Repeated Games: A State Space Approach

OLOF LEIMAR*

Department of Zoology, Stockholm University, S-106 91, Stockholm, Sweden

(Received on 19 April 1996, Accepted in revised form on 30 September 1996)

Repeated games, such as the Iterated Prisoner’s Dilemma (IPD), are often used as idealised models of social interactions. Here, I develop a state space approach to the study of two-player, repeated games. A strategy is represented by way of a state space, where a player’s choice of action depends on the current state, and where the actions performed can cause transitions from one state to another. This kind of representation provides a possible link between a game theoretical analysis and concepts from mechanistically oriented ethology, where an individual’s state is viewed as made up of motivational variables. Using the concept of a limit ESS for a game in the extensive form, I derive a number of fundamental results for state space strategies. Conditions ensuring purity of a limit ESS are given, as well as a characterisation of the most important class of pure limit ESSs. The theoretical analysis covers games with and without a role asymmetry and also games where players move alternately. To illustrate the state space approach, I apply the theoretical results to three examples. First, for the role symmetric IPD, I find a great number of pure limit ESSs, and relate these to the strategies most frequently studied previously. I also discuss whether there can be randomised limit ESSs, concluding that although this is possible, none have been found so far. Second, as a game possessing a role asymmetry, I study a simplified model of social dominance. I concentrate on the question of whether punishment administered by a dominant can determine the allocation of a resource between the dominant and a subdominant. The game turns out to have limit ESSs with this property, but there are also stable strategies where the dominant lacks control. Third, I analyse an alternating Prisoner’s Dilemma, which is a natural model to investigate the evolution of reciprocal altruism. No stable strategy for this game has been described previously. Of the limit ESSs I find, one is of particular interest, in that it closely corresponds to the notion of reciprocal altruism as conceived by Trivers.

1. Introduction

A repeated game, where two players repeat the same component game a number of times, is perhaps the simplest possible evolutionary model of a social interaction. The Iterated Prisoner’s Dilemma (IPD) is the most frequently studied example (Axelrod & Hamilton, 1981). Although such models may well be too simplified to capture the essence of real social interactions, they still have some conceptual interest. My intention here is to present a state approach which is useful for the analysis of such games, and which may also be of value for more realistic models of social interactions.

* E-mail: Olof.Leimar@zool.su.se
by an appropriate choice of a sequence of actions (Whittle, 1982). For optimality, the current choice of action should in principle be based on all information available to the decision maker. However, sometimes it will be sufficient to base decisions on only certain aspects of the information, with the remaining aspects being irrelevant for the purposes of the decision maker. The relevant aspects are called a state, and the essence of the concept lies in economy of description. In principle, “all available information” can always be regarded as a state, but our chances of actually computing an optimal strategy will be much greater if the problem allows a simple state space. The method of choice for determining an optimal strategy is dynamic programming (Whittle, 1982).

For social behaviour, a decision maker interacts with other decision makers, and a sufficient state must summarise all available information about the states of a social partner, and these states should in turn be sufficient for the partner. There is a relexivity in this situation, in that the states of the partners must contain sufficient information about each other. Of greatest interest will be fairly simple state spaces that are reflexive in this way and capable of supporting the social behaviour in question. These can be seen as a kind of minimal psychologies for the social interaction.

In the IPD and similar repeated games the players in principle know the past sequence of actions, and there are no variables, such as a player’s need for the help offered by a partner, influencing the pay-off received from some specified sequence of actions. A state must then be determined by the past sequence of actions, and is entirely “public”: it cannot have any components that are private to a player. This greatly simplifies the analysis, but might also be the point where most biological realism is lost; communication between players in such games is entirely trivial.

An example of a public state space would be the moves used in the previous round, and this has been used to analyse the IPD (Nowak & Sigmund, 1993; Nowak et al., 1995). Restriction of the players’ memory horizon to a certain number of rounds can be viewed as an assumption of a certain state space for the players, but most state spaces do not have this interpretation. For instance, perhaps the most interesting strategy suggested for the IPD is “contrite tit-for-tat” (Boyd, 1989), and as I will show it can be viewed as having a state space with three states, but it cannot be interpreted as having a memory limited to some finite number of rounds (it will “remember” a defection until it has been “compensated” for the loss incurred by the partner).

When a player only has a finite number of actions to choose from in the component game, as in the IPD, it is natural to consider strategies that can be represented using a finite state space, and I refer to these as state space strategies. As far as I am aware, all strategies for the IPD analysed in the biological literature are state space strategies. Although many strategies without a finite state space exist (for instance, to defect when the current round number is prime and cooperate otherwise), it seems unlikely that much of interest would be lost by looking for evolutionarily stable strategies among the state space strategies.

My aim is to develop a general method of analysis of state space strategies for repeated games. The component game can be any finite matrix game, either symmetric or having a role asymmetry. I will also deal with games where players move alternately. These have less often been used as biological models, but are sometimes quite natural and lend themselves well to a state space analysis.

For games with a time structure (extensive form games) a suitable definition of evolutionary stability presents certain problems. Maynard Smith’s (1982) second condition for an evolutionarily stable strategy (ESS) can be extremely restrictive for extensive form games, and it is important to apply the condition in a way that best represents the situation one intends to describe. Selten (1983) used a more liberal notion of a stable strategy, referred to as a limit ESS, in his theory of two-person, extensive form games endowed with a natural symmetry. In this theory, the possibility that a player makes mistakes when attempting to execute a strategy is taken into account. Loosely speaking, the concept of a limit ESS implies that one regards a game where players do not make mistakes as an idealisation of a situation where they may be slightly fallible.

Selten’s (1983) theory is somewhat complex but has the advantage of conceptual power and precision, and for this reason I use limit ESS as the stability criterion in my analysis of repeated games. Although I do not characterise all limit ESSs of all repeated games, for the purposes of using such games as biological models my treatment is essentially complete.

To illustrate the most important concepts entering into a state space analysis, I first go through a number of examples, at the same time discussing the nature of certain of the limit ESSs. For the IPD, I show the state space representation of a number of pure strategies that are well known from the literature. For “tit-for-tat” (Axelrod & Hamilton, 1981), I point out the connection between its lack of evolutionary stability (Selten & Hammerstein, 1984) and the lack of reflexivity of its state space. For strategies like
“contrite tit-for-tat” (Boyd, 1989) and “Pavlov” (Nowak & Sigmund, 1993). I place them in the context of other pure limit ESS state space strategies of simple structure, of which there are thousands. I also discuss the reasons for the apparent lack of randomised limit ESSs for the IPD, and describe the evolutionary instability of the most well-known suggestion (Nowak & Sigmund, 1992, 1993).

As an example of a game with a role asymmetry, I then analyse a simple model of social dominance, investigating the question of whether a dominant individual can use a threat of punishment to control a subdominant’s access to a resource (Clutton Brock & Parker, 1995). I find a number of limit ESSs having this property, but also some where the dominant lacks control, as well as cases where the control can switch because of random errors in the behavioural sequence.

Finally, to illustrate games with alternating moves I return to the Prisoner’s Dilemma, analysing a model of reciprocal altruism called the alternating Prisoner’s Dilemma (Nowak & Sigmund, 1994). Again, I find many limit ESSs, none of which seem to have been noted previously. One of these is particularly interesting, in that it accurately captures the notion of reciprocal altruism as developed by Trivers (1971). I also point out that an alternating game has a role asymmetry and therefore all limit ESSs must be pure, and I briefly describe the lack of stability of the randomised strategy suggested by Nowak & Sigmund (1994).

I then proceed to a formal and rigorous treatment of state space strategies for infinitely repeated games, which appears in a number of appendices. The treatment presupposes some familiarity with at least the basic concepts of Selten’s (1983) evolutionary game theory. This theory is also described by van Danne (1987), Selten (1988), and Hammerstein & Selten (1994). The repeated games I study are basically special cases of the general games treated by Selten, but differ in that they are not finite. For this reason, I supply proofs also of some results that essentially are special cases of results in Selten (1983).

There are two main outcomes of my analysis, both for simultaneous and alternating moves. First, Theorem 1 (Appendix E) and Theorem 3 (Appendix H) characterise the most important class of pure limit ESS state space strategies. Reflexivity, which forms part of this characterisation, is perhaps the most significant new concept introduced in my treatment. Second, Theorem 2 (Appendix E) and Theorem 4 (Appendix H) deal with conditions ensuring that a limit ESS state space strategy must be pure, and severely restrict the possibility for a randomised strategy to be a limit ESS.

2. The Iterated Prisoner’s Dilemma

The IPD consists of a number of rounds, with the probability $\omega$ of continuing to the next round satisfying $0 < \omega < 1$. In a round, the players simultaneously perform actions, where an action can be either $C$ or $D$, having the usual interpretation of cooperate and defect. A player’s pay-off from a round is $e(aa')$, where $a$ is the player’s own action and $a'$ is the partner’s action. The pay-offs are often written as

$$e(CC) = R, \quad e(CD) = S, \quad e(DC) = T, \quad e(DD) = P,$$

and we have a Prisoner’s Dilemma when $T > R > P > S$. The additional inequality $2R > T + S$ is usually also assumed for the IPD, and it implies that, on average, mutual cooperation is a more “desirable” outcome than alternating $CD$ and $DC$.

A player’s state space $X$ consists of a finite number of states $x$. The state at the beginning of a round determines the player’s choice of action in the round by way of an action rule. With a pure action rule $r$, the player selects the action

$$a = r(x)$$

when in state $x$. We can also consider randomised action rules, where a player draws the action from a distribution which depends on the state.

Initially, a player is in some given state $x_0$. The current state $x$ together with the performed actions $a$ and $a'$ determine the next state, and this relationship is called a state space dynamics:

$$x_{next} = f(x, aa').$$

Clearly, an action rule and a state space dynamics specify a strategy for the player, and this type of strategy is referred to as a state space strategy. Note that the dynamics should be defined for all possible action combinations $aa'$, even if it were the case that the action rule prevented certain action combinations from occurring. This is because we consider the possibility of errors in performing the actions; a player who intends to cooperate may defect by mistake. For the following examples, the probability of mistake should be thought of as infinitesimal.

2.1. State Space Strategies

The simplest possible case is when there is only one state. Two pure strategies for the IPD, often denoted AllD and AllC, can be represented using a single state. Some additional examples of state space
strategies with a low number of states are given in Table 1.

The first strategy illustrated is Pavlov, which was promoted by Nowak & Sigmund (1993) as a particularly likely outcome of evolution. Pavlov has two states, with \( x_0 = 1 \) as the initial state. In the following rounds, the current state is entirely determined by the action combination in the previous round. The symmetric combinations \( CC \) and \( DD \) lead to state \( x = 1 \), where the action rule \( r(x) \) specifies cooperation, and the asymmetric combinations \( CD \) and \( DC \) lead to state \( x = 2 \), where defection is specified. Thus, Pavlov does not take into account who the “guilty party” is when a defection occurs.

When two players of Pavlov meet they keep cooperating, unless there is a mistaken defection by one of them, in which case they both switch to state \( x = 2 \) in the next round. After a round of mutual defection, they then switch back to \( x = 1 \). We can illustrate this as a sequence of action combinations \((CC, CD, DD, CC, CC, \ldots)\), where the underlined \( D \) represents a mistaken defection (I arbitrarily assume the partner to make a mistake in round 2). I will refer to such a sequence as the signature of the strategy. The signature does not amount to a complete description of the strategy, different strategies can have the same signature, but it illustrates an important property of a strategy.

By modifying Pavlov, so that \( DD \) instead leads to \( x = 2 \), we arrive at Grim (Table 1). The main difference from Pavlov is that Grim will not switch back from \( x = 2 \) to \( x = 1 \) unless the combination \( CC \) occurs, but for this to happen a Grim player must cooperate by mistake. When two players of Grim meet, a mistaken defection puts both players in state \( x = 2 \), which leads to mutual defection, and they will not get out of this state unless both simultaneously cooperate by mistake. The signature of Grim is \((CC, CD, DD, DD, DD, \ldots)\).

If we were to change Grim so that when in state \( x = 2 \) no action combination causes a transition to \( x = 1 \) (thus, changing \( f(2, CC) \) from 1 to 2), we get an even more “grim” strategy, where a player cannot get out of the defecting state \( x = 2 \). I use the technical term reducible for a state space dynamics which can get trapped in a state or in a subset of the state space in this way. On the other hand, if some sequence of action combinations connects any two states, the dynamics is called irreducible. I have chosen to only consider state space strategies with irreducible dynamics as examples, mainly because stronger results can be obtained for them, but also because they appear more natural as model psychologies.

The third example is a strategy first suggested by Sugden (1986) and given the name “contrite tit-for-tat” (CTFT) by Boyd (1989). Looking at the state space dynamics in Table 1, one sees that if two players of CTFT meet, their states correspond to each other as follows: if one of them is in \( x = 1 \) the other must also be in this state, whereas if one of them is in \( x = 2 \) the other must be in \( x = 3 \), and vice versa. This kind of correspondence is an expression of the natural symmetry of the game, and I use the term reflected state, \( x^R \), for a state which corresponds to another state, \( x \), in this way. So, for the states of CTFT we have \( 1^R = 1, 2^R = 3, \) and \( 3^R = 2 \) (for both Pavlov and Grim, \( x^R = x \) holds for all the states). Not all state space dynamics are reflexive in this way, but these are of relatively little interest.

CTFT implements reciprocity for the IPD in perhaps the simplest possible way. Using reflection, it is easy to see that the signature of CTFT is \((CC, CD, DC, CC, CC, \ldots)\). Thus, the player receiving a sucker’s payoff \( S \), because of the mistaken defection, is compensated by the partner, who accepts being played for a sucker in the next round. Without an act of contrition from the partner, in the form of a \( C \) that does not require reciprocation, the player remains in the defecting state \( x = 3 \), so CTFT has a memory that potentially stretches arbitrarily far back in time.

Sugden (1986) and Boyd (1989) described CTFT by introducing a distinction of whether or not a player is in “good standing”. Since there are two players, each of which can be in good standing or not, we are

### Table 1

Examples of state space strategies for the IPD

<table>
<thead>
<tr>
<th>Strategy</th>
<th>State*</th>
<th>Action rule†</th>
<th>Next state‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>CD</td>
<td>DC</td>
</tr>
<tr>
<td>Pavlov</td>
<td>1</td>
<td>1 2 2 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 2 2 1</td>
<td></td>
</tr>
<tr>
<td>Grim</td>
<td>1</td>
<td>1 2 2 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 2 2 2</td>
<td></td>
</tr>
<tr>
<td>CTFT</td>
<td>1</td>
<td>1 3 2 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 1 2 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1 3 1 3</td>
<td></td>
</tr>
<tr>
<td>CTFT-like</td>
<td>1</td>
<td>1 3 2 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 1 1 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1 1 1 3</td>
<td></td>
</tr>
<tr>
<td>TFT</td>
<td>1</td>
<td>1 2 1 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 2 1 2</td>
<td></td>
</tr>
<tr>
<td>TFT-like</td>
<td>1</td>
<td>1 3 2 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3 3 1 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2 1 2 3</td>
<td></td>
</tr>
</tbody>
</table>

* Current state \( x \); the initial state is \( x_0 = 1 \) in all examples.
† Action rule \( r(x) \).
‡ Next state \( x_{new} = f(x, aa') \), where \( aa' \) is \( CC, CD, DC, \) or \( DD \).
led to a state space strategy with four states. As noted by Boyd (1989), there is a redundancy in this description, and three states are sufficient. Comparing with Table 1, the state \( x = 1 \) corresponds to both players in good standing or neither player in good standing, the state \( x = 2 \) corresponds to the player of the strategy not in good standing and the partner in good standing, and \( x = 3 \) corresponds to the reversed situation.

As this example illustrates, a given strategy for the game may have different but equivalent state space representations. However, one can remove any redundancy by selecting a state space representation with the lowest possible number of states. The method used to arrive at the unique minimal state space representation of a state space strategy is explained in Appendix D. The examples in Table 1 are all minimal representations.

The strategies mentioned so far are well known from the literature, but the total number of state space strategies is of course infinite. More or less arbitrarily, I have selected a fourth example, called CTFT-like in Table 1. As is easy to see, we have the reflection correspondence \( 1^R = 1, 2^R = 3, \) and \( 3^R = 2 \) for the three states. The strategy is called CTFT-like because its signature, \((CC, CD, DC, CC, \ldots)\), is the same as CTFT’s. Nevertheless, it is a different strategy: it reacts differently to the action combinations \( CD \) when in state \( x = 3 \) and \( DC \) when in \( x = 2 \). There are in fact a great number of different “CTFT-like” strategies, and the situation is similar for Pavlov and Grim.

The famous tit-for-tat (TFT) is also a state space strategy. It differs from the others in Table 1 in that it is not possible to define reflection on its minimal state space; it is not a reflexive strategy. With TFT, a state is determined by the partner’s most recent move \((x = 1 \text{ if } C, x = 2 \text{ if } D)\) so for two TFT players meeting each other, their state transitions are caused by entirely different events. The lack of reflexivity makes TFT a virtually hopeless limit ESS candidate. TFT is well known to have problems (Selten & Hammerstein, 1984), and lack of reflexivity is the root of these problems.

Strategies like TFT which only take the partner’s most recent move into account have been called “reactive” (Nowak & Sigmund, 1992, 1994). The arguments concerning TFT directly extends to any repeated game with simultaneous moves, and we can conclude that strategies taking only the co-player’s most recent move into account cannot be reflexive.

The signature of TFT is \((CC, CD, DC, CD, \ldots)\), so that a mistaken defection causes two players of TFT to start alternating between defection and cooperation. This inability to get cooperation back on track after a disturbance has been regarded as a weakness of TFT. In any case, there are reflexive strategies with this same signature, and the simplest such example is the one labelled TFT-like in Table 1: \( 1^R = 1, 2^R = 3, \) and \( 3^R = 2 \) holds for this strategy.

2.2. Evolutionary Stability for Pure Strategies

Consider a player whose partner uses a reflexive, pure state space strategy with minimal representation given by the state space \( X \) with dynamics \( f \) and action rule \( r \). Dynamic programming is a convenient method of determining optimal strategies for the player in this situation. Let us regard the player as also using the state space \( X \) with dynamics \( f \). Because of reflexivity, for a round where the player is in state \( x \), the partner is in state \( x^R \) and will use the action \( a' = r(x^R) \). If the player then uses the action \( a \), there is a pay-off increment \( e(aa') \) from the current round, followed by a transition to the state \( x_{next} = f(x, aa') \), given that the game continues to the next round. Dynamic programming is performed by solving the so-called optimality equation:

\[
W(x) = \max_a \left[ e(aa') + \omega W(x_{next}) \right], \tag{1}
\]

where \( a' = r(x^R) \) and \( x_{next} = f(x, aa') \). The equation has a unique solution \( W \), and one can interpret \( W(x) \) as the expected future pay-off when in state \( x \) for a best reply to the partner’s strategy. An action \( a \) for which the maximum of the r.h.s. of (1) occurs is called an optimal action in state \( x \), and the optimality equation is said to have strict optimal actions if, for each state \( x \), there is only one optimal action.

There is a similar optimality equation when the partner uses a non-reflexive strategy with state space \( X^R \), dynamics \( f^R \) and action rule \( r^R \). However, in this case the player’s state \( x \) is not viewed as an element in \( X \), but instead as an element in a reflected state space \( X^R \) with dynamics \( f^R \), which can be interpreted as a representation of the distinctions made by the partner when the partner uses a strategy with state space \( X \) (for a reflexive strategy, \( X^R = X \); this is explained in Appendix D).

As I will show (Theorem 1 in Appendix E), if the optimality equation corresponding to \( X, f, r \) has strict optimal actions, then the strategy is a limit ESS if and only if it is reflexive and the optimal action at \( x \) is given by \( r(x) \). With small errors of execution, such a strategy will in fact be a strict Nash equilibrium. The condition in Theorem 1 is only sufficient; there might be non-reflexive limit ESSs for which the optimality equation lacks strict optimal actions, but these ESSs occur only for marginal parameter combinations, and are of little interest.
2.2.1. Examples of cooperative limit ESSs

Call a state space strategy cooperative if the action rule specifies $C$ for the initial state $x_0$ and the dynamics satisfies $x_0 = f(x_0, CC)$, so the player stays in the cooperative initial state as long as the partner cooperates; the examples in Table 1 are of this kind.

Let us now analyse the optimality equation (1) for Pavlov. One readily finds that for $\omega > (T - R)/(R - P)$ eqn (1) has the solution $W(1) = R/(1 - \omega)$, $W(2) = P + \omega R/(1 - \omega)$, with strict optimal actions given by Pavlov’s action rule, so for this parameter range Theorem 1 shows that Pavlov is a limit ESS. On the other hand, for $\omega < (T - R)/(R - P)$ eqn (1) has the solution $W(1) = (T + \omega P)/(1 - \omega^2)$, $W(2) = (\omega T + P)/(1 - \omega^2)$, with a strict optimal action equal to $D$ in both states, so for this parameter range Theorem 1 shows that Pavlov is not a limit ESS. For the marginal case, $\omega = (T - R)/(R - P)$, the optimality equation lacks strict optimal actions, and we cannot use Theorem 1. Closer consideration shows that Pavlov is a limit ESS also for marginal parameter values, but since this case is of little interest, I will not go into it.

The analysis of the other examples in Table 1 proceeds in the same way. Disregarding marginal cases, one finds that Grim is a limit ESS when $\omega > (T - R)/(R - P)$. Note that Grim’s range of stability is greater than Pavlov’s, and is in fact the greatest possible range of stability for a cooperative strategy, because Grim is maximally severe in punishing defection. Since $(T - R)/(R - P) < 1$ must hold for a Prisoner’s Dilemma, Grim will be a limit ESS for large enough probability of continuation, whereas for Pavlov to be a limit ESS for large enough $\omega$ we must have $(T - R)/(R - P) < 1$, which is a relatively restrictive condition, considering that the pay-offs should result from some type of investment in helping the partner.

Similarly, CTFT is a limit ESS when both $\omega > (T - R)/(R - S)$ and $\omega > (P - S)/(R - S)$ hold, which is the condition given by Boyd (1989), and the strategy labelled CTFT-like in Table 1 has the same range of stability. Using the additional pay-off inequality for the Prisoner’s Dilemma, which can be written $T - R < R - S$, one sees that CTFT is a limit ESS for $\omega$ large enough.

Applying Theorem 1 to the non-reflexive TFT shows that TFT cannot be a limit ESS for parameter combinations where the optimality equation has strict optimal actions, which it will have except for some marginal cases. On the other hand, the strategy labelled TFT-like in Table 1 is a limit ESS when both $\omega > (T - R)/(R - S)$ and $\omega > (P - S)/(R - P)$ hold.

So, the various cooperative strategies have different ranges of stability, and strategies with the same signature may differ in the range of stability. Since the signature determines the expected pay-off from deviating once and then obeying the strategy, it gives rise to one stability condition, namely that $C$ should be the optimal action in the initial state $x_0 = 1$. In the examples above with two conditions, the first condition is the one related to the signature. The requirement of using optimal actions also in states other than the initial one can yield additional conditions.

Instead of determining the range of stability for a given strategy, one can consider a parameter combination and look for state space strategies that are limit ESSs with these parameters. Table 2 shows the result of doing this for $R = 2$, $S = -1$, $T = 3$, $P = 0$, and two different values of $\omega$. To generate Table 2, I constructed (by computer) all cooperative, irreducible state space strategies with low number of states in the minimal state space, and then applied Theorem 1 to each of them (a convenient algorithm for solving the optimality equation is the so-called policy-improvement routine; Whittle, 1983). The number of strategies to test grows very rapidly with the number of states, and I did not go beyond four states.

For the brief interaction case, $\omega = 0.34$ is close to the limit, $\omega = (T - R)/(R - P) = 1/3$, below which no cooperative strategy can be stable. As seen in Table 2, the number of limit ESSs for this parameter combination grows rapidly with the number of states. Of the examples in Table 1, all except Pavlov and TFT are present. In spite of the great number of additional limit ESSs, no new signatures appear, so all the strategies for $\omega = 0.34$ in Table 2 are either “Grim-like”, “CTFT-like” or “TFT-like”.

For $\omega = 0.99$, strategies with milder or less efficient punishment sequences can be stable, so for instance Pavlov and a great number of “Pavlov-like” strategies

<table>
<thead>
<tr>
<th>States in state space X</th>
<th>Brief interaction ($\omega = 0.34$)</th>
<th>Long interaction ($\omega = 0.99$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of limit ESSs</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>288</td>
<td>756</td>
</tr>
<tr>
<td>4</td>
<td>75,344</td>
<td>187,938</td>
</tr>
</tbody>
</table>

* Using $C$ in state $x_0$ and remaining in $x_0$ when the action combination $CC$ occurs.
† Pay-offs: $R = 2$, $S = -1$, $T = 3$, and $P = 0$. 

Number of cooperative*, irreducible, limit ESS state space strategies for the IPD†
are also limit ESSs. A number of new signatures also appear (there are a total of 11 signatures among the strategies in Table 2), and I will mention one of these. The signature \((CC, CD, DD, DD, CC, \ldots)\) differs from Pavlov’s in that it contains two rounds of mutual defection, instead of one, as a punishment sequence for a mistaken defection. One can regard Pavlov and Grim as being at the two ends of a spectrum of signatures, with the length of the \(DD\) punishment sequence varying from one to infinity.

Finally, note that the IPD has many limit ESSs of other kinds, apart from cooperative strategies, as for instance \(AllD\) and a host of “\(AllD\)-like” strategies.

2.3. Randomised Action Rules

Although one cannot rule out the possibility of state space strategies with randomised action rules being limit ESSs for the IPD, I have been unable to find an example. There is an important result in Selten (1983), analogous to the frequently quoted Selten (1980), implying that a limit ESS cannot randomise at a decision point where there is a clear-cut asymmetry between the players. This has consequences for state space strategies (see Theorem 2 in Appendix E), one being that an irreducible limit ESS state space strategy must be pure: when any two states are connected by a sequence of action combinations, each state must contain decision points where a clear-cut asymmetry has emerged.

The best known randomised state space strategy for the IPD is “generous tit-for-tat”, or GTFT (Molander, 1985; Nowak & Sigmund, 1992). GTFT bases its decisions on the action combination in the previous round, so it has a memory horizon equal to one, from which follows that it is an irreducible state space strategy (see Proposition 5 in Appendix D). Thus, Theorem 2 shows that GTFT is not a limit ESS.

The mechanism by which Theorem 2 excludes randomised strategies is that, when small errors of execution of a strategy are introduced, a randomised strategy must have alternative best replies for which Maynard Smith’s second condition becomes an equality, so the strategy can at best be neutrally stable (Maynard-Smith, 1982). This means that Theorem 2 does not conclusively settle the evolutionary outcome; conceivably, randomised strategies could be part of a set of mutually neutral strategies.

However, GTFT is not neutrally stable in this sense. A brief description of the stability properties of GTFT is as follows. GTFT cooperates after a \(CC\) or a \(DC\), and has certain probabilities \(q_{CD}\) and \(q_{DD}\) to cooperate after a \(CD\) and \(DD\). With an appropriate choice of these probabilities, and small errors of execution, one finds that GTFT is a Nash equilibrium, so that a properly defined version of GTFT is trembling-hand perfect (Nowak & Sigmund assumed \(q_{CD} = q_{DD}\), which for general pay-offs prevents trembling-hand perfection). The optimality equation of this GTFT has the rather curious property that both \(C\) and \(D\) are optimal actions in every state, which means that all strategies are best replies to GTFT (this holds also with small errors). As a consequence, all strategies that are limit ESSs by way of Theorem 1 (e.g. all in Table 2) can invade GTFT when there are small errors of execution, since these alternative best replies are then strict Nash equilibria, and will strictly violate the second condition. Thus, although GTFT is an acceptable equilibrium according to classical game theory, from an evolutionary point of view it is of no interest (cf. Nowak & Sigmund, 1993).

3. A Dominance Game

The state space approach can also be applied to repeated games with a role asymmetry. As an example, consider the component game in Fig. 1. There are two roles, dominant and subdominant. In each round the subdominant has an opportunity to appropriate a resource of value \(v\), thus lowering the dominant’s pay-off. The subdominant can either respect (\(R\)) the dominant’s rights to the resource or steal (\(S\)) the resource. The dominant can either leave (\(L\)) the subdominant in peace or punish (\(P\)) the subdominant. Punishing has a performance cost \(c\) for the dominant and inflicts the damage \(d\) on the subdominant. The probability \(\omega\) of continuation to the next round satisfies \(0 < \omega < 1\), and \(0 < c < v < d\) holds for the pay-offs.

The game is a highly idealised portrayal of a
dominance interaction. It is relatively cheap for the dominant to inflict cost on the subdominant through aggression, and the main biological interest of the game is the issue of whether the dominant’s ability to punish leads to control over the resource. This problem has recently been raised by Clutton Brock & Parker (1995), who analysed dominance interactions as games without any explicit time structure.

Considering a single round, it is clear that the dominant’s ability to punish is of no use. The game in Fig. 1 has only one ESS, where the subdominant steals and the dominant leaves it in peace. Thus, we see that \( \omega \) must be sufficiently large for a threat to have any effect.

The initial role assignment produces a clear-cut asymmetry between the players, from which follows that only pure state space strategies can be limit ESSs (Theorem 2 in Appendix E). Let \( l = 1 \) and \( l = 2 \) denote the roles of dominant and subdominant. A pure state space strategy for the game has a (finite) state space \( X_l \) and an action rule \( r_l \) for each of the two roles. Each \( X_l \) has an initial state \( x_0 \) and a dynamics \( x_{\text{next}} = f_l(x, a a') \), where \( a \) is the action of the player in role \( l \) and \( a' \) the other player’s action. Thus, the strategy consists of separate state space representations for each role, and these two components of the strategy are in principle entirely independent of each other. Although a strategy may have several state space representations, there is a unique minimal one, where the number of states in each \( X_l \) is as small as possible (Proposition 4 in Appendix D).

The most important category of state space strategies are the reflexive ones, and for these the two (minimal) state spaces \( X_1 \) and \( X_2 \) are intimately linked. For a reflexive state space strategy there is a one-to-one correspondence, referred to as reflection, between the states in \( X_1 \) and \( X_2 \): if \( x \) is a state in one of the spaces, \( x^R \) is the corresponding state in the other space. The initial states correspond to each other in this way, and when two players of the strategy meet, the state transitions produced by the dynamics \( f_1 \) and \( f_2 \) must be such that the current states of the players stay in reflection correspondence throughout the play of the game.

With \( v_l(aa') \) the pay-off to the player in role \( l \) from a round where \( aa' \) was used, there is then an optimality equation with one component for each role \( l \), given by

\[
W_l(x) = \max_a [v_l(aa') + \omega W_l(x_{\text{next}})],
\]

where \( a' = r_l(x^l) \) with \( l' \) the opposite role from \( l \) and \( x_{\text{next}} = f_l(x, aa') \). Solving the optimality equation one can, just as for the role symmetric case, use Theorem 1 (in Appendix E) to determine if the strategy is a limit ESS.

Reflexivity is conveniently expressed by numbering the states \( x \) in \( X_1 \) in some way, and then using the same numbering of the corresponding states \( x^R \) in \( X_2 \). This has been done with the examples of reflexive strategies shown in Table 3. For the first two, Quarrel and Tyranny, the no-mistake sequence of play is \( (LR, LR, LR, \ldots) \) when the strategy meets itself. For Quarrel, a mistaken \( S \) from the subdominant produces the signature \( (LR, LS, PS, LR, \ldots) \) and a mistaken \( P \) from the dominant gives \( (LR, PR, PS, LR, \ldots) \), so a mistake is followed by a round of both punishment and stealing after which the interaction stabilises. When the damage from punishment is fairly serious \( (d > 2\varepsilon) \) Quarrel is a limit ESS for sufficiently long interactions \( (\omega \) sufficiently

<table>
<thead>
<tr>
<th>Tables of Examples of Reflexive State Space Strategies for the Dominance Game</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy</td>
</tr>
<tr>
<td>Quarrel</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tyranny</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Flip</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Flip-flop</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

* Current state \( x \) for either dominant or subdominant; the initial state is 1.
† State transition produced by the action combination \( a, a' \), where \( a_l \) is the dominant’s action (\( L \) or \( P \)) and \( a_a \) is the subdominant’s action (\( R \) or \( S \)).
close to one), since the optimality equation then has the solution \( W_1(1) = v/(1 - \omega), W_1(2) = -c + ow/(1 - \omega), W_2(1) = 0, W_2(2) = v - d \), with strict optimal actions given by Quarrer’s action rule.

For Tyranny, on the other hand, the dominant is scrupulous in punishing any stealing, resulting in the signatures \((LR, LS, PR, LR, \ldots)\) and \((LR, PR, LR, LR, \ldots)\). Because the subdominant gets a \( P \) for every \( S \), Tyranny is a limit ESS when \( \omega \) is close enough to one. It takes a minimum of three states to implement this type of relationship. For punishment to be optimal there must be at least one state where the subdominant steals, and for scrupulous punishment to be optimal there must be an additional state where the subdominant “accepts” punishment without further stealing, so together with the initial state, where the dominance relationship is in force, we must have at least three states.

Quarrel and Tyranny are just two examples out of many limit ESSs where the dominant uses the threat of punishment to prevent the subdominant from stealing. With \( c = 0.3, v = 1, d = 3, \omega = 0.99 \), and considering strategies with at most three states, there are 14351 irreducible limit ESS state space strategies of this kind. Among these there are 23 distinct signature pairs, differing somewhat in the nature of the quarrel/punishment sequence produced by a mistake.

While dominance is a possibility, there are other types of solutions. The single-shot ESS has a corresponding reflexive state space strategy with one state, where the subdominant steals and the dominant leaves it in peace, and this strategy is always a limit ESS. Thus either the dominant or the subdominant can have the upper hand, and such polarity may be present within a state space strategy. For the example dubbed Flip in Table 3, the subdominant steals in the initial state, and continues to do so unless there is a mistaken \( R \), in which cases the relationship flips over in favour of the dominant. The situation is similar for the Flip-flop strategy, but here a mistaken punishment by the dominant in state 2 causes the subdominant to regain the initiative. Flip and Flip-flop are limit ESSs for \( \omega \) large enough.

4. Alternating Moves

For the kinds of social interactions that one might try to model as repeated games, a requirement of simultaneous moves may sometimes seem artificial. A strict alternation of moves is perhaps also a constrained format, but can often be a more natural modelling approach. For instance, in Trivers’ (1971) treatment of reciprocal altruism, an offer of help is to be reciprocated at some later time, when the previous recipient is in a position to help the previous donor.

Nowak & Sigmund (1994) analysed reciprocal altruism as an alternating Prisoner’s Dilemma, and I will use this game to illustrate state space strategies for alternating moves (see Appendices F to H for a formal development). An alternation of moves implies the presence of a role asymmetry. The leader role, denoted \( l = 1 \), moves in the first round and in subsequent odd-numbered rounds, whereas the follower role, denoted \( l = 2 \), moves in even-numbered rounds. After each round the game ends with probability \( 1 - \omega \) and continues with probability \( \omega \), where \( 0 < \omega < 1 \).

Since there is a role asymmetry it would be natural to let the roles differ in the costs and benefits of cooperation, but for comparison with simultaneous moves, I will assume symmetry in this regard. Let \( e_l(a') \) be the pay-off to the player in role \( l \) from a round where the player in role \( l' \) moved, using the action \( a' \). The action can be either \( C \) or \( D \), and the pay-offs are given by

\[
e_l(C) = -c, \quad e_l(D) = 0, \quad e_l(C) = b, \quad e_l(D) = 0,
\]

where \( 0 < c < b \) and \( l'' \) is used to denote the opposite role from \( l \). Thus, \( c \) is the cost of giving help and \( b \) is the benefit of receiving help. Comparing with the IPD, if the moves were simultaneous we would have had \( R = b - c, S = -c, T = b, P = 0 \), so to an alternating PD there is a “corresponding” simultaneous PD with additive pay-off matrix (cf. Nowak & Sigmund, 1994).

Just as for the dominance game above, the presence of a role asymmetry means that only pure state space strategies can be limit ESSs (Theorem 4 in Appendix H). A pure state space strategy for the game has a state space \( X_l \) and an action rule \( r_l \) for each role \( l \). For the leader role there is an initial state \( x_0 \) in \( X_1 \), but the starting state in \( X_1 \) for the follower role is some function \( \xi_0(a) \) of the leader’s initial action \( a \). To describe the state space dynamics, let \( aa' \) be a sequential action combination, i.e. the action \( a \) in one round followed by the action \( a' \) in the next round. State transitions for the two roles are then written as \( x_{next} = f_l(x, aa') \). Note that a player in role \( l \) is only assigned a state in every second round, when the player moves. As before, a state space strategy has a unique minimal representation, where the number of states in each \( X_l \) is as small as possible.

A simple and important example of a state space strategy for the alternating PD is shown in Table 4. The strategy is essentially a direct implementation of reciprocal altruism as described by Trivers (1971), and for this reason I have named it Reciprocity.
(Table 4 shows the minimal representation of Reciprocity). As leader, a player of Reciprocity starts out helping the partner, and offers help whenever the partner provided help in the previous round. Reciprocity also has a contrition mechanism, similar to the one in CTFT. A player who accidentally defects will accept not being helped in the next round, after which the offering of help is resumed. The signature of reciprocity can be written \((C, D, D, C, \ldots)\), where I assumed the follower to accidentally defect in the second round.

Just as for simultaneous moves, reflexivity turns out to be a crucial property of a strategy. However, when the players are assigned states only in alternating rounds, there is no mapping of a state to its reflected state. Instead, reflexivity is a property of the structure of the state space dynamics (see Appendix G). When two players use the same reflexive strategy, their state spaces are linked to each other in such a way that each action can be seen as causing a transition from the state of one role to a state of the other role. Such transitions can be expressed as mappings \(x' = g(x, a)\), where \(x\) and \(a\) are a state and an action of the player in role \(l\), and \(x'\) is the subsequent state of the player in the other role. These transitions are illustrated for Reciprocity

**Table 4**

<table>
<thead>
<tr>
<th>Role</th>
<th>State*</th>
<th>Action rule</th>
<th>Next state†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CC</td>
<td>CD</td>
</tr>
<tr>
<td>Leader</td>
<td>1</td>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>D</td>
<td>1</td>
</tr>
<tr>
<td>Follower</td>
<td>1</td>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>D</td>
<td>1</td>
</tr>
</tbody>
</table>

* Leader state, \(x = 1, 2\), where \(1\) is the initial state, or follower state, \(x = 1, 2\), where an initial \(C\) by the leader leads to 1 and an initial \(D\) leads to 2.
† State transition produced by sequential action combination \(aa' = CC, CD, DC, \text{ or } DD\).

Reflexivity of the dynamics it is, however, not sufficient that it can be represented in this way. The description of the transition \(x_{next} = f_i(x, aa')\) as going by way of a state in another space must also be maximally economic, i.e. this other space must have as few states as possible for such a description. To see that Reciprocity has this property, look at Table 4 and consider a leader in state 1 having performed \(C\). We wish to characterise the situation with regard to the possible future behaviour of the leader in a complete but maximally economic way. The follower has the options \(C, D\) leading, respectively, to the leader states 1, 2. This pattern is the sought for characterisation, and corresponds to the follower state 1 (see Table 5). Similarly, for a leader in state 1 having performed \(D\), the actions \(C, D\) by the follower puts the leader in states 1, 1, and this pattern corresponds to the follower state 2. For a leader in state 2 having performed either \(C\) or \(D\) we again obtain the pattern where \(C, D\) by the follower produces leader states 1, 2, which then corresponds to the follower state 1. Since two distinct patterns were produced, a follower needs to distinguish only two types of situations to predict what the leader will do in the future, and these situations in fact correspond to the two follower states. Turning the roles around, we similarly see that the leader needs to distinguish two situations to predict what the follower will do in the future. Finally, a state space strategy is reflexive if the dynamics of its minimal representation is reflexive. Since Table 4 shows the minimal representation of Reciprocity, we can conclude that Reciprocity is reflexive.

For a reflexive state space strategy, we have an optimality equation with one component for each role,

\[
W_i(x) = \max_a [e_i(a) + \omega(e_i(a') + \omega W_i(x_{next}))], \tag{3}
\]

where \(a' = r_i(x'), x' = g_i(x, a), x_{next} = f_i(x, aa')\), and \(l'\) is the opposite role from \(l\). Because of the alternation of moves, a two-round step is needed to connect the states of one role. Just as before, the equation is said to have strict optimal actions if there is only one maximising action for each state. For a non-reflexive strategy, there is a very similar optimality equation, with \(W_i(x)\) being defined on a reflected state space (see Appendix H).

With regard to evolutionary stability, the situation is entirely analogous to the one for simultaneous
moves. If the optimality equation has strict optimal actions, then the strategy is a limit ESS if and only if it is reflexive and the optimal action at each \(x\) is given by the action rule \(r(x)\) of the strategy (Theorem 3 in Appendix H).

Analysing the optimality equation for Reciprocity, one finds that for \(\omega > c/b\) it has the solution

\[
W(1) = \frac{c - b}{c - b - \omega^2}, \quad W(2) = \frac{c - b - \omega^2 c}{1 - \omega^2},
\]

with strict optimal actions given by Reciprocity's action rule. Theorem 3 then shows that reciprocity is a limit ESS when \(\omega > c/b\). Similarly, one finds that Reciprocity is not a limit ESS when \(\omega < c/b\), whereas \(\omega = c/b\) is a marginal case not covered by Theorem 3. One readily sees that the range of stability is the greatest possible for a cooperative strategy for the alternating PD: for \(\omega < c/b\) all out defection can invade any cooperative strategy. Comparing with the corresponding simultaneous PD, Reciprocity has the same range of stability as Grim and CTFT (these have the same ranges when the pay-off matrix is additive).

In spite of its naturalness, Reciprocity appears not to have been described previously, and for this reason I have analysed it in some detail. However, there are many other cooperative limit ESSs for the alternating PD, particularly if one includes those where leader and follower roles differ in the structure of their state spaces. For instance, there are (reflexive) limit ESSs where the number of states in \(X_l\) and \(X_r\) differ, but I refrain from giving additional examples.

An interesting and general consequence of reflexivity for alternating moves is that a non-trivial reflexive strategy cannot have a finite memory horizon (Proposition 7 in Appendix G). Intuitively, if my partner only cares about the last \(n\) rounds, I should only care about the last \(n - 1\) rounds when it is my turn to move, and proceeding in this manner one finds a strategy which does not care about the past as the only possibility. Note that Reciprocity has a potentially infinite memory horizon: for a sequence ending in a string of defections, one needs to know the state before the first \(D\) to determine the current state (Table 4). In this way, a player of Reciprocity can keep the “memory of having been played for a sucker” for an arbitrary number of rounds.

So, for alternating moves the practice of restricting the horizon to the most recent rounds is not a good method of analysis, since pure limit ESS state space strategies cannot be found this way (except perhaps for marginal cases). Such a method was used by Nowak & Sigmund (1994), although for a different game than the one I have considered, in that they assumed \(\omega = 1\). Nowak & Sigmund suggested a state space strategy with randomised action rule, having some similarity to “generous tit-for-tat” (GTFT) for the IPD, as a likely outcome of evolution. This strategy takes into account only the two most recent rounds, and cooperates after a \(CC\) or a \(DC\), defects after a \(CD\), and has a probability \(q\) to cooperate after a \(DD\). Because of the role asymmetry such a strategy cannot be a limit ESS (Theorem 4), but it could conceivably be part of a set of mutually neutral strategies. However, the situation is quite analogous to the one for GTFT. With an appropriate choice of \(q\) the “generous” strategy is trembling-hand perfect. Introducing small errors, the set of best replies to this strategy is somewhat less extensive than is the case for GTFT, but the set does include Reciprocity, which is a strict Nash equilibrium for small errors. Thus, with small errors of execution, Reciprocity strictly violates the second ESS condition for the “generous” strategy, and can invade.

5. Discussion

The idea to use a state space to represent a strategy for a repeated game is not new to the treatment here. The common practice of restricting an analysis to strategies taking only the most recent rounds into account can be viewed as a state space approach. There is also a body of work in classical game theory where a strategy is implemented through an automaton (e.g. Rubinstein, 1986; Abreu & Rubinstein, 1988), and this approach is sometimes put under the heading of “bounded rationality.” Building on this work, Nowak et al. (1995) used two-state automata to represent pure strategies taking only the most recent round into account, and performed an evolutionary analysis on this strategy set. Of previous state space approaches, the one closest in spirit to the development here is that by Boyd (1989).

The main addition to previous work in my analysis lies in a thorough application of Selten’s (1983) evolutionary game theory to state space strategies, with a concomitant stringency deriving from the conceptual precision of the extensive form. Although this method of analysis has a cost in complexity, the rewards in generality and forcefulness of results seem well worth the effort.

My use of a state space to represent strategies is not intended as an assumption of limited mental abilities of biological organisms, but rather as a method of unveiling the properties of certain solutions to a game-theoretical problem. Whether these solutions have a particular significance cannot be determined from the formal reasoning involved, but must rest on a non-trivial step of biological interpretation, involving both empirical facts and theoretical evaluation of the consequences of the extreme
simplifications involved. Let me elaborate on this step.

A basic aspect of social interactions extending over time is the opportunity to react to behavioural variability in a social partner. The factors behind such variability could be genetic variation, unavoidable randomness deriving from the imperfection of the real world, or systematic differences between individuals in the fitness consequences of different outcomes of the interaction. Genetic variation is the most fundamental, in that it is a prerequisite for an evolutionary analysis, and in an attempt at parsimony one might then dispense with other types of variability. This was done in the analysis of the IPD by Axelrod & Hamilton (1981), and in much subsequent theoretical work. The approach has been plagued by uncertainties concerning its predictions of evolutionary outcomes (Boyd & Lorberbaum, 1987; Farell & Ware, 1989; Lorberbaum, 1994), but recently these problems seem to be resolved. A class of strategies sharing the properties of being “nice” and “retaliatory”, i.e. cooperating unless provoked by defection, can dominate in a peaceful coexistence when the probability of continuing to the next round is sufficiently close to one (Bendor & Swistak, 1995). The class includes non-reflexive strategies like TFT and some (perhaps all) cooperative limit ESSs.

Whether much of importance is gained by also including the possibility of mistaken actions in a model, an assumption which probably cannot be faulted on empirical grounds, depends on the nature of the changes in predictions produced by such an assumption. Although perhaps somewhat subtle, these changes imply a dramatic shift in perspective. If mistaken actions occur on a regular basis in relationships, a strategy can be regarded as being finely adapted to its current social environment, and not just to a fairly broad and unspecified spectrum of present and past variation in genotypes. Reflexivity expresses the mutual fit to each other of the social partners’ behavioural mechanisms.

Strategies for the IPD are often described and given names that suggest motivation and psychological mechanisms. Usually, these descriptions are little more than explanatory or rhetorical devices, involving no further discussion of the circumstances when such a description is meaningful. In contrast, Trivers (1971) claimed that a number of motivational variables, like moralistic aggression and guilt, which are known to be present in humans, are aspects of adaptations for reciprocal altruism. Thus, Trivers seemed to regard the presence of certain types of motivation as an important empirical support for his evolutionary idea.

As with adaptations in general, an evolutionary treatment of social motivation probably has greatest chance of success when psychological mechanisms can be regarded as tools for an organism to deal with its current environment. In my opinion, the main biological value in modelling social interactions as repeated games of the kind studied here lies in constructing the simplest possible situation where such a perspective is meaningful.

Of the motivational variables mentioned by Trivers (1971), only moralistic aggression and guilt have a possible connection with the IPD. Considering the state space and dynamics of a strategy like CTFT in Table 1, and referring to states 2 and 3 as “guilt” and “moralistic aggression”, we have a simple psychological mechanism with the ability to regulate the relationship in a way that is evolutionarily stable. For the alternating PD, the strategy Reciprocity in Table 4 is even simpler, in that no state is associated with “guilt”; this is because individuals were assigned states only in the rounds when they move. It is quite possible, although not demonstrated in the present analysis, that a similar simple mechanism could be evolutionarily stable in more general situations, for instance with graded variation in the amount of help provided and a less strict format for when it is an individual’s turn to move. A point to note is that if we were to interpret guilt and moralistic aggression in this manner, we would be assuming that they are adaptations to regulate a relationship in the face of purely random events, i.e. events unrelated to any characteristics of the partners.

Another simple mechanism is provided by the spectrum ranging from Pavlov to Grim (Table 1), where a random disturbance results in a cessation of the relationship for a number of rounds. Pavlov has been interpreted as expressing the simple learning rule of win-stay, lose-shift (Nowak & Sigmund, 1993). Although learning is likely to be quite important in real relationships, the interpretation seems inappropriate for the IPD, since in a social environment consisting of another player of Pavlov, there is nothing to learn.

As there are so many limit ESSs for a game like the IPD, one could question the value of focusing attention on just a few of them, since no strong principle picks out one ESS in favour of another. The observation that many strategies share the same signature brings about some reduction, allowing the interpretation of certain aspects of a strategy as behavioural mechanisms that are rarely used and for which our predictions are weak. Some ESSs probably depend on the strict format of the game for their stability, and these would then be of less interest.
ESSs with a particularly simple structure have the advantage of clearly illustrating the outcome of the analysis, but in general, considerations outside the formal analysis are needed to justify why certain ESSs might be more worthy of attention.

While it is clear that simple behavioural mechanisms can regulate noise in a cooperative relationship, or a dominance relationship, in an evolutionarily stable manner, other factors that are absent from a two-player, repeated game may be equally or more important in real social relationships. For instance, mobile organisms have the possibility to choose partners and to terminate relationships. Once in a relationship, variation between individuals in needs and abilities, rather than random errors, may be the main reason for variation in behaviour, leading to communication of private variables. Repeated games are highly idealised models of social interactions. Because of their simplicity, an unusually complete understanding can be achieved.

This study was supported by the Swedish Natural Science Research Council.

REFERENCES


APPENDIX A

Description of a Simultaneous Moves Game

The formal description of a game in the extensive form contains, among other elements, a game tree with nodes and branches. An information set is a set of nodes of the game tree that a player cannot distinguish, and represents the player’s position when about to move. In the description of the game, I do not explicitly represent the game tree. Because of the simple structure, an information set can be identified with the history of moves leading up to it.

In the main, I will use Selten’s (1983) notation, although I follow van Damme (1987) in using transposition to denote the natural symmetry operation of the game.

At the start of the game each of the two players is assigned a role $l$ from the set of roles $L$. We are mainly interested in two cases: a single role, $L = \{l\}$, or two roles, $L = \{l_1, l_2\}$. The assignment $\lambda = (l, l')$ means that player 1 has role $l$ and player 2 has role $l'$. The transposed, or reversed, assignment is defined as $\lambda^T = (l', l)$. The initial role assignment $\lambda$ is randomly chosen from a set $\Lambda = L \times L$ of possible assignments ($\times$ denotes Cartesian product), and $\Lambda$ transposes to itself, $\Lambda^T = \Lambda$. The probability $p(\lambda)$ of $\lambda$ occurring satisfies $p(\lambda^T) = p(\lambda)$. The set $\Lambda$ is such that the role of one player determines the role of the other player: any given role $l \in L$ occurs once for player 1 and once for player 2 among the assignments $\lambda \in \Lambda$. As a consequence, the transposition $l'$ of a role $l$ becomes determined by the requirement that an assignment can be written as $\lambda = (l, l')$.

The examples I considered previously were of this
kind. If there is only one role, like in the IPD, we have \( \Lambda = \{(i, l)\} \) and a role assignment must be symmetric: \( l_i = l_t \). In this case the initial assignment is redundant and could be dropped from the game. A single clear-cut role asymmetry is given by \( \Lambda = \{(i, l), (l, i)\} \), with each of the two assignments having a probability one half of occurring, and with role transposition \( \overline{\ell_i} = l_t, \overline{l_t} = l_i \).

Following the role assignment there are one or more rounds of play. After a round the game continues with probability \( \omega \) and ends with probability \( 1 - \omega \), where \( 0 < \omega < 1 \). In a round the players perform role-specific actions. The finite set of available actions \( a \) in role \( i \) is denoted \( A_i \), and is assumed to contain at least two actions. With the role assignment \( \lambda = (l, l') \), the set \( A_i \) of possible action combinations \( \alpha = (a, a') \) is given by \( A_i = A_i \times A_{l'} \).

The first component of \( \alpha \) is the action by player 1 and the second that by player 2. Transposition of the action combination \( \alpha \in A_i \) is defined as \( \alpha' = (a, a') = (a', a) \in A_l'i \). For brevity, the combination \( (a, a') \) will sometimes be written \( a a' \).

With the role assignment \( \lambda = (l, l') \), the pay-off to player 1 from a round with the action combination \( \alpha \in A_i \), is denoted \( U_{i}(\alpha) \). The pay-off to player 2 from the round is \( e_{i}(\alpha') \), i.e. it is equal to the pay-off player 1 would obtain in the transposed situation.

A history \( h \) records the past sequence of events in the game, at a point when a player is about to move. Initially, a history consists of the role assignment \( h = \lambda \), and after \( t \) rounds it is given by \( h = (\lambda, a_1, \ldots, a_t) \), where each action combination in the sequence is a member of \( A_i \). Let \( H \) be the set of all histories, and \( H_l \) all that start with \( \lambda \). We define transposition of the history \( h \in H_l \) as \( h' = (\lambda, a_1, \ldots, a_t)' = (\lambda', a_1', \ldots, a_t') \in H_{i'} \). In the following, we will often consider extending a history \( h \in H_l \), with an action combination \( \alpha \in A_i \), and the notation \( (h, \alpha) \) is used to indicate the element in \( H_l \) formed in this way.

Although the players move simultaneously and know the history, their positions are regarded as formally distinct in the extensive form. Using \( i \) to denote player number, we can achieve this by forming two copies, \( U_i, i = 1, 2 \) of the set \( H \) of histories. For instance, let \( U_i \) have elements \( u = (h, i) \). With transposition of players, \( i' \), given by \( 1' = 2 \) and \( 2' = 1 \), we define transposition of the history \( u = (h, i) \in U_i \), for player \( i \) as \( u' = (h, i)' = (h', i) \in U_{i'} \). Let \( U_{i'} \) be the set of histories for player \( i \) starting with the assignment that puts the player in role \( l \), i.e. \( \lambda = (l, l') \) for player 1 and \( \lambda = (l', l) \) for player 2. Finally, for \( u = (h, i) \) the notation \( (u, \alpha) \) indicates the extended history \(((h, \alpha), i)\) for player \( i \).

**INFORMATION SETS, CHOICES AND PAY-OFFS**

We are now ready to identify some essential elements of the extensive form of the game. First, \( U_i \) for \( i = 1, 2 \) is the set of information sets \( u \) of player \( i \). We also need the set \( C_i \) of choices of player \( i \). With \( u \) an information set of player \( i \) and \( l \) the role of the player at \( u \), we define \( C_{i, u} = A_i \times \{ u \} \) as the set of choices at \( u \), and \( C_{i} \) is then the union of the \( C_{i, u} \) over \( u \in U_i \). For a choice \( c = (a, u) \in C_i \), the transposed choice is defined as \( c' = (a', u') \in C_{i'} \). The choice \( c \) thus corresponds to taking the action \( a \) at the information set \( u = (h, i) \) and the transposed choice \( c' \) corresponds to taking the same action \( a \) at the transposed information set \( u' = (h', i') \).

The pay-off to player 1 is obtained by summing the pay-off increments \( e_{i}(\alpha) \) over all rounds \( t \) until the game ends, and the pay-off to player 2 is obtained by similarly summing \( e_{i'}(\alpha') \). Since the probability \( \omega \) of continuation is assumed to satisfy \( 0 < \omega < 1 \), the probability that the game ends after a finite number of rounds is equal to one.

In the extensive form, random events are modelled as choices by a special player, \( i = 0 \), having no strategic interests. The sets \( U_0 \) and \( C_0 \) of information sets and choices for this player, together with a transposition operation, are easily constructed, but since the related notation is not needed in the following I refrain from doing so.

It is now straightforward to see that we have a symmetric, extensive form, two-person game, as defined by Selten (1983). The symmetry of the game is the mapping defined by transposition of choices, which then extends to transposition of information sets, histories, etc., as given above.

**STRATEGIES**

A local strategy \( b_i \), of player \( i \) at the information set \( u \in U_i \), is a probability distribution over the choices at \( u, c \in C_i \), and \( b_i(c) \) is the probability assigned to \( c \). With \( l \) the player’s role at \( u \), we can identify the local strategy with a distribution over the set \( A_i \) of available actions and without risk of confusion we can write \( b_i(a) \) for the probability assigned to \( a \in A_i \). The local strategy is pure if the distribution is concentrated on a single action. A behaviour strategy \( b \) of player \( i \) is then a specification of a local strategy \( b_i \) for each of the player’s information sets.

Transposition induces a one-to-one mapping between the behaviour strategies of players 1 and 2; the transposition of a local strategy has the same distribution over available actions and is at an information set with a transposed history. The set \( B \) of behaviour strategies \( b \) of player 1 are then simply
referred to as behaviour strategies for the game, with the understanding that they should be transposed when used by player 2, i.e. $b^T$ is a behaviour strategy of player 2.

**PERTURBED GAME**

Following Selten (1983), a perturbation $\eta$ is a collection of minimum probabilities $\eta_c \geq 0$ for each $c \in C_1 \cup C_2$, satisfying $\eta_c \geq \eta$ and $\sum_{c \in C} \eta_c < 1$. The set $B(\eta)$ consists of the behaviour strategies of player 1 assigning at least probability $\eta_c$ to $c \in C_1$. In a perturbed game, only behaviour strategies in $B(\eta)$ are allowed, and the interpretation is that players may select choices by mistake. A positive perturbance is allowed, and the interpretation is that players may perform at least probability $h$ and interpret it as the probability of actually assigning at least probability $h$. The unperturbed game is regarded as special perturbed game where all $\eta_c$ are zero, denoted $\eta = 0$. The supremum norm, 

$$|\eta| = \sup_{c \in C_1} |\eta_c|,$$

will be used to measure the size of $\eta$.

A $c \in C_1$ corresponds to an action $a$ at an information set $u$, so without risk of confusion we can also denote $\eta_c$ as $\eta_{u,a}(a)$, where $a \in A$, and $l$ is the role of player 1 at $u$. To conveniently represent strategies in $B(\eta)$, we can define $d_a(a'|a)$ for $a', a \in A$, by

$$d_a(a'|a) = \eta_{u,a}(a') \text{ for } a' \neq a, \quad \sum_{a' \in A} d_a(a'|a) = 1,$$

and interpret it as the probability of actually performing $a'$ when intending to perform $a$. The requirement $\sum_{a \in A} \eta_{u,a}(a) < 1$ means that $d_a(a|a) > \eta_{u,a}(a)$. Regarding $d_a(a'|a)$ as a function of $a'$ as a local strategy, the local strategies at $u$ assigning at least probability $\eta_{u,a}(a')$ to $a'$ is the convex hull of the set of the $d_a(a'|a)$. We then have a linear mapping $D_u$ between the set of local strategies $b_u$ at $u$ and the set of local strategies $\tilde{b}_u$ at $u$ assigning at least probability $\eta_{u,a}(a')$ to $a'$, $\tilde{b}_u = D_u b_u$, given by

$$\tilde{b}_u(a') = \sum_{a \in A} d_a(a'|a) b_u(a),$$

where $l$ denotes the role of player 1 at $u$. The mapping $D_u$ is one-to-one; its determinant is readily computed to $(1 - \sum_{a \in A} \eta_{u,a}(a))^{n_l-1} > 0$, where $n_l$ is the number of actions in $A$. Thus, given a behaviour strategy $b \in B$, we can find a corresponding perturbed behaviour strategy $\tilde{b} = D b \in B(\eta)$ by applying $D_u$ at each information set $u$. This mapping

$$D : B \to B(\eta) \quad (A.1)$$

is one-to-one, since it is one-to-one for each $u \in U_1$.

When talking about the mapping $D$, it will be understood that it refers to a perturbance $\eta$ currently under consideration.

Finally, for $b', b \in B$, let $|b' - b|$ be the natural supremum norm: with

$$|b' - b| = \max_{a \in A} |b'_a - b_a|$$

we have

$$|b' - b| = \sup_{a \in A} |b'_a - b_a|.$$

One easily sees that $|D b - b|$ goes to zero with $|\eta|$, and we can in fact give the bound

$$|D b - b| \leq K_i |\eta|, \quad (A.2)$$

where $K_i$ is a positive constant. For instance, with $n_i$ the number of actions in $A_i$ and

$$n = \max_{i \in L} n_i,$$

we can choose $K_i = n - 1$.

**DEFINITION OF LIMIT ESS**

The games with simultaneous moves considered here are infinite, whereas Selten (1983) dealt with finite games. The lack of an upper bound on the number of rounds does, however, not cause any difficulties. Because the probability $\omega$ of continuation is less than one, the expected pay-off of one behaviour strategy against another is well defined as an absolutely convergent sum.

For $b', b \in B$, let $E(b', b)$ be the expected pay-off of using $b'$ against $b$. We need not specify player numbers since, from the symmetry of the game, the expected pay-off to player 1 when player 1 uses $b'$ and player 2 uses $b''$ is equal to the expected pay-off to player 2 when player 2 uses $b''$ and player 1 uses $b$. For $b', b \in B(\eta)$, $b'$ is called a best reply in $B(\eta)$ to $b$ if $E(b', b) \geq E(b'', b)$ for all $b'' \in B(\eta)$. If the inequality is strict when $b'' \neq b'$, the best reply $b'$ is said to be strict in $B(\eta)$.

Selten used the term direct ESS for a strategy satisfying Maynard Smith’s (1982) conditions restricted to the space $B(\eta)$. Thus, $b \in B(\eta)$ is a direct ESS of the perturbed game if, first, it is a best reply to itself in $B(\eta)$ and, second, for every alternative best reply $b'' \in B(\eta)$ the inequality $E(b, b') > E(b'', b')$ holds. Note that zero perturbance is included as a special case. To define a limit ESS for the infinite games considered here, I have interpreted the convergence criteria in Selten’s definition using supremum norm.

**Definition.** A behaviour strategy $b \in B$ is a limit ESS if there is a sequence of direct ESSs $b^i \in B(\eta^i)$ of
perturbed games for which $|\eta^k| \to 0$ as $k \to \infty$, such that $|b^k - b| \to 0$ as $k \to \infty$.

APPENDIX B

Criteria for Limit ESS

In this appendix, the optimality equation for a simultaneous moves game is studied, leading to two propositions about limit ESSs.

Reflection of information sets

With $u = (h, 1)$ an information set of player 1, I will call $u^h = (h^r, 1)$ the reflected information set, which belongs to player 1. During a play of the game, player 2 is at $(h, 2)$ when player 1 is at $u = (h, 1)$. If player 2 uses the behaviour strategy $b \in B$, i.e. uses $b^r$, the local strategy of player 2 is the one indicated by $b^r$ at the information set $(h, 2)$, and it specifies the same distribution over actions as $b$ does at $u^h = (h, 2)^r$. So in this sense, when player 1 is at $u$, the local strategy of player 2 is $b^r$. Strategically, player 1 may “think of” player 2 as being at $u^h$.

DYNAMIC PROGRAMMING

Expected pay-offs and best replies are conveniently computed using dynamic programming, i.e. by solving the optimality equation. The lemmas in this section will be needed for subsequent propositions about limit ESSs. They are of a somewhat technical nature and are variants of well-known results in optimisation theory (Whittle, 1982, 1983).

Value functions

For games with simultaneous moves, a history $h$ defines (the start of) a subgame. Let $v_u(b', b)$, for $b', b \in B$ and $u = (h, 1) \in U_1$, be the expected future pay-off to player 1 in the subgame when player 1 uses $b'$ and player 2 uses $b$. The subgame might not be reached with positive probability when $b', b$ is played, but $v_u(b', b)$ is a well-defined quantity. Note that $v_u(b', b)$ does not depend on the local strategies of $b'$ and $b''$ at information sets outside the subgame. Considering the dependence on $u$, we can regard $v_u(b', b)$ as an element in the set $B$ of real-valued functions on $U_1$, bounded in supremum norm (the set $B$ a Banach space). Since the expected number of rounds until the game ends is $1/(1 - \omega)$, we have

$$|v(b', b)| = \sup_{u \in U_1} |v_u(b', b)| \leq \frac{M}{1 - \omega}.$$  \hfill (B.1)

where $M$ is the maximum of all $|c_i(x)|$. I will call $v(b', b)$ a value function.

From its definition, the value function $v(b', b)$ satisfies the equation

$$v_u = \sum_{a \in A} b_u(a) \sum_{a' \in A} b_{a'}(a') (c(aa') + \omega v_{u(a,a')})$$  \hfill (B.2)

where $l$ and $l'$ are the roles of players 1 and 2 at $u$ ($l'$ is the role of player 1 at the reflected information set $u^r$) and $aa'$ is the action combination $x = (a, a')$. As is well known, the equation has a unique solution in $B$, since the right-hand side of (B.2) defines a contraction operator, say $f$, on $B$; for any $v', v'' \in B$ we have $|fv' - fv''| \leq \omega |v' - v''|$, which means that $f$ has a unique fixed point in $B$; $v = fv$.

The expected pay-off of using $b'$ against $b$ can now be written

$$E(b', b) = \sum_{\lambda \in \Lambda} p(\lambda)v_{U_1\lambda}(b', b),$$  \hfill (B.3)

where $(\lambda, 1)$ is the information set of player 1 immediately after the initial role assignment $\lambda$.

For $b \in B$, define

$$V_u(b) = \sup_{b' \in B} v_u(b', b)$$  \hfill (B.4)

for each $u \in U_1$, and call $V(b)$ a supremal value function. Given a perturbance $\eta$ and the corresponding mapping $D$ in (A.1) we also define, for $b \in B$, a supremal value function

$$\tilde{V}_u(\eta, b) = \sup_{b' \in \mathcal{B}} v_u(Db', Db) = \sup_{\tilde{b} \in \tilde{B}(b)} v_{u}(\tilde{b}', \tilde{b}),$$  \hfill (B.5)

where $\tilde{b} = Db$. Note that for zero perturbance we have $\tilde{V}(0, b) = V(b)$. From (B.1), $|\tilde{V}(\eta, b)| \leq M(1 - \omega)$ so that $\tilde{V}(\eta, b) \in B$.

Optimality equation

Consider a behaviour strategy $b \in B$. For the unperturbed game, we have the optimality equation

$$V_u = \max_{a \in A} \sum_{a' \in A} b_{a'}(a') (c(aa') + \omega V_{u(a,a')})$$  \hfill (B.6)

where $l$ and $l'$ denote the roles of players 1 and 2 at $u$. Also, given a perturbance $\eta$ and the corresponding mapping $D$, we have the optimality equation

$$\tilde{V}_u = \max_{a \in A} \sum_{a' \in A} d_{a'a'}(a') \sum_{a \in A} \tilde{b}_{a'}(a') (c(aa') + \omega \tilde{V}_{u(a,a')})$$  \hfill (B.7)

where $\tilde{b} = Db$.

Lemma B1. For $b \in B$, the supremal value function $V(b)$ in (B.4) is the unique solution in $B$ to the
optimality equation (B.6). Also, the supremal value function \( \hat{V}(\eta, b) \) in (B.5) is the unique solution in \( \mathcal{B} \) to the optimality equation (B.7).

**Proof.** This is a standard dynamic programming result. That \( V(b) \) and \( \hat{V}(\eta, b) \) satisfy (B.6) and (B.7) follows by taking the supremum in (B.2) above. The equation (B.6) is a special case of (B.7), for which \( \eta = 0 \). The right hand side of (B.7) defines a contraction operator \( F \) on \( \mathcal{B} \). For any \( \hat{V}', \hat{V}'' \in \mathcal{B} \) we have \( |\hat{V}' - \hat{F}\hat{V}''| \leq \omega |\hat{V}' - \hat{V}''| \), which means that \( \hat{F} \) has a unique fixed point in \( \mathcal{B} \): \( \hat{V} = \hat{F}\hat{V} \). \( \square \)

Let us write (B.6) as

\[
V_a = \max_{a' \in A} \phi_a(a', b, V),
\]

and (B.7) as

\[
\hat{V}_a = \max_{a' \in A} \phi_a(a', \eta, b, \hat{V}).
\]

Clearly, (B.6') is a special case of (B.7'): \( \phi_a(a, b, V(b)) = \phi_a(a, 0, b, \hat{V}(0, b)) \). Now, given an optimality equation (B.7), with solution \( \hat{V} \), an action \( a \) which maximises \( \phi_a(-, \eta, b, \hat{V}) \), i.e. one for which \( \phi_a(a, \eta, b, \hat{V}) = \hat{V}_a \), is called an optimal intended action at \( u \) or, for the case \( \eta = 0 \), an optimal action at \( u \). We say that the optimality equation has strict optimal intended actions if, for each \( u \), only one intended action is optimal, and that it has uniformly strict optimal intended actions if there is an \( \epsilon > 0 \) such that for each \( u \) there is an intended action \( a \) such that for \( a'' \neq a \) the inequality \( \hat{V}_u \geq \phi_a(a'', \eta, b, \hat{V}) + \epsilon \) holds.

**Best replies**

**Lemma B2.** Consider a perturbation \( \eta \), a strategy \( b \in \mathcal{B} \), and the optimality equation (B.7) for \( \hat{V}(\eta, b) \).

(i) For any \( b' \in \mathcal{B} \) which assigns positive probability only to optimal intended actions, \( b' = Db' \) is a best reply to \( b = Db \) in \( B(\eta) \) and \( v(b', \hat{b}) = \hat{V}(\eta, b) \). (ii) For a \( b' \in \mathcal{B} \) such that \( b' = Db' \) is a best reply to \( b = Db \) in \( B(\eta) \) and a \( u \) which has a positive probability of being reached when \( b', \hat{b} \) is played, the local strategy \( b_r' \) must assign zero probability to non-optimal intended actions at \( u \).

**Proof.** (i) One easily verifies that \( v = \hat{V}(\eta, b) \) solves (B.2) for the value function of the strategy pair \( b', \hat{b} \), i.e. \( v(b', \hat{b}) = \hat{V}(\eta, b) \). From (B.3) and the definition (B.5) of \( \hat{V}(\eta, b) \), it then follows that \( b' \) is a best reply to \( b \) in \( B(\eta) \). (ii) From part (i) we know there is a best reply in \( B(\eta) \) to \( b' \) with value function equal to \( \hat{V}(\eta, b) \). From the definition of the supremal value function, for any \( \hat{b} \in \mathcal{B}(\eta) \) we have \( v_s(b', \hat{b}) \leq \hat{V}_s(\eta, b) \). Suppose \( b_r' \) gives positive weight to a non-optimal intended action at \( u \); we must then have \( v_s(b', \hat{b}) < \hat{V}_s(\eta, b) \). If \( u \) is reached with positive probability, then for those information sets \( u' \) whose history is part of the history of \( u \) the inequality \( v_s(b', \hat{b}) < \hat{V}_s(\eta, b) \) must hold, and (B.3) then shows \( b' \) not to be a best reply, which is a contradiction.

Because of the possibility of unreached information sets, Lemma B2 does not imply that any best reply in \( B(\eta) \) to \( b \) must use only optimal intended actions.

**Lemma B3.** For \( b \in \mathcal{B} \) and a perturbation \( \eta \), use the simplifying notation \( \tilde{V} = \hat{V}(\eta, b) \) and \( V = V(b) \).

(i) There is a positive constant \( K_1 \) such that

\[
|\tilde{V} - V| \leq K_1 |\eta|
\]

for all \( b \in \mathcal{B} \) and perturbances \( \eta \). (ii) There is a positive constant \( K_2 \) such that

\[
\sup_{\eta} \max_{a''} |\phi_a(a'', \eta, b, \tilde{V}) - \phi_a(a'', b, V)| \leq K_2 |\eta|
\]

for all \( b \in \mathcal{B} \) and perturbances \( \eta \).

**Proof.** (i) let \( F \) and \( \hat{F} \) be the operators corresponding to the optimality equations (B.6) and (B.7). We have \( V = FV \) and \( \hat{V} = \hat{F}V \), so that \( |\tilde{V} - V| = |FV - FV| = |F\hat{V} - \hat{F}V + \hat{F}V - FV| \). From the triangle inequality and contraction \( |\tilde{V} - V| \leq |\hat{F}V - FV| + \omega |V - V| \), leading to

\[
|\tilde{V} - V| \leq \frac{1}{1 - \omega} |\hat{F}V - FV|.
\]

Using (A.2) and that \( |\hat{V}| \leq M/(1 - \omega) \) one easily finds the bound

\[
|\phi_a(a'', \eta, b, \tilde{V}) - \phi_a(a'', b, V)| \leq 2nK_1 |\eta| \frac{M}{1 - \omega},
\]

where

\[
n = \max_{i \in I} n_i
\]

with \( n_i \) the number of actions in \( A_i \). This same bound also holds for \( |F\hat{V} - \hat{F}V| \), so we can choose \( K_2 = 2nK_1 M/(1 - \omega)^2 \).

(ii) From part (i) one finds that \( |\phi_a(a'', b, \tilde{V}) - \phi_a(a'', b, V)| \leq 2nK_1 |\eta| \). Together with (B.8), and with \( K_2 = 2nK_1 M/(1 - \omega)^2 \), one then finds that we can choose \( K_1 = K_2 \).

**Lemma B4.** Consider \( b \in \mathcal{B} \) and a positive perturbation \( \eta \). If the optimality equation (B.7) for \( \hat{V}(\eta, b) \) has strict optimal intended actions specified by the (pure) strategy \( b' \in \mathcal{B} \), then \( b' = Db' \) is a strict best reply in \( B(\eta) \) to \( b = Db \).

**Proof.** Part (i) of Lemma B2 shows that \( b' \) is a best reply. For a positive \( \eta \), all information sets have positive probability of being reached when \( b', \hat{b} \) is played. Part (ii) of Lemma B2 then shows that a best reply to \( b \) must specify optimal intended actions at each information set, and these are uniquely given by \( b' \), so \( b' \) is the only possible best reply. \( \square \)
Lemma B5. Consider \( b \in B \) and zero perturbation. If the optimality equation (B.6) for \( V(b) \) has uniformly strict optimal actions specified by the (pure) strategy \( b' \in B \), then there is an \( \epsilon > 0 \) such that for all positive perturbances \( \eta \) with \( |\eta| < \epsilon \), \( b' = Db' \) is a strict best reply in \( B(\eta) \) to \( b = Db \).

Proof. Let the excess \( \epsilon > 0 \) satisfy the condition for uniform strictness for the optimality equation (B.6). From part (ii) of Lemma B3, for \( |\eta| < \epsilon \), \( b' = Db' \) is a non-optimal intended action of the optimality equation. Since \( b \) is played, we must have \( b' \) played, \( b' = Db' \) is a strict best reply in \( B(\eta) \) to \( b = Db \).

Proof. Let the excess \( \epsilon > 0 \) satisfy the condition for uniform strictness for the optimality equation (B.6). From part (ii) of Lemma B3, for \( |\eta| < \epsilon \), \( b' = Db' \) is a non-optimal intended action of the optimality equation. Since \( b \) is played, we must have \( b' \) played, \( b' = Db' \) is a strict best reply in \( B(\eta) \).

**Limit ESS Results**

The two propositions in this section apply to the supremal value function \( V(b) \) in (B.4) and its optimality equation (B.6). The propositions form the basis for Theorem 1 in Appendix E.

**Proposition 1.** For a pure behaviour strategy \( b \in B \), if the optimality equation for \( V(b) \) has uniformly strict optimal actions which are specified by \( b \), then \( b \) is an ESS.

Proof. This follows directly from Lemma B5: for all sufficiently small and positive perturbances \( \eta \), \( b = Db \) is a strict best reply to itself in \( B(\eta) \), and thus a direct ESS.

**Lemma B6.** If \( b \) is a direct ESS of a perturbed game, then any \( u \in U \) is reached with positive probability when \( b \) is played.

Proof. Assume \( b \) a direct ESS and that \( u \in U \) cannot be reached when \( b \) is played. Let \( t_u \) be the round number of \( u \), i.e. the history of \( u \) contains \( t_u \) rounds. Clearly, modifying the strategy of player 1 and/or player 2 at information sets with round numbers greater than or equal to \( t_u \) cannot affect the probability of reaching \( u \). Construct \( b' \) by modifying \( b \) only at \( u \). Since \( u \) cannot be reached when \( b' \) is played, we must have \( E(b', b) = E(b, b) \), and \( b' \) is then an alternative best reply. Furthermore, since \( u \) cannot be reached when either \( b \), \( b' \) or \( b'' \) are played, we must have \( E(b, b') = E(b', b') \), and \( b' \) then violates the second condition, which is a contradiction.

**Lemma B7.** Consider \( b \in B \) and a perturbation \( \eta \). If the corresponding \( \tilde{b} = Db \in B(\eta) \) is a direct ESS of the perturbed game then \( b \) must assign zero probability to non-optimal intended actions of the optimality equation (B.7) for \( \tilde{V}(\eta, b) \).

Proof. For \( \tilde{b} \) a direct ESS, Lemma B6 states that all information sets are reached with positive probability when \( \tilde{b} \) is played, and part (ii) Lemma B2 then implies that \( b \) must assign zero probability to non-optimal intended actions.

**Proposition 2.** A limit ESS \( b \in B \) can use only optimal actions of the optimality equation for \( V(b) \).

Proof. For a limit ESS \( b \in B \), let \( \tilde{b} \in B(\eta) \) be the required sequence of direct ESSs, i.e. with \( |\tilde{b} - b| \to 0 \) and \( |\eta| \to 0 \) as \( k \to \infty \). Let \( b_k \in B \) be the sequence of “unperturbed” strategies corresponding to \( \tilde{b}_k \), i.e. \( \tilde{b}_k = \tilde{D}b_k \) where \( D_k \) denotes the one-to-one mapping \( D \) (see (A.1)) for the perturbation \( \eta_k \). Because of (A.2), the sequence \( b_k \) is convergent and has \( b \) as limit. Now assume that there is an information set \( u \) such that \( b_k(a) > 0 \) for a non-optimal action \( a \) of the optimality equation (B.6) for \( V(b) \). Since \( b_k \) converges to \( b \), we must have \( b_k(a) > 0 \) for \( k \) large enough. Also, from part (ii) of Lemma B3, for \( k \) large enough \( a \) must be a non-optimal intended action of the optimality equation (B.7) for \( \tilde{V}(\eta_k, b_k) \). Lemma B7 then implies \( b_k(a) = 0 \), which is a contradiction.

**APPENDIX C**

**Image Detachment and Limit ESS**

Selten (1983) introduced the concept of an image detached information set, and this concept will be used in the following. An information set \( u \) is image detached if no play intersects both \( u \) and its symmetric image \( u^c \). Thus, an information set \( u = (h, i) \) is image detached if and only if the history \( h \) satisfies \( h^c \neq h \). With a clear-cut role asymmetry, all information sets are then image detached, because \( h^c = \lambda \). With only one role, so that \( h^c = \lambda \), the histories containing at least one asymmetric action combination, i.e. one with \( x^c \neq x \), correspond to image detached information sets. From the definition of reflection in Appendix B, \( u \in U \) is image detached if and only if \( u^b \neq u \).

The importance of the concept comes from Theorem 5 in Selten (1983), which states that a limit ESS must prescribe a pure local strategy at an image detached information set. The result also holds for the games with simultaneous moves considered here, and for completeness I give a proof. The proof uses some of the lemmas in Appendix B.

**Lemma C1.** If \( u \in U \) is image detached and \( b, b' \in B \) satisfy \( b_c = b_c \) for all \( u' \neq u \), then \( v_u(b_c, b) = v_u(b_c, b') \) for all \( b'' \in B \).

Proof. Since \( b_c \) and \( b_c^c \) can differ only at the information set \( u' \), which is not part of the subgame starting with the history of \( u \), the lemma is an immediate consequence of the definition of the value function.

**Lemma C2.** Consider \( b \in B \) and a perturbation \( \eta \). If the corresponding \( \tilde{b} = Db \in B(\eta) \) is a direct ESS of the perturbed game and the information set \( u \in U \) is
image detached, then the local strategy $b_x$ must be pure.

**Proof.** Assume $\hat{b}$ a direct ESS, $u$ image detached, and $b_\nu$ properly randomised. From Lemma B7, the optimality equation (B.7) for $\hat{V}(\eta, b)$ must have more than one optimal intended action at $u$. Construct $\hat{b}'$ by modifying $b$ only at $u$ in a way such that $\hat{b}'$ uses only optimal intended actions at $u$. From Lemma B7, $b$ uses only optimal intended actions, so that $\hat{b}'$ also uses only optimal intended actions. Part (i) of Lemma B2 then shows $\hat{b}' = Db'$ to be an alternative best reply to $\hat{b}$ in $B(\eta)$, and also that $v(\hat{b}', \hat{b}) = v(\hat{b}, \hat{b}) = \hat{V}(\eta, b)$.

Let $t_u$ be the round number of $u$. We wish to determine for which $u' \in U_l$ the equality

$$v_u(\hat{b'}, \hat{b}) = v_u(\hat{b}, \hat{b})$$

(C.1)

holds. Let $t_u$ be the round number of $u'$. Since $\hat{b}_u = \hat{b}_v$, for all $u'$ with $t_u > t_u$, the definition of the value function shows (C.1) to hold for these $u'$. Considering (B.2), since $\hat{b}_u = \hat{b}_v$ for all $u'$ with $t_u > t_u$ and $u' \neq u$, (C.1) then holds also for these $u'$. From Lemma C1, $v_u(\hat{b}, \hat{b}) = v_u(\hat{b}, \hat{b})$ and $v_u(\hat{b'}, \hat{b}) = v_u(\hat{b'}, \hat{b})$, and we know that $v_u(\hat{b}, \hat{b}) = v_u(\hat{b}, \hat{b})$, so (C.1) holds also for $u' = u$. Again considering (B.2), since $\hat{b}_u = \hat{b}_v$ for all $u'$ with $t_u < t_u$, (C.1) then holds also for these $u'$. Thus, (C.1) holds for all $u' \in U_l$, and (B.3) shows that $E(\hat{b}, \hat{b'}) = E(\hat{b}, \hat{b'})$, which violates the second condition, contradicting the assumption that $\hat{b}$ is a direct ESS.

** Proposition 3.** A limit ESS must specify a pure local strategy at an image detached information set.

**Proof.** Just as in the proof of Proposition 2 in Appendix B, let $\hat{b}$ be the required sequence of direct ESSs and $\hat{b}'$ the corresponding sequence of “unperturbed” strategies, which converges to $\hat{b}$. We know from Lemma C2 that $\hat{b}_u$ must be pure at an image detached $u$, so the limit $b_x$ must also be pure. □

**APPENDIX D  
State Space Strategies**

A state space strategy is a behaviour strategy where the set $U_l$ of a player’s information sets is divided into a finite number of classes, such that the “same” local strategy (the same distribution over actions) is used at the information sets in a class. A class then corresponds to a state. Only certain types of divisions into classes are of interest to us here. When the game is played, a player passes through a sequence of information sets, and thus through a sequence of states. To make dynamic programming a powerful tool for determining best replies, state transitions should be Markovian.

**Definition of a State Space Strategy**

Consider the role assignment $\lambda = (l, f')$. Remember that $U_l$ is the part of $U_l$ where player 1 has role $l$. Without risk of confusion, we can say that the local strategies at information sets $u'$, $u \in U_l$, are the same, $b_u = b_v$, if they both specify the same distribution over $A_1$. Also, remember that for $u \in U_l$, and $x \in A_1$, $(u, x)$ denotes the information set in $U_u$ obtained by extending the history of $u$ with the action combination $x$. Finally, given a partition $X_l$ of $U_u$, i.e. given a division of $U_u$ into disjoint, non-empty classes $x \in X_l$, whose union is $U_u$, write $u' \sim u$ for the equivalence relation: $u'$ and $u$ belong to the same class $x \in X_l$.

**Definition.** A state space strategy of player 1 is a behaviour strategy $b$ of player 1 such that for each assignment $\lambda = (l, f')$ there is a finite partition $X_l$ of $U_l$ with the following property: for information sets $u'$, $u \in U_l$, the implications

$$u' \sim u \Rightarrow b_u = b_v$$

(D.1)

and

$$u' \sim u \Rightarrow (u', x) \sim (u, x)$$

for all $x \in A_1$ hold. The partitions $X_l$, as well as their union $X$, are called state spaces, and an element $x \in X_l$ is called a state.

**Partitions $X_l$** satisfying (D.2), without reference to a behaviour strategy, are also called state spaces of player 1. For player 2, state space strategies and state spaces are defined by similarly partitioning $U_u$.

**Action Rules and State Space Dynamics**

Because of (D.1), a state space strategy $b$ of player 1 specifies the same distribution over actions at all $u$ belonging to the same state $x$. The randomised action rule $\rho_l$ is then a mapping from the state space $X_l$ to the set $\Delta(A_1)$ of distributions over the actions in $A_1$,

$$\rho_l: X_l \rightarrow \Delta(A_1),$$

where $\rho_l(x)$ is the distribution specified by $b$ for $u \in x'$; the probability assigned to $a \in A_1$ is written $\rho_l(x, a)$. By the action rule $\rho$ is meant the collection of mappings $\rho_l$, $l \in L$.

For a pure state space strategy, I will introduce a special notation for the action rule. A pure action rule is a mapping,

$$r_l: X_l \rightarrow A_1,$$

which maps a state $x$ to one of the available actions, $a = r(x)$, and $r$ is the collection of the $r_l$.

The implication (D.2) expresses the Markov
property, and means that the partition $X_i$ induces a state space dynamics, i.e. a mapping
\[
f_i: X_i \times A_i \rightarrow X_i,
\]
where $\hat{\lambda} = (\lambda, \Gamma)$. Thus, if the player’s current state is $x \in X_i$ and the action combination $x \in A_i$ is performed in the current round, the player’s next state will be $x' = f_i(x, \lambda)$. The player’s state immediately after the initial role assignment is called the initial state and denoted $x_0$. By the dynamics $f_i$ is meant the collection $f_i, l \in L$, and similarly for $x_0$.

We then say that $(X, \rho, f, x_0)$ is a state space representation of the state space strategy $b$. Strictly speaking, $f$ and $x_0$ are superfluous, since they are determined by $X$, so we can also say that $(X, \rho)$ is a representation of $b$. However, as was done in the examples (e.g., Table 1), it can be convenient to regard the states as abstract points instead of as classes of information sets, and $f$ and $x_0$ can then be used to reconstruct the partitions of the $U_U$.

**UNIQUE MINIMAL STATE SPACE REPRESENTATION**

A state space strategy $b$ will have many state space representations, but among the representations one is more natural than the others. Considering the set of all state space representations of $b$, a representation $(X, \rho)$ is said to be minimal for $l$ if the number of states in $X_l$ is minimal (over all representations), and the representation is minimal if it is minimal for each $l \in L$. As a preliminary to formulating the proposition, with two partitions of the same set, one is said to be a refinement of the other if each of its classes is a subset of one of the classes of the other.

**Proposition 4.** A state space strategy of player 1 has a unique minimal state space representation. The state space of any representation of the strategy is a refinement of the state space of the minimal representation.

**Proof.** The proof hinges on the fact that for two finite partitions with the same number of elements (classes), if one is refinement of the other, they must be equal.

Let $b$ be a state space strategy of player 1 and consider the assignment $\hat{\lambda} = (\lambda, \Gamma)$. Define the sequence $X_{l}^{k}, k = 0, 1, \ldots$, of partitions of $U_U$, with corresponding equivalence relations $\sim_{k}$, by saying that $u' \sim_{0} u$ holds when
\[
b_{u'} = b_u,
\]
and, for $k > 0$, $u' \sim_{k} u$ holds when
\[
u' \sim_{k-1} u \text{ and } (u', \lambda) \sim_{k-1} (u, \lambda) \text{ for all } \lambda \in A_i.
\]
We do not yet know if any of the $X_{l}^{k}$ can be regarded as a (Markov) state space for $b$, but we can clearly define action rules $\rho^{k}_{l}$. Now, let $(X, \rho)$ be a representation of $b$ (since $b$ is a state space strategy, it must have at least one representation). From (D.1), $X_l$ must be a refinement of $X_{l}^{0}, \text{ and from (D.2), given that } X_l \text{ is a refinement of } X_{l}^{k}, \text{ it must be a refinement of } X_{l}^{k+1}. \text{ Thus, induction shows } X_l \text{ to be a refinement of } X_{l}^{k} \text{ for all } k. \text{ Since the partitions } X_{l}^{k} \text{ are successive refinements of each other and } X_l \text{ is finite, there must be a } K \text{ such that } X_{l}^{k+1} = X_{l}^{k} \text{ for } k \geq K, \text{ and then } X_{l}^{K} \text{ is a state space for } b \text{ with action rule } \rho_{l}^{K}. \text{ We also see that } X_l \text{ has at least as many elements as } X_{l}^{K}, \text{ and if it has the same number of elements, it is equal to } X_{l}^{K}. \text{ Using the notation } X_l \text{ and } \rho_{l}^{K} \text{ for } X_{l}^{K} \text{ and } \rho_{l}^{K}, \text{ and going through the argument for each } l \in L, \text{ we find } (X_l, \rho_{l}) \text{ as the unique minimal representation, and also that } X_l \text{ is a refinement of } X_l. \quad \blacksquare

The procedure of successive refinement in the proof is useful as a numerical algorithm for constructing the minimal representation of a state space strategy, i.e. given that a state space strategy $b$ is specified in some manner, for instance by way of a representation, one can compute the minimal representation with this algorithm.

**SYMmetry**

The symmetry of the game expresses that player number (1 or 2) has no strategic relevance. We can speak of a behaviour strategy $b$ for the game, with the understanding that it should be transposed when used by player 2, and we want to do the same for states, action rules, and state space dynamics.

Since state space strategies are behaviour strategies, their transposition is already defined, but to fully appreciate how the states of the two players are related during a play of the game, some more details are helpful.

Let $(X, \rho, f, x_0)$ be the representation of a state space strategy $b$ of player 1. Remembering that a state $x$ is a class of information sets of player 1, we immediately have a definition of $x^T$ as a class of information sets of player 2. Here it should be noted that the set of histories corresponding to $x^T$ might be different from each of the sets of histories that correspond to the states of player 1. In any case, considering the assignment $\hat{\lambda} = (\lambda^T, \Gamma)$, where $\Gamma$ is the role of player 2, $X^T$ with
\[
X_l^T = (X_l)^T,
\]
will be a state space for the behaviour strategy $b^T$ of player 2. So, if $x \in X_l^T$ we have $x^T \in X_l$. Next, the action rule $\rho^T$ on $X^T$ is a collection of mappings
\[
\rho^T_l: X_l^T \rightarrow \Delta(A_l)
\]
\[
    f^\lambda : X^\lambda \times A \to X_{\lambda}^\lambda
\]

satisfying \( f^\lambda(x, z) = (f^\lambda(x^\lambda, z^\lambda))^\lambda \). Similarly, the reflection of the initial states \( x_0 \) is \( x_0^\lambda \) given by \( x_0^\lambda = (x_0^\lambda)^\lambda \). The interpretation of the reflected state space \( X^\lambda \) is that player 1 is “using” the state space of player 2, who is using \( X^\tau \) as defined in the previous section, i.e., player 1 is making the same distinctions as player 2.

**Definition.** A state space \( X \) of player 1 is said to be reflexive if \( X^\tau = X \), and a state space strategy of player 1 is said to be reflexive if the state space of its minimal representation is reflexive.

For a reflexive state space \( X \), reflection of states is then a one-to-one mapping from \( X \) to \( X^\tau \). The state space dynamics of a reflexive state space satisfies \( f^\lambda = f^\lambda \).

The important consequence of reflexivity is that if both players use the same reflexive state space strategy \( b \), i.e., player 2 uses \( b^\tau \), their current states during the play of the game change in parallel, so they can be regarded as making the same distinctions. If \( x^\tau \in X \) is the current state of player 1 and \( x^\tau \in X^\tau \) is the current (player 2) state of player 2, we clearly have \( x^\tau = (x^\tau)^\tau \). We can avoid reference to player number by saying that player 2 is in the state \( x^\tau \). Thus, the states of the two players are related by reflection during a play of the game.

**IMAGE DETACHMENT, IRREDUCIBILITY AND MEMORY HORIZON**

Since a state space strategy uses the same distribution over actions at information sets belonging to the same state, Proposition 3 in Appendix C shows that a limit ESS state space strategy must be pure at a state containing an image detached information set, and it is of interest to determine when this is the case. For a state space \( X \), let \( X^\lambda \subseteq X \) be the set of states containing at least one image detached information set. For asymmetric assignments, i.e. \( \lambda = (i, l^\prime) \) with \( i \neq i^\prime \), all information sets in \( U_{l^\prime} \) are image detached, so \( X^\lambda = X \), but for symmetric assignments \( X^\lambda \) depends on the dynamics of \( X \). To elucidate this dependence I will introduce a few concepts.

A sequence of action combinations \( z = (x_1, \ldots, x_k) \in A^k, \lambda = (i, l^\prime), \) and a state \( x_i \in X \), define a sequence of states by way of the dynamics \( f^\lambda \) of \( X \) for \( x_{k+1} = f^\lambda(x_k, z_k), k = 1, \ldots, K \). We can write the state \( x_{k+1} \) as a function of \( x_k \) and \( z_k \), \( x_{k+1} = f^\lambda(x_k, z_k) \), in this way defining a \( K \)-step state dynamics \( f^\lambda \) for \( X \). For states \( x, x' \in X \), if there is a \( K > 0 \) and an \( z \in A^k \) such that \( f^\lambda(x', z) \) we say that \( x \) can be reached from \( x' \). By the definition of \( X \), all \( x \in X \) can be reached from the initial state \( x_0 \), except possibly \( x_0 \) itself.

A set \( Z \subseteq X \) is said to be invariant under the dynamics \( f^\lambda \) if \( x \in X \) if the image of \( Z \times A \) is in \( Z \). The state space \( X \) is called irreducible if no proper subset \( Z \subseteq X \) is invariant. Irreducibility is equivalent to the requirement that any state in \( Z \) can be reached from any other state: if \( X \) is reducible with \( Z \subseteq X \) invariant, \( Z \times A \) cannot be reached from an \( x' \in Z \); if \( X \) is irreducible, the invariant set \( Z \), consisting of all states that can be reached from \( x \) must be equal to \( X \).
To verify irreducibility, it is enough to verify that the initial state $x_0$ can be reached from any $x \in X_i$.

The set $X''_i$ is invariant: for each $x \in X''_i$ there is an image detached information set $u \in x_i$, i.e. the history $h$ of $u$ satisfies $h \neq h''$, which means that the history $h''$ of $u'' = (u, z) \in x' = f(x, z)$ also satisfies $h'' \neq h''', \ldots$ so that $x' \in X''_i$. Thus, if $X_i$ is irreducible then $X''_i = X_i$. Also, since an invariant set containing $x_0$ must be equal to $X_i$, we have that $X''_i = X_i$ if and only if $x_0 \in X''_i$.

Let us call a state permanent if it contains infinitely many information sets, and otherwise call it temporary. Clearly, the set of all permanent states in $X_i$ form an invariant set. A state space $X_i$ is called permanent if all its states are permanent, and we see that an irreducible $X_i$ is permanent. The initial state $x_0$ again plays a special role. One readily sees that $x_0$ is permanent if and only if $x_0 = \{\\lambda, 1\}\}$, where $(\lambda, 1)$ with $\lambda = (l, l')$ is the information set of player 1 immediately after the initial role assignment: if some sequence of action combinations reaches $x_0$ from $x_0$, each repetition of the sequence gives an additional information set in $x_0$. Since all states in $X_i$ except perhaps $x_0$ can be reached from $x_0$, it then follows that $X_i$ is permanent if and only if $x_0 \neq \{\\lambda, 1\}\}$. For $l = l'$ the information set $(\lambda, 1)$ is not image detached, so for a symmetric assignment, $X''_i = X_i$ implies that $X_i$ is permanent.

For an information set $u$, let $t_u$ denote its round number, i.e. the history of $u$ contains $t_u$ rounds. For $u, u' \in U_i$ with $t_u, t_{u'} \geq K \geq 0$, say that $u$ and $u'$ have the same last $K$ rounds if the two sequences of the last $K$ action combinations of the histories of $u$ and $u'$ are identical. A permanent $x \in X_i$ is said to have finite memory horizon if there is a $K \geq 0$ such that for any information set $u \in x$ with $t_u \geq K$, all information sets $u' \in U_i$ with round number $t_{u'} \geq K$ and the same $K$ last rounds as $u$ also belong to $x$. The smallest such $K$, denoted $K_x$, is called the memory horizon of $x$. If there is no such $K$ for a permanent $x$, the state is said to have infinite memory horizon. For a temporary state $x$ the concept of a memory horizon is not so meaningful; the above definition would formally lead to a finite horizon, since for $K$ greater than the maximum round number of the information sets in $x$ there are no $u \in x$ with $t_u \geq K$. In any case, a state space $X_i$ is said to have finite memory horizon if each permanent $x \in X_i$ has finite memory horizon.

Consider a permanent $x \in X_i$ with finite memory horizon $K \geq 0$. For each $K \geq 0$, there must be some $x' \in X_i$, $z \in A_x$ such that $x = f(x', z)$, and for such an $z$ we must in fact have $x = f(x', z)$ for all $x' \in X_i$. It is now easy to see that a permanent $X_i$ with finite memory horizon must be irreducible: for each $x \in X_i$ and any positive $K \geq K$, there is some $z \in A_x$ such that $x = f(x', z)$ for all $x' \in X_i$, so that all states are connected by a (non-empty) sequence of action combinations. Let me summarise the main conclusions obtained.

**Proposition 5.** Let $X_i$ be a state space of player 1, and consider the conditions: (i) $X_i$ is permanent with finite memory horizon, (ii) $X_i$ is irreducible, (iii) $X''_i = X_i$, and (iv) $X_i$ is permanent. For $l = l'$ the implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) hold. For $l \neq l'$ (iii) holds and the implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iv) hold.

The proposition might seem to be of limited interest since it does not sharply delimit the condition $X''_i = X_i$. However, finite memory horizon is often assumed in the literature dealing with the IPD and irreducibility is a quite natural condition to consider, so for these reasons I have dealt with the matter.

**APPENDIX E**

**Simultaneous Moves: State Space ESS Results**

My analysis of games with simultaneous moves leads to two main results. Theorem 1 characterises the most important class of pure limit ESS state space strategies, namely those where dynamic programming specifies a unique optimal action at every state, and reflexivity turns out to be a crucial property of such strategies. Theorem 2 is essentially an application of Theorem 5 in Selten (1983), and it severely limits the scope for limit ESS state space strategies with properly randomised actions rules.

**Dynamic Programming**

For $b \in B$ a state space strategy with a not necessarily minimal representation $(X, \rho)$, let us now define a supremal value function on the reflected state space $X^B$. For a state $x \in X^B$, let $\lambda = (l, l')$ denote the role assignment, which is the same for all $u \in x$, and let $f^B$ be the dynamics of $X^B$. Then define the supremal value function $W(x)$ for $x \in X^B$ as the solution to the equation

$$W(x) = \max_{\alpha, a' \in A} \sum_{x \in X^B} f(x, a') \rho(x') \times [\epsilon(aa') + \omega W(f^B(x, aa'))]. \quad (E.1)$$

Just as in the development in Appendix B, an action $a \in A$, for which the maximum on the r.h.s. of (E.1) occurs is called an optimal action at the state $x \in X^B$. If, for each $x \in X^B$, there is only one optimal action, the optimality equation (E.1) is said to have strict optimal actions.
Lemma E1. Let \( b \in B \) be a state space strategy with representation \((X, \rho)\). (i) The optimality equation \((E.1)\) has a unique solution \( W \) in the set of bounded real-valued functions on \(X^b\). (ii) The supremal value function \( V(b) \) in \((B.4)\) is equal to \( W \), in the sense that \( V_r(u) = W(x) \) for \( u \in x \). (iii) The optimal actions at \( x \in X\) for the optimality equation \((E.1)\) are the same as the optimal actions at \( u \in x \) for the optimality equation \((B.6)\) for \( V(b) \). (iv) If the optimality equation \((E.1)\) has strict optimal actions then the optimality equation \((B.6)\) for \( V(b) \) has uniformly strict optimal actions.

Proof. (i) The r.h.s. of \((E.1)\) defines a contraction operator. (ii) For an information set \( u \in x \in X^b \), we have \( u^x \in X^b \), so that \( b_{\rho(x)}(a') = \rho_{\rho(x)}(x^a, a') \). For \( u \in x \) we also have \((u, aa') \in f^t(x, a a')\). Thus, defining \( V \) by \( V_r(u) = W(x) \) for \( u \in x \), we get a solution in \( B \) to \((B.6)\), which we know (Lemma B1) is unique and equal to \( V(b) \). (iii) Evident from the proof of (ii). (iv) Since there are only finitely many states in \( X^b \), uniformity follows. ∎

The restriction of \( W \) to \( X^b \) is conveniently written \( W_l \). Note that \((E.1)\) can be seen as a number of separate equations, one for each role \( l \), i.e. one for each assignment \( l = (l, l') \).

Stability Condition for Pure State Space Strategies

For a pure state space strategy, with representation \((X, r)\), we can write the optimality equation \((E.1)\) as

\[
W(x) = \max_{a \in A_l} [v_r(aa') + \omega W(x')] \tag{E.2}
\]

with the notation \( a' = r_r(x^a), x' = f^l(x, a a') \), and where \( l \) is the role of player 1 at \( x \in X^b \).

Theorem 1. For a game with simultaneous moves, let \( b \in B \) be a pure state space strategy with minimal representation \((X, r)\) for which the optimality equation \((E.2)\) has strict optimal actions. Then \( b \) is a limit ESS if and only if, first, \( b \) is reflexive and, second, the optimal action at each \( x \in X^b = X \) is given by \( r_r(x) \).

Proof. From part (iv) of Lemma E1 the optimality equation \((B.6)\) for \( V(b) \) has uniformly strict optimal actions. Propositions 1 and 2 in Appendix B then yield that \( b \) is a limit ESS if and only if the optimal actions are specified by \( b \). Thus, for a reflexive \( b \) the result follows immediately from part (iii) of Lemma E1. Suppose now \( b \) is not reflexive and a limit ESS, i.e. with the optimal action at \( u \in U_l \) specified by \( b_{\rho} \). From part (iii) of Lemma E1, there is then some action rule \( \bar{r} \) on \( X^a \) specifying the optimal action \( \bar{r}(x) \in A_l \) at \( u \in x \in X^a \). Since \( X^a \) is a state space, different from \( X \) but with the same number of states, we see that \((X^a, \bar{r})\) is an alternative minimal representation of \( b \), which contradicts Proposition 4 in Appendix D.

For pure state space strategies, strict optimal actions for the optimality equation \((E.2)\) can often be regarded as a generic case. For instance, for the IPD Theorem 1 will be applicable except for marginal values of the parameters \( \omega, R, S, T, P \), and these are of rather little interest. Nevertheless, for completeness I will briefly discuss what can happen when Theorem 1 does not apply.

Backtracking along the trail of reasoning leading to Theorem 1, one eventually passes Lemmas B4 and B5, so a state space strategy that is a limit ESS by way of Theorem 1 is a strict Nash equilibrium, and thus a direct ESS, of the perturbed game resulting from any sufficiently small, positive perturbation. This means that stability will be quite insensitive to the detailed structure of a perturbation. However, the definition of a limit ESS (Appendix A) only requires that it is a limit of one sequence of direct ESSs of perturbed games, so in general a limit ESS will be sensitive to the structure of the perturbation. As pointed out by Selten (1983), in such a case one needs to evaluate the reasonableness of the type of perturbation leading to stability.

As an illustration, I have investigated two of the examples in Table 1, Pavlov and TFT, at marginal parameter combinations, and found that it was possible to tailor perturbations of the required kind, but these perturbations were fairly unreasonable, with the probability of mistaken defection much smaller than the probability of mistaken cooperation.

Consequences of Image Detachment

According to Proposition 3 in Appendix C, a limit ESS must prescribe a pure local strategy at every image detached information set. A state space strategy prescribes the same distribution over available actions at the information sets belonging to the same state, so if a state \( x \) has at least one image detached information set belonging to it, the action rule of a limit ESS must be pure at \( x \). With \( X^a \) the set of states \( x \in X_l \) containing at least one image detached information set, Proposition 5 in Appendix D directly yields the following:

Theorem 2. For a game with simultaneous moves, let \((X, \rho)\) be a representation of a limit ESS state space strategy \( b \in B \), and consider a role \( l \in L \). (i) For \( x \in X^a \), \( \rho_{\rho(x)}(x) \) must be pure. (ii) If \( l \neq l' \), \( \rho_{l'}(x) \) must be pure. (iii) If \( X_l \) is permanent with finite memory horizon, \( \rho_{l'} \) must be pure. (iv) If \( X_l \) is irreducible, \( \rho_{l'} \) must be pure.

The theorem lists a number of conditions ensuring pureness of the action rule. The conditions are not
mutually exclusive (see Proposition 5), but are given because they can be useful in the analysis of a game.

Theorem 2 does not exclude the possibility of limit ESS state space strategies with randomised action rules. Although I have been unable to find any interesting examples, one can certainly construct trivial examples by resorting to temporary states (defined in Appendix D). For instance, consider a repeated role-symmetric Hawk-Dove game for parameter values where the single-shot ESS is mixed. With a temporary initial state $x_0$ such that all transitions from this state lead to some state $x_1$, it is easy to find a limit ESS where the single-shot ESS is played in the first round. In effect, a preliminary round, the outcome of which is ignored, is tackled on at the beginning of the game. Since $x_1 \in X^u$ must hold, the action rule must be pure in all states except $x_0$.

APPENDIX F

Description of an Alternating Moves Game

The alternation of moves means that there must be at least two roles, and for simplicity I limit myself to exactly two roles: $l = l_1$ means to move first and $l = l_2$ to move second. There are only two possible assignments, $A = \{(l_1, l_2), (l_2, l_1)\}$, and each assignment has a probability one half of occurring. The assignment where player 1 moves first is denoted $\lambda_1 = (l_1, l_2)$ and that where player 2 moves first $\lambda_2 = (l_2, l_1)$, and $l_1 = l_2'$.

Let $A_l$ be the set of available actions in role $l$. In a round of the game only one of the players moves, and $\omega$ is the probability of continuation ($0 < \omega < 1$). A history is given by $h = (\lambda, a_1, \ldots, a_t)$, where $t$ is the number of rounds of the run and $a_{t-1} \in A_{l_1}$, $a_t \in A_{l_2}$, $m = 1, 2, \ldots$. The transposed history is defined as $h^T = (\lambda', a_1, \ldots, a_t)$, i.e. it has the same sequence of actions but the reversed role assignment.

The pay-off to a player in role $l$ from a round where a player in role $l'$ moved is $e_\lambda(a')$, where $l'$ is either $l$ or $l'$ and $a' \in A_{l'}$. So if player 1 is in role $l$, the pay-offs from the round to players 1 and 2 will be $e_\lambda(a')$ and $e_{\lambda T}(a')$.

Let $H$ be the set of all histories and $H_l$ those that start with $\lambda$. Since players move alternately, we can identify an information set $u$ with a history $h$. The set $U_1$ of information sets for player 1 consists of all histories in $H_{l_1}$ with an even number of rounds together with all histories in $H_{l_2}$ with an odd number of rounds. The set $U_2$ of information sets for player 2 consists of the remaining histories, i.e. those in $H_{l_2}$ with an even number of rounds together with those in $H_{l_1}$ with an even number of rounds. Just as for simultaneous moves, we let $U_u$ be the part of $U$, where player $i$ has role $l$.

It will be convenient to use notation like $(u, a)$ and $(u, a, a')$ to indicate the extension of a history with one or more actions. With the assignment $\lambda = (l, l')$ and with $a \in A_1, a' \in A_{l'}$, if $u \in U_1$ then $(u, a) \in U_2$, and if $u \in U_2$, then $(u, a, a') \in U_1$. Also, we have $(u, a)^T = (u^T, a)$. Going one more step, we can achieve harmony with the notation for simultaneous moves by introducing the action combination $\alpha = (a, a')$, where the first component is an action by player 1 and the second one by player 2. At an information set $u \in U_1$, where player 1 moves, we interpret $\alpha$ as the action $a$ by player 1 followed by the action $a'$ by player 2, and at a $u \in U_2$ the order of the actions is reversed. Then, with the assignment $\lambda$ and $\alpha \in A_1 = A_l \times A_{l'}$, if $u \in U_1$ then $(u, \alpha) \in U_2$, and if $u \in U_2$, then $(u, \alpha) \in U_2$. With this interpretation we have $(u, \alpha)^2 = (u^T, \alpha^T)$, where $\alpha^T = (a, a')^T = (a, a')$. For brevity, $\alpha = (a', a)$ is sometimes written as $aa'$.

Finally, defining the set of choices at an information set $u$ of a player in role $l$ as $C_u = A_l \times \{u\}$, we have a symmetric, extensive form, two-person game. Behaviour strategies, and their transposition, are defined just as in Appendix A, and $B$ is again the set of behaviour strategies of player 1. The previous discussion of perturbances and the definition of limit ESS (in Appendix A) apply without change to games with alternating moves.

DYNAMIC PROGRAMMING

Since a history defines (the start of) a subgame, we have a value function $v(b', b)$ for $b', b \in B$, where $v_u(b', b)$ for $u \in U_1$ is the expected future pay-off to player 1 from the subgame starting at $u$ when player 1 uses $b'$ and player 2 uses $b$. With $M$ the maximum of all $|v_u(a)|$, the value function satisfies

$$|v(b', b)| \leq \frac{M}{1 - \omega},$$

(F.1)

corresponding to (B.1), and it is the unique solution in $B$ to the equation

$$v_u = \sum_{a \in A_{l_1}} b_u'(a)[e_\lambda(a)$$
$$+ \omega \sum_{a' \in A_{l'}} b_{u', a}(a')(e_{\lambda T}(a') + \omega v_{u(a', a')})],$$

(F.2)

where $l$ and $l'$ denote the roles of players 1 and 2 at $u$; this equation corresponds to (B.2). The expected
The pay-off of using \( b' \) against \( b \) then becomes

\[
E(b', b) = \frac{1}{2} \left( v_i(b', b) + \omega \sum_{a \in A_i} b_i(a) v_{i_{a', a}}(b', b) \right),
\]

corresponding to (B.3).

The supremal value functions \( V(b) \) and \( \hat{V}(q, b) \) are defined exactly as in (B.4) and (B.5):

\[
V_v(b) = \sup_{b' \in B} v_v(b', b)
\]

and

\[
\hat{V}_v(q, b) = \sup_{b' \in B} v_v(Db', Db) = \sup_{b' \in B} v_v(b', \hat{b}),
\]

where \( \hat{b} = Db \). They satisfy the optimality equations

\[
V_v = \max_{a \in A_1} \left[ e_v(a) + \omega \sum_{a' \in A_i} b_{0a'}(a') (e_v(a') + \omega V_{a, a'}) \right]
\]

and

\[
\hat{V}_v = \max_{a' \in A_1} \sum_{a \in A_i} d_{0a'}(a') \left[ e_v(a) + \omega \sum_{a'' \in A_i} \hat{b}_{0a''}(a') (e_v(a') + \omega \hat{V}_{a, a''}) \right].
\]

It is now straightforward to see that Lemmas B1 to B5 apply also to games with alternating moves, if references to eqns (B.2–B.7) are replaced by references to eqns (F.2–F.7).

**APPENDIX G**

**State Space Strategies for Alternating Moves**

Using the interpretation of \( \alpha \) as a sequential action combination, a state space strategy of player 1 is defined exactly as for simultaneous moves (in Appendix D).

**Definition.** A state space strategy of player 1 is a behaviour strategy \( b \) of player 1 such that for each assignment \( \lambda = (l, l') \) there is a finite partition \( X_l \) of \( U_l \) with the following property: for information sets \( u', u \in U_l \), the implications

\[
u' \sim u \Rightarrow (u', \alpha) \sim (u, \alpha) \text{ for all } \alpha \in A_j\]

hold. The partitions \( X_l \), as well as their union \( X \), are called a state spaces, and an element \( x \in X_l \) is called a state.

The action rule \( \rho \) of a state space strategy \( b \) of player 1 with state space \( X \) is defined just as in Appendix D, and \( (X, \rho) \) is a representation of the strategy. Furthermore, the state space dynamics on \( X_l \) can again be expressed as \( x' = f(x, \alpha) \), where \( \alpha \in A_j \), \( \lambda = (l, l') \), is a sequential action combination. However, player 1 has an initial state only for the assignment \( \lambda_1 \), where player 1 moves first, and this state will be denoted \( x_0 \in X_{l_1} \). With the assignment \( \lambda_2 \), we instead have a mapping, \( \xi_0 : A_{l_1} \rightarrow X_{l_1} \), where \( \xi_0(a) \) is the state of player 1 when player 2 used the action \( a \) in the first round. Thus, a representation of a state space strategy \( b \in B \) can also be written as \( (X, \rho, f, x_0, \xi_0) \), although \( f, x_0 \), and \( \xi_0 \) are determined by \( X \). For the corresponding representation \( (X', \rho', f', x_0', \xi_0') \) of the transposed strategy \( b' \), we find \( X', \rho', \) and \( f' \) exactly as in Appendix D, and we have \( x_0' = (x_0)^{r'} \in (X, \rho, f, x_0, \xi_0)^{r'} = X_{l_1}^{r'} \) and \( \xi_0' : A_{l_1} \rightarrow X_{l_1}^{r'} \) with \( \xi_0'(a) = (\xi_0(a))^{r'} \).

Finally, for a game with alternating moves, a state space strategy of player 1 has a unique minimal representation, i.e. Proposition 4 holds; the proof of Proposition 4 in Appendix D remains valid for alternating moves.

**Reflexivity**

The appropriate definition of reflection of a state space for alternating moves is somewhat different from the one for simultaneous moves. During the play of a game with alternating moves, a round is associated with only one information set, belonging to one of the players, and the distinctions made by the other player at this point must be inferred from the
future. The natural definition of reflexivity is found by inspecting the optimality equation (F.6).

Given a state space $X$ of player 1, i.e. given partitions $X_i$ of $U_i$, with corresponding equivalence relation $\sim$ satisfying (G.2), define the reflected partitions $X^r_i$ of $U_i$ with equivalence relation $\sim_r$ as follows: for $u', u \in U_i$, 
\[ u' \sim_r u \text{ when } (u'^r, a) \sim_r (u^r, a) \text{ for all } a \in A_i, \] (G.3)
Note that for $u \in U_i$, the information set $(u^r, a)$ belongs to $U_i$. By $X^r$ is then meant the union of the $X^r_i$.

**Proposition 6.** Let $X$ be a state space of player 1. (i) $X^r$ is also a state space player 1. (ii) For $u', u \in U_i$, if $u' \sim u$ then $(u'^r, a) \sim_r (u^r, a)$ for all $a \in A_i$. (iii) If $X^r$ is a refinement of $X$ then $X^r = X$. (iv) $X$ is a refinement of the doubly reflected state space $X^{rr} = (X^r)^r$.

**Proof.** Let $\lambda = (I, I')$. (i) Consider $u', u \in X^r_i$. From (G.2), we need to verify that $u' \sim_r u$, i.e. that $(u'^r, a) \sim_r (u^r, a)$ for all $a \in A_i$, implies $(u', a) \sim (u, a)$ for all $a \in A_i$, i.e. implies $(u'^r, a)$ as defined by (G.3) is the same as $(u^r, a)$ for all $a \in A_i$. (ii) From (G.2) we see that $u' \sim u$ implies $(u', a, a') \sim_r (u, a, a')$ for all $a, a' \in A_i$, which can be expressed as $(u'^r, a) \sim_r (u^r, a)$ for all $a \in A_i$. (iii) That $X^r$ is a refinement of $X$ can be expressed as $u' \sim_r u$ using this in part (ii) of the lemma and considering (G.3), one then sees that $u' \sim u \Rightarrow u' \sim_r u$. Thus $u' \sim u \Leftrightarrow u' \sim_r u$, which is the same as $X^r = X$. (iv) Part (ii) of the lemma can be interpreted as stating: $u' \sim r u \Rightarrow u' \sim_r u$. The defining condition (G.3) induces mappings
\[ g^r_i : X^r_i \times A_i \to X^r, \] (G.4)
where for a $u \in X^r_i$ we have $g^r_i(u, a)$ as the state in $X^r_i$ containing $(u^r, a)$. Similarly, part (ii) of Proposition 6 induces mappings
\[ g^r_i : X_i \times A_i \to X^r, \]
where for a $u \in X_i$ we have $g_i(u, a)$ as the state in $X^r_i$ containing $(u^r, a)$. The state space dynamics of $X^r$ is then given by
\[ f^r_i : X^r_i \times A_i \to X^r, \] (G.5)
with $\lambda = (I, I')$ and $f^r_i(x, aa') = g^r_i(g_i(x, a), a')$. Making the reversed composition one finds that the dynamics $f$ of $X$ can be written as $f_i(x, aa') = g^r_i(g_i(x, a), a')$, for $x \in X_i$, $aa' \in A_i$. Taking into account the symmetry of the game, one then sees that the state transitions on $X$ can be regarded as going “by way of” states in $X^r$. Although this is a property of the reflected state space, it does not uniquely characterise $X^r$. Since transitions on $X^r$ can also be regarded as going by way of states in $X$, we would then have $X = X^{rr}$, which does not hold generally. From part (iv) of Proposition 6, states (classes) in $X$ are subsets of states in $X^{rr}$, but $X$ may contain more states than $X^{rr}$. Instead, $X^r$ as defined by (G.3) is the simplest state space with the property of inducing the mappings $g^r_i$ and $g_i$. This point will be brought up again in connection with the memory horizon of a state space. Reflexivity is now defined just as previously.

**Definition.** A state space of player 1 is said to be reflexive when $X^r = X$, and a state space strategy is reflexive if the state space of its minimal representation is reflexive.

For a reflexive state space $X^r_i = X_i$, so that $g^r_i = g_i$ holds. Thus, we have the two mappings $g_i$ and $g^r_i$:
\[ g_i : X_i \times A_i \to X^r, \]
and for $x \in X_i$, $a = aa' \in A_i$,
\[ x' = f_i(x, aa') = g^r_i(g_i(x, a), a'). \]
Concerning initial states, the mapping $\xi_0$ can be written $\xi_0(a) = g_i(x_0, a) \in X_i$. So, when two players using the same reflexive state space $X$ meet, a play of the game can be seen as an alternation of states in $X_i$ and $X^r$ (note that $(X^r)^r = X^r_i$), produced by the mappings $g_i$.

**REFLEXIVITY AND MEMORY HORIZON**

The concept of memory horizon was introduced for games with simultaneous moves in Appendix D, and was shown to be of interest in relation to image detachment. For alternating moves, the memory horizon instead has a connection with reflexivity.

The definitions of an invariant set, irreducibility, and permanence introduced in Appendix D apply also to alternating moves. In particular, a state $x \in X_i$ is called permanent if it contains infinitely many information sets. The concept of a memory horizon differs slightly in that only one player moves during a round. For $u, u' \in U_i$ with round numbers $t_u, t_{u'} \geq K$, say that $u$ and $u'$ have the same last $K$ rounds if the two sequences of the last $K$ actions of the histories of $u$ and $u'$ are identical. As before, a permanent $x \in X_i$ is then said to have finite memory horizon if there is a $K \geq 0$ such that for any information set $u \in x$ with $t_u \geq K$, all information sets $u' \in U_i$ with $t_{u'} \geq K$ and the same $K$ last rounds as $u$ also belong to $x$. The smallest such $K$, again denoted
$K_\ell$, is the memory horizon of $x$, and if there is no such $K$ for a permanent $x$, the state is said to have infinite memory horizon. A state space $X_\ell$ has finite memory horizon if each permanent $x \in X_\ell$ has finite memory horizon, and the horizon $K_\ell$ of $X_\ell$ is the maximum of the $K_\ell$ over the permanent states $x \in X_\ell$.

**Proposition 7.** For a game with alternating moves, let $X$ be a state space of player $1$, with components $X_l$ and $X_r$. If at least one $X_l$ has finite memory horizon $K_l = 0$, $X$ cannot be reflexive.

**Proof.** Suppose $X$ is reflexive and $X_l$ has finite memory horizon $K_l > 0$. Let $Z_l \subseteq X_l$ be the set of permanent states in $X_l$. One readily sees that $Z_l$ is invariant, and that $g_l$ in (G.6) maps $Z_l \times A_l$ into (in fact onto) $Z_l$: if $x \in X_l$ is permanent then $g_l(x, a) \in X_l$ is permanent. Since the relations $\sim$ and $\sim$ are the same, applying (G.3) for $l = l'$ shows that $X_l' = X_r'$ has a finite memory horizon $K_l' \leq K_l - 1$. Again applying (G.3), this time with $l = l'$, yields $K_l \leq K_l' \leq K_l - 1$, which is a contradiction. $\square$

Note that we can have $X$ reflexive when each $X_l$ has memory horizon zero; in this case each $X_l$ consists of a single state. State spaces with finite positive memory horizon illustrate the point made earlier, that the doubly reflected state space $X_{\omega \omega}$ might contain fewer states than $X$. For instance, consider the state space $X_l$ given (somewhat loosely) by the actions in the two most recent rounds, which has a finite horizon equal to two. Because of the alternation of moves, the reflected state space $X_l'$ is then given by the action in the most recent round, i.e. it has horizon one, and $X_{\omega \omega}$ similarly has horizon zero, and thus consists of a single state.

**APPENDIX H**

**Alternating Moves: State Space ESS Results**

The main results for alternating moves closely parallel those for simultaneous moves: Theorems 3 and 4 below correspond to Theorems 1 and 2 in Appendix E. Only pure strategies can be limit ESSs, and reflexivity is again a crucial condition for the most important class of limit ESSs. As a point of special interest, state space strategies with a finite memory horizon cannot be reflexive (Proposition 7 in Appendix G), so restricting attention to strategies taking only the most recent moves into account is not a useful method of analysis for alternating moves.

**Dynamic Programming**

In analogy with the development in Appendix E, for $b \in B$ a state space strategy with representation $(X, \rho)$ let us define a supremal value function on the reflected state space $X^\omega$. For a state $x \in X^\omega$, let $\lambda = (l, l')$ denote the role assignment. Define the supremal value function $W(x)$ for $x \in X^\omega$ as the solution to the equation

$$W(x) = \max_{a \in \Delta_l} \left[ e_r(a) + \omega \sum_{a' \in \Delta_{l'}} \rho_{l'}(g^\lambda_l(x, a)) e_r(a') + \omega W(x') \right] + \omega W(f_l^\lambda(x, aa')),$$

(H.1)

where $g^\lambda_l$ and $f_l^\lambda$ are given in (G.4) and (G.5); note that $g^\lambda_l(x, a) \in X_l'$ and $f_l^\lambda(x, aa') \in X_l'^\omega$. Equation (H.1) corresponds to (E.1), and we can again define optimal and strict optimal actions. A result corresponding to Lemma E1 also holds for alternating moves, with only a slight change of the proof:

**Lemma H1.** Let $b \in B$ be a state space strategy with representation $(X, \rho)$. (i) The optimality equation (H.1) has a unique solution $W$ in the set of bounded real-valued functions on $X^\omega$. (ii) The supremal value function $V(b)$ in (F.4) is equal to $W$, in the sense that $V_b(x) = W(x)$ for $x \in X$. (iii) The optimal actions at $x \in X^\omega$ for the optimality equation (H.1) are the same as the optimal actions at $u \in x$ for the optimality equation (F.6) for $V(b)$. (iv) If the optimality equation (H.1) has strict optimal actions then the optimality equation (F.6) for $V(b)$ has uniformly strict optimal actions.

**Proof.** (i) The right hand side of (H.1) defines a contraction operator. (ii) For an information set $u \in x \in X^\omega$, we have $(u', a) \in g^\lambda_l(x, a)$, so that $v_{u', o}(a') = \rho_{l'}(g^\lambda_l(x, a), a')$. For $u \in x$ we also have $(u, aa') \in f_l^\lambda(x, aa')$. Thus, defining $V$ by $V_b(x) = W(x)$ for $x \in X$, we get a solution in $B$ to (F.6), which we know is unique and equal to $V(b)$. (iii) Evident from the proof of (ii). (iv) Since there are only finitely many states in $X^\omega$, uniformity follows. $\square$

**Stability Condition for Pure State Space Strategies**

For a pure state space strategy with representation $(X, r)$ we can write the optimality equation (H.1) as

$$W(x) = \max_{a \in \Delta_l} \left[ e_r(a) + \omega e_r(a') + \omega W(x') \right] + \omega W(x'),$$

where $\lambda = (l, l')$ is the role assignment at $x \in X^\omega$.

In complete analogy with Theorem 1 in Appendix E we then have the following:

**Theorem 3.** For a game with alternating moves, let $b \in B$ be a pure state space strategy with minimal representation $(X, r)$ for which the optimality equation (H.2) has strict optimal actions. Then $b$ is a
limit ESS if and only if, first, \( b \) is reflexive and, second, the optimal action at each \( x \in X^b = X \) is given by \( r_t(x) \).

Proof. From part (iv) of Lemma H1 the optimality equation (F.6) for \( V(b) \) has uniformly strict optimal actions. Propositions 1 and 2 in Appendix B, which hold also for alternating moves, then yield that \( b \) is a limit ESS if and only if the optimal actions are specified by \( b \). Thus, for a reflexive \( b \) the result follows immediately from part (iii) of Lemma H1. Suppose now \( b \) is not reflexive and a limit ESS, i.e. with the optimal action at \( u \in U_1 \) specified by \( b_u \). From part (iii) of Lemma H1, there is then some action rule \( \tilde{r} \) on \( X^b \) specifying the optimal action \( \tilde{r}(x) \in A_i \) at \( u \in x \in X^b \). From part (i) of Proposition 6 in Appendix G, \( X^b \) is a state space, so that \( (X^b, \tilde{r}) \) is an alternative representation of \( b \), and Proposition 4 in Appendix D states that \( X^b \) is a refinement of \( X \). From part (iii) of Proposition 6 we then conclude that \( X^b = X \), which is a contradiction. \( \square \)

CONSEQUENCES OF IMAGE DETACHMENT

As noted in Appendix F, for a game with alternating moves all information sets are image detached, and from Proposition 3 in Appendix B, which holds also for alternating moves, we then know that a limit ESS must be pure. Of course, this applies also to the special case of state space strategies.

Theorem 4. For a game with alternating moves, the action rule of a limit ESS state space strategy must be pure.