

Mistakes Allow Evolutionary Stability in the Repeated Prisoner's Dilemma Game

ROBERT BOYD

*Department of Anthropology, University of California, Los Angeles,
California 90024, U.S.A.*

(Received 8 April 1988, and accepted in revised form 5 August 1988)

The repeated prisoner's dilemma game has been widely used in analyses of the evolution of reciprocal altruism. Recently it was shown that no pure strategy could be evolutionarily stable in the repeated prisoner's dilemma. Here I show that if there is always some probability that individuals will make a mistake, then a pure strategy can be evolutionarily stable provided that it is "strong perfect equilibria" against itself. To be a strong perfect equilibrium against itself, a strategy must be the best response to itself after every possible sequence of behavior. I show that both unconditional defection and a modified version of tit-for-tat have this property.

1. Introduction

Co-operative relationships between unrelated individuals are an important feature of the behavior of social mammals. Several authors have argued that reciprocal co-operation is likely to evolve whenever pairs of individuals interact over a lengthy period of time (Axelrod & Hamilton, 1981; Axelrod, 1984; Brown *et al.*, 1982; Peck & Feldman, 1985). This conclusion is based, in part, on an analysis of the repeated prisoner's dilemma game which indicates that behavioral strategies that lead to reciprocal co-operation are evolutionarily stable (Maynard Smith, 1982). Recently, Jeffery Lorberbaum and I (Boyd & Lorberbaum, 1987) showed that no pure strategy could be evolutionarily stable in the presence of arbitrary patterns of recurrent mutation. Farrel & Ware (in press) have extended this result, showing that no finite mixture of strategies can be stable either. This result is quite discouraging because it suggests that the nature of the strategies that will be able to persist in real populations depends on the distribution of rare phenotypic variants actually produced by mutation or environmental variation. Thus, predictions about social behavior depend on factors about which we usually have little or no knowledge. All of this analysis assumes that individuals never make mistakes. Here I extend the work of Sugden (1986), to show that if individuals sometimes intend to co-operate, but instead mistakenly defect or *vice versa*, then pure strategies, including those leading to reciprocal co-operation, can be evolutionarily stable.

Evolutionary analyses of the repeated prisoner's dilemma assume that pairs of individuals are sampled from a population and interact t or more times with probability w^t . Each interaction, or "turn", is a single period prisoner's dilemma. An individual is characterized by an inherited strategy that determines whether it

co-operates (*C*) or defects (*D*) depending on the sequence of interactions up to that point. Such strategies can be simple unconditional rules like always defect (*ALLD*), or they can be contingent rules like tit-for-tat (*TFT*) which co-operates on the first move and then copies the other player's behavior on the previous move. Strategies are also categorized as pure or mixed. The behavior of an individual using a pure strategy is uniquely determined by the history of the pair's interactions up to that turn, while an individual using a mixed strategy may choose from several behaviors at random. Here I will restrict attention to pure strategies. The incremental effect of a single interaction on the fitness of an individual is given in Table 1. An individual's expected fitness depends on its own strategy and the strategy of the individual with which it interacts. After social interaction, individuals reproduce. The representation of a particular strategy in the next generation is proportional to the average fitness of all individuals using that strategy.

TABLE 1

Each player has the choice of two strategies, C for co-operate and D for defect. The pairs of entries in the table are the payoffs for players 1 and 2 respectively associated with each combination of strategies. In the case of Prisoner's dilemma it is assumed that $T > R > P > S$, and $2R > S + T$

		Player 2	
		C	D
Player 1	C	R, R	S, T
	D	T, S	P, P

To see why no pure strategy can be evolutionarily stable, consider a population in which *ALLD* is common and two other strategies are maintained at low frequencies by recurrent mutation: tit-for-two-tats (*TF2T*) which allows two consecutive defections before retaliating, and suspicious tit-for-tat (*STFT*) which defects on the first move and then plays tit-for-tat. *ALLD* will be evolutionarily stable as long as it has higher average fitness than either rare strategy. *ALLD* will have higher average fitness than *STFT* when:

$$\begin{aligned}
 & p_{ALLD}[V(ALLD|ALLD) - V(STFT|ALLD)] \\
 & + p_{TF2T}[V(ALLD|TF2T) - V(STFT|TF2T)] \\
 & + p_{STFT}[V(ALLD|STFT) - V(STFT|STFT)] > 0
 \end{aligned}
 \tag{1}$$

where p_{ALLD} , p_{STFT} , and p_{TF2T} are the frequencies of the three strategies. Suspicious tit-for-tat co-operates only if its opponent co-operates first. This means that when *STFT* is paired with *ALLD*, both individuals defect forever. The same thing happens

when *STFT* is paired with *STFT*. Thus $V(ALLD|ALLD) = V(STFT|ALLD) = V(ALLD|STFT) = V(STFT|STFT)$, and, therefore, eqn (2) reduces to:

$$V(ALLD|TF2T) > V(STFT|TF2T). \quad (2)$$

ALLD is evolutionarily stable only if it does better against *TF2T* than does *STFT*. For large values of enough w , this condition will be violated because *STFT* gets the benefits of long term co-operation when interacting with *TF2T* while *ALLD* gets only a short term advantage from its initial defection.

This example illustrates a more general principle: no pure strategy can be evolutionarily stable because no pure strategy is the *uniquely* best reply to itself. For any given strategy E , there is always some other strategy I that responds to E the same way that it does to itself. This means that the relative fitness of E and I will depend on how they do against other strategies that cause them to behave differently. When w is large enough, no strategy can be best against every possible third strategy (Axelrod, 1984). Thus, no pure strategy can be evolutionarily stable against every mix of strategies. (For a proof of this assertion see Boyd & Lorberbaum, 1987. Farrel & Ware (in press) show that this conclusion can be extended to finite mixtures of pure strategies.)

2. The Effect of Mistakes

All this analysis assumes there are no mistakes. When an individual's strategy calls for co-operation, it actually co-operates, and when its strategy calls for defection, the individual defects. In real world situations it seems likely that individuals will sometimes make mistakes—individuals who intend to co-operate will instead defect, and individuals who intend to defect will co-operate. For example, an animal trying to help another in a coalition may make a tactical error that causes its support to be ineffective. I will model mistakes by assuming that there is some probability that an individual who means to co-operate defects instead, and some probability that an individual who means to defect actually co-operates. The probability of a mistake may depend on the turn, or the sequence of behaviors up to that turn, but not on the identity of the player.

If individuals make mistakes, pure strategies can be evolutionarily stable because certain patterns of mistakes allow a strategy to be the uniquely best reply against itself. In what follows, I will first show that if a strategy E is what I will call a "strong perfect equilibrium" against itself, and if there is a positive probability of both types of mistakes (C for D and D for C) on every turn regardless of the sequence of behaviors up to that point, then for every possible distinct strategy I , $V(E|E) > V(I|E)$, or, in other words, E is the uniquely best reply against itself. Next, I will show that when this is the case, E is evolutionarily stable against any mix of invading pure strategies provided that they are sufficiently rare. Finally, I will show that two simple strategies, *ALLD* and a modified version of tit-for-tat proposed by Sugden (1986) are evolutionarily stable under the right conditions.

It is useful to describe games like the repeated prisoner's dilemma using a game tree. Each node in the tree represents a point at which the two players choose to co-operate or to defect. Each combination of choices leads to one of four "successor" nodes depending on whether zero, one, or two mistakes are made. The interaction between a particular pair of individuals is described as a path through the tree. A strategy specifies what action an individual should take at each node in the tree. Thus, in a world without mistakes, a particular pair of pure strategies will always lead to the same path through the tree. When there are mistakes, however, the same pair of strategies can generate a variety of paths depending on the pattern of mistakes made during a particular interaction. Each node can be thought of as the beginning of a new game—called the "subgame" which begins at that node.

A pure strategy E can be the uniquely best reply to itself only if the pattern of errors guarantees that there is a positive probability of reaching every node in game tree. Suppose that this is not the case. Consider a strategy I , that behaves exactly the same as E at every reached node, but differently at one of the unreached nodes. I is distinct, but has exactly the same expected payoff against E as E has against itself. Thus E cannot be the uniquely best reply against itself. Any pattern of errors that insures that there is a positive probability of both types of errors (C for D and D for C) at every node in the game tree will lead to a finite probability of reaching every node for every pair of interacting strategies.

Next, I will specify what needs to be true for a strategy to be a strong perfect equilibrium against itself. Suppose that you know that your opponent is committed to using a particular strategy E , and you have the task of designing an optimal counter strategy. Because your opponent's strategy is fixed, this is just a dynamic optimization problem. One way to solve such problems is using dynamic programming. Using this technique you would determine the decision (C or D) at each node that will lead to the highest expected pay-off in all succeeding nodes. Thus you would determine an optimal decision at every node whether or not that node is ever reached. Suppose that having done this you find that at every node (i) the optimal action has strictly greater expected payoff than the alternative action, and (ii) the optimal action is the same action specified by E . If this were the case, E is a strong perfect equilibrium against itself. (Game theorists in economics make extensive use of the concept of a perfect equilibrium which requires only that the optimal action at each node have no worse pay-off than alternative actions. In game theoretic terminology, a perfect equilibrium is a Nash equilibrium in every subgame. I use the modifier "strong" to indicate that here I require that every subgame be a strong Nash equilibrium. See Rasmusen (in press) for a very clear discussion of the concept of perfection.)

If a strategy E is a strong perfect equilibrium against itself, and if there is a positive probability of both kinds of mistakes at every node in the game tree, then for every possible distinct strategy I , $V(E|E) > V(I|E)$. Because there is a positive probability of both kinds of errors, any pair of strategies must have some probability of reaching every node in the game tree. Any distinct strategy, I , must behave differently than E at some node in the tree. Since E is a strong perfect equilibrium against itself, I 's behavior must lead to a lower expected pay-off in the subgame

beginning at that node than E 's behavior. Thus E is the uniquely best reply against itself.

Any strategy that is the uniquely best reply against itself can resist invasion by any combination of strategies when there are mistakes if the invading strategies are sufficiently rare. Consider a population in which there is a common strategy I_0 , with frequency p_0 and n invading strategies I_1, I_2, \dots, I_n with frequencies p_1, p_2, \dots, p_n . The common strategy can resist invasion by the rare strategies if it has higher expected fitness than each of them, or, for each j :

$$p_0[V(I_0|I_0) - V(I_j|I_0)] + \sum_{k=1}^n p_k[V(I_0|I_k) - V(I_j|I_k)] > 0. \quad (3)$$

If I_j is distinct from I_0 , and if I_0 is the uniquely best reply against itself, then the first term on the left hand side of this inequality is positive for every possible strategy I_j . There is no guarantee about the sign of the terms in the summation. However, if the p_k are sufficiently small, the inequality will be satisfied in any case.

It might seem that it would be nearly impossible for a strategy to be a strong perfect equilibrium against itself. However, even the very simple strategy, *ALLD*, has this property when errors lead to every sequence of behavior. Suppose that you know that your opponent is playing *ALLD*, and you are given the task of designing an optimal counter strategy. An *ALLD* individual will defect, no matter what you do. Nothing in your behavior can induce an *ALLD* individual to co-operate. The definition of the prisoner's dilemma guarantees that you are better off defecting whether your opponent co-operates or defects. Thus the best reply to *ALLD* is to always defect. Moreover, every distinct strategy has a lower expected pay-off. To be distinct, there must be some sequence of behaviors that would induce it to co-operate. Since every sequence of behavior can occur as a result of a sequence of errors, such a strategy must have a lower expected fitness against *ALLD* than *ALLD* itself (see Sugden, 1986, p. 109, for a similar argument).

To show that reciprocity can be evolutionarily stable, we must find a reciprocal strategy that is a strong perfect equilibrium against itself when there is a positive probability of each type of mistake on each turn. Consider the following ingenious modification of tit-for-tat proposed by Sugden (1986, p. 110). This strategy, which I will label "contrite tit-for-tat (*CTFT*)", depends on the notion of good standing. An individual is always in good standing on the first turn. It remains in good standing as long as it co-operates when *CTFT* specifies that it should co-operate. If an individual is not in good standing it can get back in good standing by co-operating on one turn. Then *CTFT* specifies that an individual should co-operate (i) if it is not in good standing, or (ii) if its opponent is in good standing; otherwise the individual should defect. To see why I call this contrite, consider the sequence of behaviors following a mistake by one member of a pair both playing *CTFT*. On turn t both players 1 and 2 intend to co-operate, but player 1 mistakenly defects. On turn $t+1$, player 2 is in good standing and player 1 is not. Thus player 1 co-operates, and player 2 defects. On turn $t+2$, player 2 is still in good standing despite his defection on the previous turn because *CTFT* called for defection. Thus player 1 absorbs the sucker's pay-off as an act of contrition.

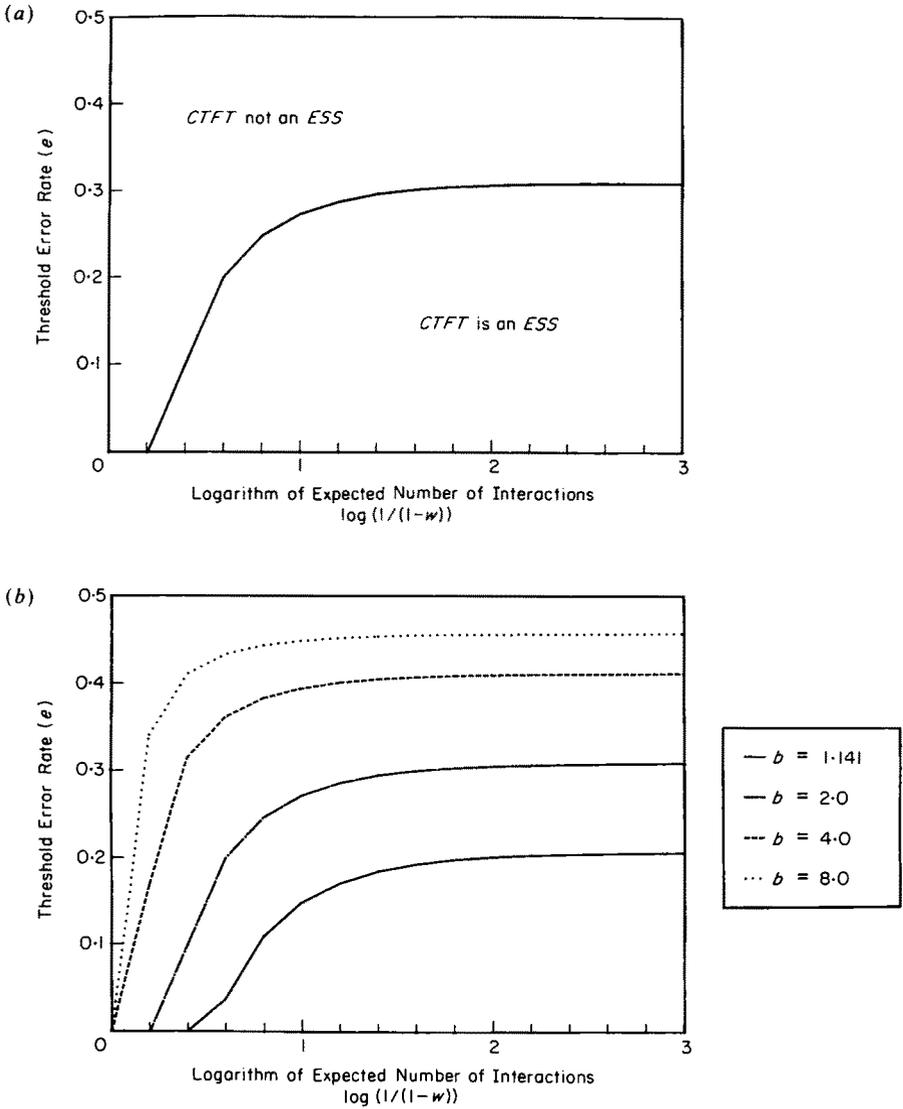


FIG. 1. The threshold error rate at which *CTFT* becomes evolutionarily unstable is plotted as a function of the logarithm (base 10) of the expected number of interactions. It is assumed that $T = b$, $R = b - 1$, $P = 0$, and $S = -1$. Part (a) shows the region of evolutionary stability of *CTFT*, and part (b) shows how the shape of this region is affected by changing the benefit resulting from co-operative behavior. Note that as co-operation becomes more beneficial, *CTFT* is evolutionarily stable for a wider combination of parameter values.

Sugden (1986) argues that *CTFT* is the best reply to itself, and therefore is evolutionarily stable. However, his proof does not allow for the possibility that errors generate every sequence of behavior, and depends on the assumption that the probability of mistakes is arbitrarily small. Here I require that the mistakes are sufficiently likely that they have a much greater effect on the fitness than encounters with rare invading types. In the appendix, I extend Sugden's proof to show that if the constant independent probability e of each type of mistake is less than a threshold value, and

$$w > \max \left\{ \frac{P-S}{R-S}, \frac{T-R}{R-S} \right\}, \quad (4)$$

then *CTFT* is the uniquely best reply against itself. If T , R , P , and S satisfy the definition of a prisoner's dilemma, this inequality is satisfied for w sufficiently close to one. The threshold error rates are shown in Fig. 1 as a function of the logarithm of the expected number of interactions for several different pay-off matrices. Notice that the range of error rates that allow *CTFT* to be evolutionarily stable increases with the expected number of interactions, and that *CTFT* can be stable even if error rates are quite substantial, particularly if mutual co-operation results in sizable long term benefits.

3. Discussion

Mistakes allow evolutionary stability when two conditions are met: there must be a positive probability of both types of mistakes at every node in the game tree, and there is some strategy that is the best reply to itself at every node in the game tree. Notice that these are sufficient conditions for evolutionary stability. Evolutionary stability may be possible under less stringent conditions.

These two conditions do not seem particularly restrictive. In real world situations it seems likely that mistakes are always possible. As long as there is some chance of both types of errors occurring, then every possible sequence of behavior will be generated with some probability. Since simple strategies like *ALLD* and *CTFT* can each be strong perfect equilibrium against themselves, this condition does not seem hard to fulfill either.

It is important to note that strategies fulfilling these conditions are guaranteed to be stable only if invading strategies are maintained at arbitrarily low frequencies. In fact, mutation will maintain invading strategies at some low, but finite frequency. At such a frequency the effect of interactions with rare invading types on the expected fitness of individuals is bound to be larger than the effect due to interactions with the common type that result in rarely reached nodes in the game tree. It seems possible that there could be an invading type that would be inferior to the common type at some rarely reached nodes of the tree, but still have higher expected fitness than the common type because the rare type did better against other invading types. However, the selection resulting from such fitness differences would be exceedingly weak.

I am very grateful to Eric Rasmusen for his clear and patient explanations of game theory as practiced by economists. I also thank Eric, Mark Ramsayer, and John Wiley for useful discussions of the ideas contained in this paper.

REFERENCES

- AXELROD, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
 AXELROD, R. & HAMILTON, W. D. (1981). *Science*, NY. **211**, 1390.
 BOYD, R. & LORBERBAUM, J. (1987). *Nature*, Lond. **327**, 58.
 BROWN, J. S., SANDERSON, M. J. & MICHOD, R. E. (1982). *J. theor. Biol.* **99**, 319.
 FARREL, J. & WARE, R. Evolutionary stability in the repeated prisoner's dilemma *Theor. pop. Biol.* (in press).
 MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games*. London: Cambridge University Press.
 PECK, J. & FELDMAN, M. W. (1985). *Am. Nat.* **127**, 209.
 RASMUSEN, E. *Games and Information*. Oxford: Basil Blackwell (in press).
 SUGDEN, R. (1986). *The Economics of Rights, Co-operation and Welfare*. Oxford: Basil Blackwell.

APPENDIX A

There are two players, the *CTFT* player, and the "other" player whose strategy is to be determined. The behavior of a *CTFT* player depends only on the standing of both players. The game is stationary because w , the probability that the pair will continue to interact, and e , the probability of a mistake, are both constant. Thus there are only four types of sub-games: those beginning at a node at which both players are in good standing, the *CTFT* player is in good standing and the other player is not, the *CTFT* is not in good standing and the other player is in good standing, and neither is in good standing. I will label these types of nodes *GG*, *NG*, *GN*, and *NN*, respectively. To show that *CTFT* is a strong perfect equilibrium against itself we need to show that the optimal behavior for the other player at each of these nodes is the same as the behavior specified by *CTFT*. I will label the expected payoffs associated with the optimal decision at each type of node, V_{GG} , V_{GN} , V_{NG} , and V_{NN} .

First, consider nodes at which the *CTFT* player is not in good standing, but the other player is in good standing. From the principle of optimality, the optimal expected payoff for the other player at this node is given by:

$$V_{GN} = \max [(1-e)^2(R + wV_{GG}) + e(1-e)(S + wV_{GN} + T + wV_{GG}) + e^2(P + wV_{GN}), \\ (1-e)^2(T + wV_{GG}) + e(1-e)(P + wV_{GN} + R + wV_{GG}) + e^2(S + wV_{GN})] \quad (A1)$$

where "max" denotes the greater of the two terms in brackets. Defection is the optimal action if the first term on the right-hand side of eqn (A1) is less than the second term, or if

$$(1-e)^2(R - T) + e(1-e)(S + T - P - R) + e^2(P - S) > 0 \quad (A2)$$

when T , R , P , and S satisfy the prisoner's dilemma conditions, this condition is satisfied as long as $e < 1/2$. When a *CTFT* individual is not in good standing, it will allow the other individual to defect without reprisal. Thus the optimal behavior for the other individual is to defect. As long as $e < 1/2$, choosing defection will maximize the chance of actually defecting. Notice that this is the same action specified by *CTFT*.

Next, consider nodes at which the *CTFT* player is in good standing, but the other player is not. The optimal expected pay-off for the other player at this node is given by:

$$V_{NG} = \max [(1-e)^2(S + wV_{GG}) + e(1-e)(R + wV_{GG} + P + wV_{NG}) + e^2(T + wV_{NG}), \\ (1-e)^2(P + wV_{NG}) + e(1-e)(T + wV_{NG} + S + wV_{GG}) + e^2(R + wV_{GG})]. \quad (A3)$$

Co-operation is the optimal action if the first term on the right hand side of eqn (A3) is greater than the second term. Assuming that $e < 1/2$, this requires:

$$(1-e)(S - P) + e(R - T) + w(V_{GG} - V_{NG}) > 0. \quad (A4)$$

Finally, consider nodes at which both players are in good standing, or both players are not in good standing. Since the *CTFT* player behaves exactly the same way in both cases, the optimal behavior for the other player and the expected payoff associated with the optimal action are the same at both of these kinds of nodes. The optimal expected payoff for the other player, $V_{GG} = V_{NN}$, is

$$V_{GG} = \max [(1-e)^2(R + wV_{GG}) + e(1-e)(S + wV_{GN} + T + wV_{NG}) + e^2(P + wV_{NN}), \\ (1-e)^2(T + wV_{NG}) + e(1-e)(P + wV_{NN} + R + wV_{GG}) + e^2(S + wV_{GN})]. \quad (A5)$$

Co-operation is the optimal action if the first term on the right hand side of eqn (A5) is greater than the second term. Once again assuming that $e < 1/2$, this will be true if

$$(1-e)[R - T + w(V_{GG} - V_{NG})] + e[S - P + w(V_{GN} - V_{GG})] > 0. \quad (A6)$$

If eqns (A4) and (A6) are satisfied, and if $e < 1/2$, then the optimal response to *CTFT* at every node is the same behavior specified by *CTFT*, and, therefore, *CTFT* is a strong perfect equilibrium against itself. To find the range of parameters which allow these inequalities to be satisfied, we assume that they are satisfied, and then solve eqns (A1), (A3) and (A5) to find expressions for V_{GG} , V_{NG} , and V_{GN} . We then substitute these expressions into eqns (A4) and (A6), to find the range of values of e , w , T , R , P , and S that allow these inequalities to be satisfied. Assuming that the best response to *CTFT* at each type of node is *CTFT* yields the following expressions for the optimal expected payoff at each node.

$$V_{GG} = \frac{we(1-e)(T+S) - 2we^2(1-e)^2(P+R-T-S) + (1-ew)[R(1-e)^2 + e(1-e)(T+S) + e^2P]}{(1-w)[1+ew(1-2e)]} \quad (A7)$$

$$V_{NG} = \frac{(1-e)^2S + e(1-e)[P+R] + e^2T + (1-e)wV_{GG}}{1-ew} \quad (A8)$$

and

$$V_{GN} = \frac{(1-e)^2T + e(1-e)[P+R] + e^2S + (1-e)wV_{GG}}{1-ew}. \quad (A9)$$

Now assume that e is small enough that terms of order e^2 can be ignored. With this assumption eqn (A4) is satisfied if

$$S - P + w(R - S) + e[P - S + R - T + w(T + 3S - 3R - P) + w^2(T - R)] > 0.$$

Thus if eqn (4) is satisfied, then eqn (A4) is satisfied for small enough e . Similarly, eqn (A6) is satisfied if

$$R - T + w(R - S) + e[P - S + R - T + w(2T + 2S - 3R - P) + w^2(T - R)] > 0.$$

Thus, if eqn (4) is satisfied, then eqns (A4) and (A6) are satisfied for sufficiently small e . For $e = 1/2$, $CTFT$ cannot be the uniquely best response to itself since every strategy must have the same expected pay-off against every other strategy. Thus by continuity, there must be some value of $e < 1/2$ for which eqns (A4) and (A6) are no longer satisfied. This threshold value can be found by substituting eqn (A7) through eqn (A9) into eqns (A4) and (A6) and solving for the value of e that makes these equalities. This was done numerically to obtain the results plotted in Fig. 1.