 game*, Behavioural Science 12 : 81 – 84.

Rapoport, A. (1974), *Game theory as a theory of conflict resolution*, D. Reidel Publishing Company.

Rapoport, A. and Chammah, A.M. (1965), *Prisoner's dilemma*, Ann Arbor, The University of Michigan Press.

Rapoport, A. and Guyer, M. (1966), *A taxonomy of 2×2 games*, General Systems 11 : 203 – 214.

Ross, M.S. (1983), *Introduction to stochastic dynamic programming*, Academic Press.

Savage, L.J. (1954), *The foundations of statistics*, Wiley & Sons, New York.

Selten, R. (1964), *Valuation of n person games*, in Advances in Game Theory (eds. Dresher, Shapley and Tucker), Princeton University Press, 577 – 626.

Selten, R. (1975), *Reexamination of the perfectness concept for equilibrium points in extensive games*, International Journal of Game Theory 4 : 25 – 55.

Selten, R. (1978), *The Chain store paradox*, Theory and Decision 9 : 127 – 159.

Shapley, L.S. (1953), *A value for n person games*, in Contributions to the Theory of Games, II, Princeton University Press, 307 –317.

Shin, H.S. (1988a), *Bayes–rationality as the stability of decisions and correlated equilibria*, unpublished, Oxford University.

Shin, H.S. (1988b), *Logical structure of common knowledge*, unpublished, Oxford University.

Shin, H.S. (1989), *Two notions of ratifiability and equilibrium in games*, unpublished, Oxford University.

Shubik, M. (1970), *Game theory, behaviour, and the paradox of the Prisoner's dilemma: three solutions*, Journal of Conflict Resolution 14 : 181 – 194.

Shubik, M. (1981), *Perfect or robust noncooperative equilibrium: a search for the Philosopher's stone?*, in Essays in Game Theory and Mathematical Economics, Bibliographisches Institut : 153 – 180.

Shubik, M. (1982), *Game theory in the social sciences*, M.I.T. Press, Cambridge.

Shubik, M. (1983), *Comment on 'The confusion of is and ought in game theoretic contexts'*, Management Science 29 : 1380 – 1383.

Shubik, M. and Sobel (1980), M.J., *Stochastic games, oligopoly theory and competitive resource allocations*, in Dynamic Optimization and Mathematical Economics (ed. P.-T. Liu), Plenum, New York.

Simon, H. (1957), *Models of man: social and rational*, Wiley, New York.

Smale, S. (1980), *The Prisoner's dilemma and dynamical systems associated to non–cooperative games*, Econometrica 48 : 1617 – 1634.

Smith, J.Q. (1984), *A statistical approach to the analysis of repeated two player games*, Research Report 62, Statistics Department, University of Warwick.

Smith, J.Q. (1985), *Diagnostic checks of non-standard time series models*, Journal of Forecasting 4 : 283 – 291.

Smith, J.Q. (1987), *Influence diagrams for Bayes decision analysis*, Research Report 100, Statistics Department, University of Warwick.

Smith, J.Q. (1988), *Models, optimal decisions and influence diagrams*, in Proceedings of the 3rd Conference on Bayesian Statistics at Valencia, Oxford University Press.

Smith, J.Q. and Young, S.C. (1987), *Stochastic experimental games – A Bayesian approach*, Research Report 128, Statistics Department, University of Warwick.

Taylor, M. (1976), *Anarchy and cooperation*, Wiley, New York.

Terhune, K.W. (1974), *"Wash-In", "Wash-Out" and systemic effects in extended Prisoner's dilemma*, Journal of Conflict Resolution 18 : 656 – 685.

Thomas, L.C. (1984), *Games, theory and applications*, Wiley, New York.

van der Wal, J. (1981), *Stochastic dynamic programming*, Mathematical Centre Tracts 139, Amsterdam.

West, M. (1986), *Bayesian model monitoring*, Journal of the Royal Statistical Society, B, 48 : 70 – 78.

Whittle, P. (1983), *Optimisation over time*, Vol. 2, Wiley.

Wilson, J.G. (1986), *Subjective probability and the Prisoner's dilemma*, Management Science 32 : 45 – 55.

Wilson, W. (1971), *Reciprocation and other techniques for inducing cooperation in the Prisoner's dilemma game*, Journal of Conflict Resolution 15 : 167 – 195.

Young, S.C. and Smith, J.Q. (1988a), *Suboptimality of m-step back strategies in Bayesian games*, Research Report 139, Statistics Department, University of Warwick (submitted to the International Journal of Game Theory).

Young, S.C. and Smith, J.Q. (1988b), *Deriving and analysing optimal strategies in Bayesian models of games*, Research Report 156, Statistics Department, University of Warwick (submitted to Management Science).

Zeeman, E.C. (1979), *Population dynamics from game theory*, in Global theory of Dynamic Systems, Springer–Verlag, Berlin.

Zeeman, E.C. (1981), *Dynamics of the evolution of animal conficts*, Journal of Theoretical Biology 89 : 249 – 270.