

Article

Evolutionary Exploration of the Finitely Repeated Prisoners' Dilemma—The Effect of Out-of-Equilibrium Play

Kristian Lindgren and Vilhelm Verendel

Complex Systems Group, Department of Energy and Environment, Chalmers University of Technology, SE-412 96 Göteborg, Sweden; E-Mails: kristian.lindgren@chalmers.se (K.L.); vive@chalmers.se (V.V.)

Received: 5 September 2012; in revised form: 15 November 2012 / Accepted: 21 December 2012 / Published: 4 January 2013

Abstract: The finitely repeated Prisoners' Dilemma is a good illustration of the discrepancy between the strategic behaviour suggested by a game-theoretic analysis and the behaviour often observed among human players, where cooperation is maintained through most of the game. A game-theoretic reasoning based on backward induction eliminates strategies step by step until defection from the first round is the only remaining choice, reflecting the Nash equilibrium of the game. We investigate the Nash equilibrium solution for two different sets of strategies in an evolutionary context, using replicator-mutation dynamics. The first set consists of conditional cooperators, up to a certain round, while the second set in addition to these contains two strategy types that react differently on the first round action: The "Convincer" strategies insist with two rounds of initial cooperation, trying to establish more cooperative play in the game, while the "Follower" strategies, although being first round defectors, have the capability to respond to an invite in the first round. For both of these strategy sets, iterated elimination of strategies shows that the only Nash equilibria are given by defection from the first round. We show that the evolutionary dynamics of the first set is always characterised by a stable fixed point, corresponding to the Nash equilibrium, if the mutation rate is sufficiently small (but still positive). The second strategy set is numerically investigated, and we find that there are regions of parameter space where fixed points become unstable and the dynamics exhibits cycles of different strategy compositions. The results indicate that, even in the limit of very small mutation rate, the replicator-mutation dynamics does not necessarily bring the system with Convincers and Followers to the fixed point corresponding to the Nash equilibrium of the game. We also perform a detailed analysis of how the evolutionary behaviour depends on payoffs, game length, and mutation rate.

Keywords: backward induction; rationality; prisoners' dilemma; evolutionary dynamics

1. Introduction

During the past two decades there has been a huge expansion in the development and use of agent-based models for a variety of societal systems and economic phenomena, ranging from markets of various types and societal activities such as energy systems and land use, see e.g., [1–5] for a few illustrative examples. A key issue in the construction of such models is the design of the agents. To what extent are the agents rational? And what does it mean that an agent is rational? This has divided scholars into different camps. Herbert Simon [6] introduced the concept of “bounded rationality”, which can be implemented in a variety of ways in agent-based models, assuming that there is some limitation on the reasoning capability of the agent. Game theory provides useful methods for the analysis of interaction between agents and their behaviour. But it is also well known, from experiments of human behaviour in game-theoretic situations [7], that human subjects do not always follow the behaviour predicted by game theory – and for good reasons. People can in many cases establish cooperation for mutual benefit where game theory would predict the opposite. This discrepancy between “rational” game-theoretic agents and human ones is often attributed to either limited rationality of human reasoning or to social preferences. The latter can in some cases be referred to as rule-based rationality [8] under which rules-of-thumb may have developed over time in cultural evolution under positive selective feedback from the benefits of cooperation.

In the modeling and construction of agents it is therefore of high importance that the assumptions made on rationality and the reasoning process are made explicit. Binmore discusses this in his classic papers “Modeling rational players” [9,10]. He makes the distinction between *eductive* and *evolutive* processes leading to an equilibrium in a game. The former refers to a process internal to the agent representing a reasoning process, while the latter may work with much simpler characterisation of the agents where evolutionary processes lead to an equilibrium by mutation and selection on the population level. In evolutionary game theory and agent-based modeling it is common to use a combination of these, but one seldom designs agents who carefully reason about possible actions and their consequences. And the fundamental question still remains: What does it mean for an agent to be rational?

One of the major achievements in game theory is the establishment of the Nash equilibrium concept and the existence proof that any finite game has at least one such equilibrium [11]. The Nash equilibrium is a situation where no player can gain by unilateral change of strategy, and in that sense this can be seen as a rational equilibrium in providing the player a best response to other players’ actions. The Nash equilibrium is thus often regarded as a result of rational reasoning, reflecting the behaviour of rational players. Importantly, for our discussion, this view of the Nash equilibrium has also been carried over from single-shot to repeated games.

In several finitely repeated games, in which the number of rounds is known, the solution of how players choose actions can be guided by the backward induction procedure. This is often exemplified by the Prisoners’ Dilemma, for which the single round game has a unique Nash equilibrium with both players defecting, while the indefinitely repeated game has an uncountable infinity of equilibria allowing for cooperation. However, when the exact number of rounds, n , is known, a player can start with considering the last round, in which the score is maximized by defecting. So, with both players being rational in the sense that they want to maximize their own score, the outcome of the last round is

clear—mutual defection. But then the next-to-last round turns into the last unresolved round, and the same reasoning applies again resulting in mutual defection also for round $n - 1$. The assumption needed is that each player knows that the other one is rational. The procedure then repeats all the way to the first round, showing that the Nash equilibrium is mutual defection from the start of the game. Since the result of the backward induction procedure does not seem to lead to an intuitively rational result it is often referred to as the "backward induction paradox" [12]. Note that backward induction also applies to certain one-shot games, and the discrepancy between theory and observation has been discussed also for several such examples, e.g., the "Beauty contest" [13] and the "Traveler's dilemma" [14].

There are at least two important objections against the generality of the reasoning based on backward induction. The first objection is empirical, since studies on how human players behave in the game show a substantial level of cooperation, but with a transition to lower levels of cooperation towards the end. Explanations are several, and this implies that several mechanisms are in play. For example, it has been observed in the laboratory that subjects cooperate initially but attempt to cheat each other by deviation in the end [7,15].

The second objection is conceptual and strongly connected to the notion of rationality and what can be considered as a rational way of reasoning. The only equilibrium that can exist in a given finite repetition is the Nash equilibrium, but whether that is to be considered as rational is the question. A critical point concerns what conclusion a player should draw if the opponent deviates from what backward induction implies and instead cooperates *in the first round*. It is then clear that the opponent is not playing according to the Nash equilibrium, and there is a chance to get a higher score for some time if cooperation can be established.

In this situation, the choice between (i) following backward induction and defecting from start and (ii) deviating from backward induction by starting with cooperation becomes a strategic decision. One can imagine "rational" players in both categories. In the first category, there are then two options, either one just plays defect throughout the game whatever the opponent does, as backward induction suggests, or one switches to cooperation if the opponent cooperates. In the second category, both players are again faced with the question of, provided the opponent is cooperating, when to switch to defection. Obviously, there cannot exist a fixed procedure for deciding on when it is optimal to switch from cooperation to defection, since such a strategy would be dominated by the one that switches one round before. However, the interaction and survival of different ways to handle first round cooperation can be studied using evolutionary methods.

The purpose of this paper is to investigate in an evolutionary context the performance of strategies representing the strategic choices discussed above in the finitely repeated Prisoners' Dilemma. In Binmore's terms, we focus on an evolutive process, in which each agent has a certain, relatively simple strategy for the game, and the mix of strategies and their evolution is investigated on the population level. Importantly, the chosen strategies can all be seen as components in the reasoning processes discussed above: both (i) the steps involved in the backward induction process, and (ii) the steps initiating and responding to cooperation in the first round which then reflects the possibility for strategies to deviate from equilibrium play. It is well known that evolutionary drift or mutations, at least if sufficiently strong, can drive the population away from a fixed point corresponding to the Nash equilibrium. Under what circumstances does the evolutionary dynamics lead to the same result as the backward induction process

with a Nash equilibrium as its fixed point, and when can deviation from Nash equilibrium play alter that process? The answer, which is elaborated in this paper, depends on choices of a number of critical model characteristics and parameters: selected strategy space, mutation rate, payoff matrix, and the length of the game.

We prove that, for a simple set of strategies, *i.e.*, conditional cooperators, the replicator-mutation dynamics is always characterised by a stable fixed point corresponding to the Nash equilibrium in the limit of zero (but positive) mutation rate. Numerically we show that such a result does not hold in general, even if the only Nash equilibrium is characterised by defection from the start of the game. The strategy set that allows for different reactions on the first round may lead to a path of actions different from what is considered in the backward induction process. On the population level, this turns out to destabilise the dynamics, and for a large part of parameter space, the evolution does not bring the system to a stable fixed point, even in the limit of zero (but positive) mutation rate. The dynamics is instead characterised by oscillatory behaviour.

Most of the work related to evolutionary dynamics, backward induction, and the finitely repeated Prisoners' Dilemma concerns the replicator dynamics [16] with no mutation, as this class has been examined analytically more thoroughly than the replicator-mutation dynamics [17]. Nachbar [18] studies convergence in the dynamics and shows that for 2-stage games all cooperation goes extinct when starting from a mixed population. This result is extended by Cressman [19,20] to the finitely repeated Prisoners' Dilemma of arbitrary length.

Several authors have investigated various types of evolutionary dynamics under the effect of perturbations or mutations ([21–25]). Here the focus has primarily been on other games than the finitely repeated Prisoners' Dilemma. Gintis et al. [26] consider the replicator dynamics subject to recurrent mutation when mutation rate goes to zero. For finite noncooperative games, they show that the dynamics need not converge to the subgame perfect equilibrium, and limiting equilibria can be far away from the this equilibrium. They also show that in the n -player Centipede game, there exists a unique limiting equilibrium as mutation rate goes to zero, which is far from the subgame perfect equilibrium but equal in payoffs.

Ponti [27] studies replicator-mutation dynamics in the Centipede game [28]. Using simulations, he finds recurrent phases of cooperation in the evolutionary dynamics, and for particular payoffs of the game, this is shown to depend both on mutation rate as well as the length of the game. It is left as an open question whether such behaviour would disappear in the long run, or persist for negligible mutation rate. This result is most relevant to the present paper in the context of the finitely repeated Prisoners' Dilemma.

Our work is also related to the literature on the "backward induction paradox" [12], which has focused on deviation from equilibrium play in the first round of extensive games. For the backward induction reasoning to be rational, it has been assumed that each player has full knowledge about the rationality of the other player, and that both players know this, et cetera, known as "common knowledge of rationality". It has been proved by Aumann [29] that such common knowledge implies backward induction in games with a unique subgame perfect equilibrium. However, backward induction provides no firm basis to act when a player deviates from equilibrium play [30–32]. Our study can be viewed as an evolutionary study of a population in a setting where reacting to out-of-equilibrium play in the first round is possible.

2. Evolutionary Dynamics

The evolution of strategies in the finitely repeated Prisoners' Dilemma is studied using replicator dynamics with a uniform mutation rate. This is a model of an infinite population where all interact with all, and in which each strategy i occupies a certain fraction x_i of the population. The selection process gradually changes the population structure, based on a comparison of the average score s_i of strategy i with the average score s in the population,

$$s_i = \sum_{j=1}^n u(i, j)x_j \quad (1)$$

$$s = \sum_{i=1}^n s_i x_i \quad (2)$$

where $u(i, j)$ is the score for strategy i meeting strategy j in the N -round Prisoners' Dilemma. Assuming a uniform mutation rate of ϵ , the replicator-mutation dynamics can be written,

$$\frac{dx_i}{dt} = x_i (s_i - s - \epsilon) + \epsilon/n \quad (3)$$

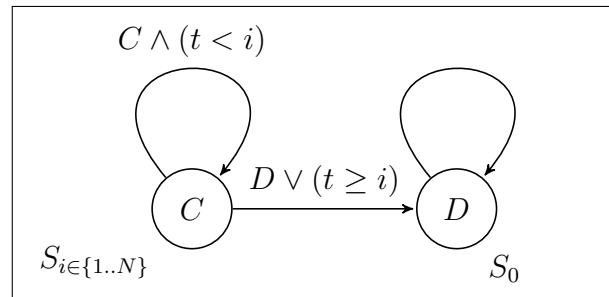
where n is the number of different strategies. The dynamics conserve the normalisation of x .

The score $u(i, j)$ between two strategies, i and j , is calculated from N rounds of the Prisoners' Dilemma with a payoff matrix for the row player given by

	C	D
C	1	0
D	T	P

This means that the reward R for mutual cooperation (C, C) has been set to 1, and the "suckers payoff" S for cooperation against defection (C, D) is 0. The temptation to defect against a cooperator (D, C) is associated with a score T , and for mutual defection (D, D) the score is P . With the usual constraints of $T > R > P > S$ and $2R > T + S$ it remains to study the parameter region given by $T \in [1, 2]$ and $P \in [0, 1]$. There are in principle three independent parameters in the game, but in combination with the replicator-mutation dynamics, the third one is incorporated in the mutation rate ϵ (which can be derived by subtraction and division by S and $R - S$, respectively, in the replicator-mutation dynamics equation, Equation (3)).

Figure 1. Finite state machine illustrating the first strategy set, Γ_1 , of the S_i strategies. The action to perform is in the node, and transitions are taken on the basis of the opponent's action in the previous round (or based on the previous round number t). All strategies start in the left node (C), except S_0 that starts with defection in the right node (D). S_i cooperates conditionally for i rounds after which it starts to defect.



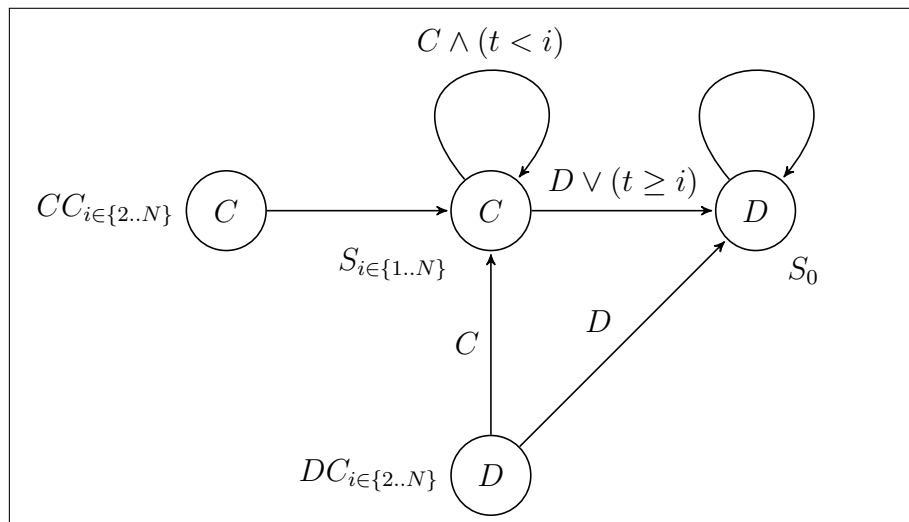
2.1. Selecting the Strategy Set

We investigate the evolutionary behaviour considering two sets of strategies. The first one is a strategy set Γ_1 that represents various levels of depth in applying the backward induction procedure to conditional cooperation. A strategy in this set is denoted S_k (with $k \in \{0, 1, \dots, N\}$), which means that the strategy is playing conditional cooperation up to a round k and then defects throughout the game, see Figure 1. Conditional cooperation means that if the opponent defects, one switches to unconditional defection for the rest of the game. For example, S_N is prepared to cooperate through all rounds (or as long the opponent does), while S_0 defects from the first round. It is then clear that strategy S_k dominates S_{k+1} (for $k \in \{0, 1, \dots, N-1\}$), and backward induction leads us to the single Nash Equilibrium in which both players choose strategy S_0 .

Note that the entire strategy set Ω for the N -round Prisoners' Dilemma is very large, as a strategy requires a specification how to react on each possible history (involving the opponent's actions) for every round of the game. This results in 2^{2^N-1} possible strategies, e.g. for $N = 10$ rounds this is in the order of 10^{308} . The selection of strategies to consider is critical. One can certainly introduce strategies so that other Nash equilibria are introduced along with those characterized by always defection. For example, selecting only three strategies, e.g., $\{S_0, S_5, S_{10}\}$ will lead to a game with three equilibria, one for each strategy. But this is an artefact of the specific selection made. Here, we do not want to create new Nash equilibria but we want to investigate how and if the evolutionary dynamics brings the population to a fixed point dominated by defect actions corresponding to the original Nash equilibrium. One way to achieve this is to make sure that iterated elimination of weakly dominated strategies can be applied to the constructed strategy set, in a way so that only strategies that defect throughout the game remain, keeping the non-cooperative characteristic of the Nash equilibrium.

The second set of strategies Γ_2 (in Figure 2) that we consider is given by extending Γ_1 to include also strategies corresponding to steps of reasoning in which one (i) tries to establish cooperation even if the opponent defects in the first round and (ii) responds to such attempts by switching to cooperation for a certain number of rounds. We refer to such strategies as "Convincers" and "Followers", respectively.

Figure 2. Finite state machine illustrating the extended strategy set Γ_2 consisting of the strategies S_i , CC_i , and DC_i . S_i are the conditional cooperators as described in Figure 1. The Convincers are denoted CC_i and the Followers DC_i . Strategies start in the state at which the name is placed. The strategies CC_i start with at least two rounds of cooperation, which may trigger DC_i to switch from defection to cooperation. After that the latter two strategies act as the conditional cooperators S_i .



A Convincer strategy CC_k starts with cooperation twice in a row, regardless of the opponent's first round action, and in this way it can be seen as an attempt to establish cooperation even with first round defectors. From round 3, the strategy plays as S_k , *i.e.*, conditional cooperation up to a certain round $k \in \{2, 3, \dots, N\}$.

A Follower strategy DC_k starts with defection, but can be triggered to cooperation by a Convincer (or S_k where $k > 0$). So the Follower switches to cooperation in the second round, after which it, like the Convincer, plays as S_k , *i.e.*, conditional cooperation up to round $k \in \{2, 3, \dots, N\}$. (A Follower strategy is also triggered to cooperation by an S_k strategy, but S_k does not forgive the first round defect action and cooperation cannot be established.)

For the extended strategy set, Γ_2 , it is straightforward to see that iterated elimination of weakly dominated strategies, starting with those cooperating throughout the game, leads to a Nash equilibrium with only defectors.

For the first strategy set, Γ_1 , the Nash equilibrium of (S_0, S_0) is strict since any player deviating would score less. For the second strategy set, Γ_2 , this Nash equilibrium is no longer strict as one of the players could switch to a Follower strategy, still defecting and scoring the same. For the first strategy set, the NE is unique, but for the second one that is not necessarily the case. Since backward induction still applies in the second set, we know that any NE is characterized by defection only, which can be represented by a pair (S_0, DC_k) . This is a NE only if the S_0 player cannot gain by switching to CC_{k-1} (or to CC_2 if $k = 2$). This translates into $kP > S + (k-2)R + T$ (for $k > 2$) or $T < kP - (k-2)$ for our parameter space, while for $k = 2$, we have $2P > S + R$ or $P > 1/2$. Furthermore, if (S_0, DC_k) is a NE, then also (DC_j, DC_k) , with $j \leq k$, is a NE.

This means that in a part of the payoff parameter region, for $P < 1/2$, there is only one NE, while for larger P values there are several NEs, with increasing number the closer to 1 the parameter P is. Note, though, that any additional NE here is characterized by defection from the first round, corresponding to the unique subgame perfect Nash equilibrium (S_0, S_0) . This illustrates the fact that in the NE one can switch from a pure defector to a Follower without reducing the payoff. If this happens under genetic drift in evolutionary dynamics, the situation may change so that Convincers may benefit and cooperation can emerge.

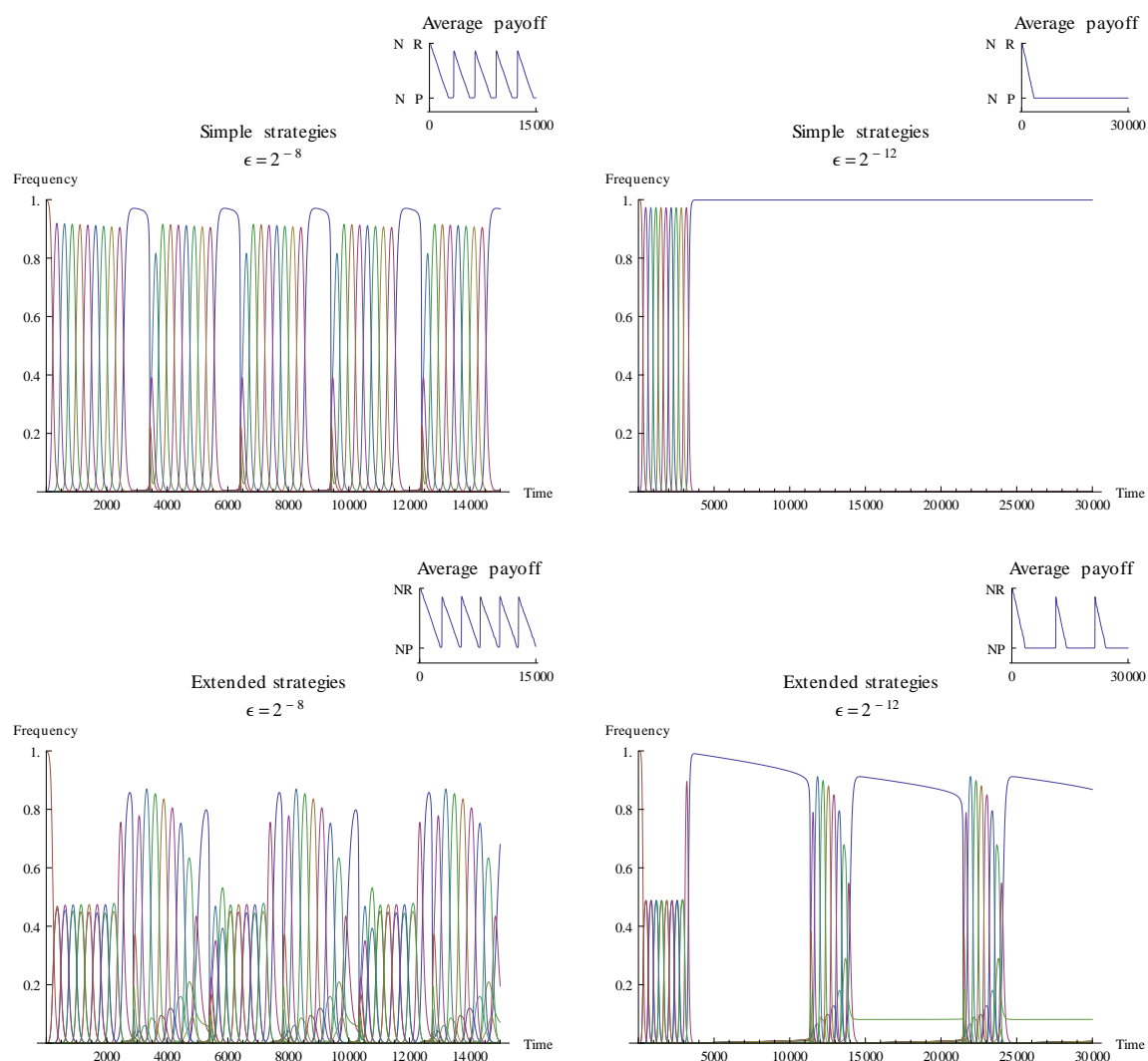
3. Dynamic Behaviour and Stability Analysis

The dynamic behaviour and the stability properties of the fixed points are investigated both analytically and numerically, for the two strategy sets presented in Section 2.1. In each case, we investigate the dependence of the payoff parameters, T and P , as well as the mutation rate and game length, ϵ and N , respectively. We are mainly interested in which influence the presence of Convincers and Followers has on the stability of fixed points and its impact on the dynamics. This will be illustrated in three different ways as follows. First, we present and examine a few specific examples of the evolutionary dynamics and discuss the qualitative difference between the strategy sets. For the simple strategy set, we show analytically that the fixed point associated with the Nash equilibrium remains stable for a sufficiently small mutation rate. A numerical investigation is then performed for the extended strategy set, characterising the fixed point stability. Finally, for different lengths of the game, we study realisations of the evolutionary dynamics from an initial condition of full cooperation ($x_N = 1$ and $x_i = 0$ for $i \neq N$, with x_k denoting the fraction of strategy S_k in the population) to examine to which degree cooperation survives and whether the dynamics is attracted to a fixed point or characterized by oscillations. By studying variation of the dynamics over the different regions in the parameter space and changing the payoff parameters T and P (by adhering to the inequalities between T, P, R, S as described in Section 2), it is studied how the population evolves in various games over time.

3.1. Dynamics with the Simple and the Extended Strategy Set

Realisations of the evolutionary dynamics, Equation (3), for both the simple and the extended strategy sets are shown in Figure 3. For particular payoff values and a 10-round game, the mutation rate ϵ is varied to illustrate how it affects the dynamics.

Figure 3. Illustration of the dynamics for a particular game ($P=0.2$, $T=1.33$) for 10-round Repeated Prisoners' Dilemma ($N = 10$). Each main plot displays how the frequency of different strategies changes in the population as a function of time. Smaller subplots show how the population average payoff changes with time, with NR and NP denoting full cooperation and defection, respectively. Top: simple strategy set (S_0, \dots, S_{10}). Below: the extended strategy set, which includes also Convincers (CC_2, \dots, CC_{10}) and Followers (DC_2, \dots, DC_{10}). When lowering ϵ , the extended strategy set exhibits meta-stability with recurring cooperation, while for the simple strategy set cooperation disappears.



First, we consider the case with the simple strategies (S_0, \dots, S_{10}) in Figure 3. From an initial state of full cooperation, with a population consisting only of fully cooperative S_{10} players, the dynamics will, for both levels of mutation rate, lead to a gradual unraveling of cooperation to a point where S_0 , full defection, dominates the population. The first step of this unraveling occurs because S_9 defecting in the final round will have higher payoff than S_{10} . At this stage S_0 is much worse off, but the population goes through a series of transitions which reminds of a backward induction process. This can also be seen in terms of average payoff, as illustrated in Figure 3. When the population is in this non-cooperative mode a positive mutation-rate ϵ may offset the situation. For the higher mutation rate, cooperative strategies re-emerge after a period of influx from other strategies. The mutations gradually introduce cooperative behaviour to a critical point where some degree of cooperation has a selective advantage over full defection, and we see a shift in the level of cooperation. After a while, cooperative behaviour is again overtaken by full defection and a cyclic behaviour becomes apparent. Comparing with the next realisation of the simple strategies we see that this can happen only when the mutation rate is high enough. As mutation rate becomes smaller, here illustrated with $\epsilon = 2^{-12}$, there is no re-appearance of cooperation. When the mutation rate gets too low, strategies other than defection are kept on a level that is too low to promote further cooperation. This demonstrates that the mutation rate can affect whether cooperation re-appears or not.

Second, we consider the case with extended strategies (the $3N - 1$ strategies $S_0, \dots, S_{10}, CC_2, \dots, CC_{10}, DC_2, \dots, DC_{10}$) shown at the bottom of Figure 3. We observe that adding Convincer and Follower strategies changes the dynamics, but with a similar unraveling of cooperation. For both mutation rates, the trajectories have periods with defection dominating in the population, indicated by average payoff, but the population does not seem to stabilize. Lowering the mutation rate increases the time between the outbreaks of cooperation. Contrary to the simple strategy set, the system does not seem to settle close to full defection. The explanation is due to the Follower strategies being able to gradually enter the population by getting the same payoff as the strategy of full defection. At a critical point, there are sufficiently many Followers, which make mutations to Convincers successful and cooperation re-enters the population for a period.

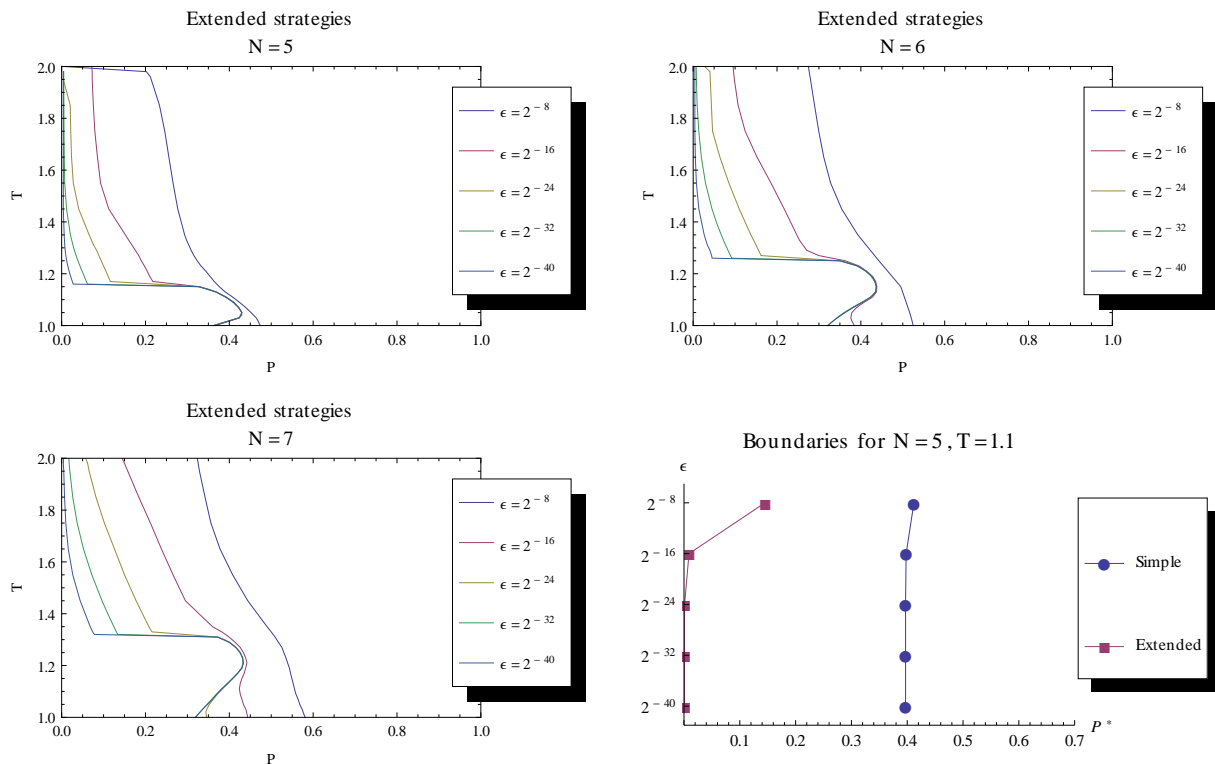
In the next section, we will investigate the dynamics and the stability characteristics of both the simple and the extended strategy sets in detail, varying the payoff parameters over the full ranges, and investigating the behaviour in the limit of diminishing mutation rate.

3.2. Existence of Stable Fixed Points

We now turn to examine the existence of stable fixed points in the dynamics for low mutation rates. For the simple strategy set the following proposition holds.

Proposition 1: For the simple set of strategies, Γ_1 , the fixed point associated with the Nash equilibrium, dominated by strategy S_0 , is stable under the replicator-mutation dynamics, if the mutation rate is sufficiently small (but positive).

Figure 4. Numerical analysis showing which games have stable fixed points. Stable fixed points have been found to the right of a given line, at which they disappear if P is decreased further. Lowering the mutation rate ϵ turns more evolutionary dynamics into having stable fixed points. Runs for $N = 5, 6, 7$ indicate a convergence, as the mutation rate is decreased, towards a fraction of games being without stable fixed points, as seen in the lower left of the parameter space for the different game lengths. To the bottom right the boundary is shown for a particular $T = 1.1$.



Proof 1: See Appendix A.

Our results for the extended strategy set, Γ_2 , are based on numerical investigations: by using an eigenvalue analysis of the Jacobian of the replicator-mutation dynamics, Equation (3), we determine for which parameters T , P and ϵ the dynamics is characterised by stable fixed points. We are especially interested in the case of lowering ϵ towards 0 to examine whether fixed points become stable when mutation rate is sufficiently small, like in the case of the simple strategy set.

This stability analysis over the parameter space and the results are presented for different lengths of the game in Figure 4. Stable fixed points exist to the *right* of the line corresponding to a specific ϵ , while to the *left* of the line no stable fixed points were found. When decreasing the mutation rate ϵ towards 0, we see that an increasing fraction of parameter space is characterised by a stable fixed point. Unlike the case for the simple strategy set, the numerical investigation shows a convergence of the delimiting line between stable and unstable fixed points, indicating that there is a remaining region in parameter space (with low P and low T) for which fixed points are unstable in the limit of zero mutation rate.

Note from the discussion in Section 1 that while these results show the existence of stable fixed points, it is left to consider whether the population dynamics would actually converge to such states. Next, we turn to consider population dynamics over the parameter space and its outcome.

3.3. Recurring Phases of Cooperation

Motivated by the findings above in Section 3.1, we now turn to study recurrent cooperative behaviour in the evolutionary dynamics. We investigate if and for which games a population, starting from initial cooperation as before, settles into a mode of oscillations that at some point in the cycle brings the system back to a higher level of cooperation¹.

The interesting case is when mutation rates are small: higher mutation rates introduce a background of all different strategies, which can be seen as artificially keeping up cooperative behaviour in the population. To avoid this effect, we investigate the dynamics with $0 < \epsilon \ll 1$, and, in particular, numerical analyses were made with a series of decreasing mutation rates. The population is initialized as before, described in Section 3.1.

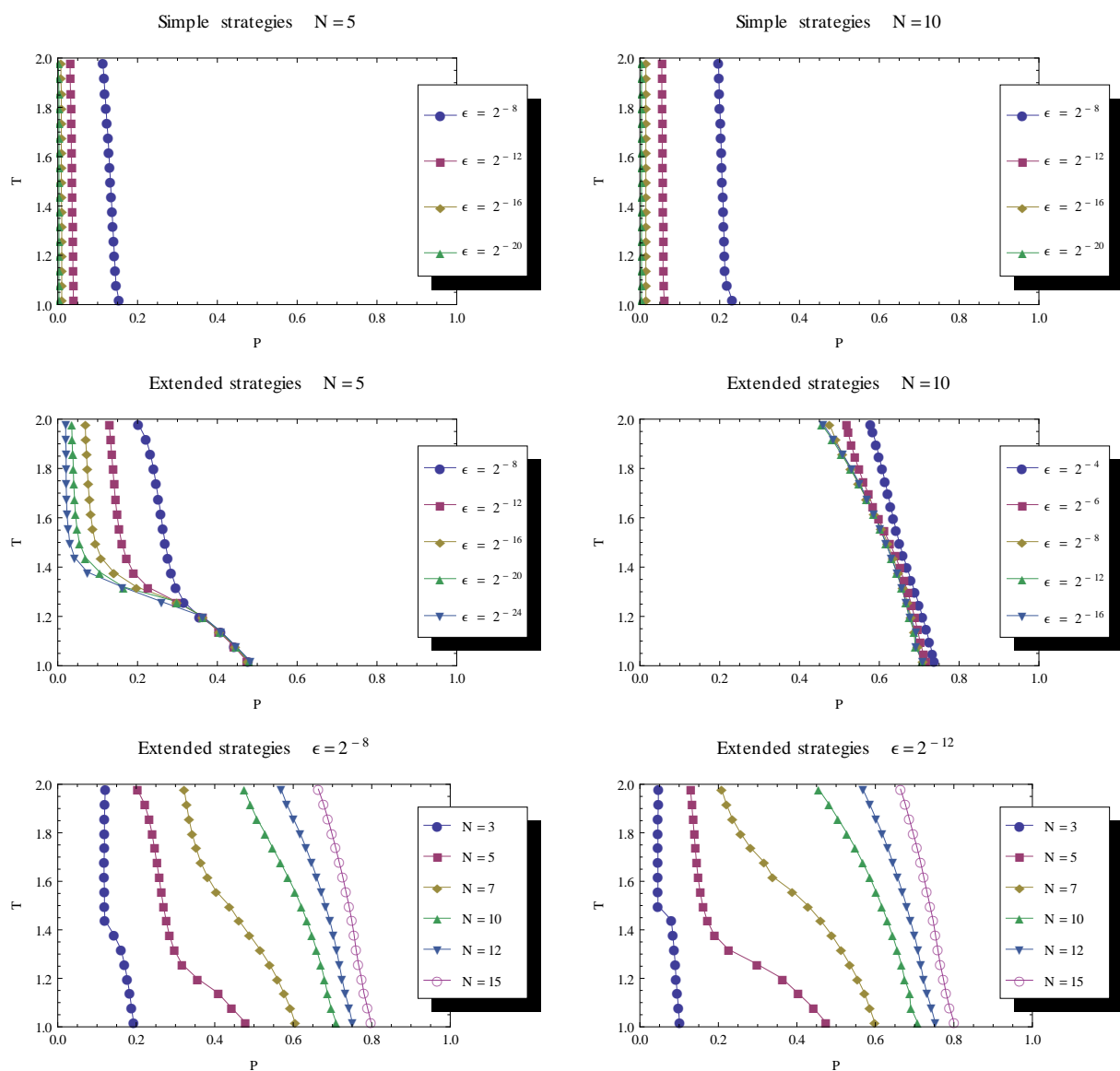
First, we characterise the simple and the extended strategy set for different game lengths, varying mutation rates, and varying the parameters of T and P . Figure 5 displays the results for games of 5 and 10 rounds in the top and the middle row. For a given T a line denotes the border for which games to the *left* have recurring phases of cooperation, while games to the *right* are characterised by a fixed point dominated by defection. For the simple strategy set, we observe that by decreasing mutation rate, the fraction of games where cooperation recurs becomes smaller and disappear. This suggests that as we make the mutation rate sufficiently small, cooperation will die out in the case of simple strategies, as is expected from Proposition 1 stating that the fixed point dominated by the S_0 strategy becomes stable. On the other hand, for the extended strategy set, a considerable fraction of games seem to offer recurring phases of cooperation despite lowering the mutation. For the 10-round game, the line describing the critical parameters is seen to converge as the mutation rate decreases, *i.e.*, there is a large part of the parameter region for which the dynamics is not attracted to a fixed point. For the 5-round game, this occurs at least in part of the parameter space. The bottom graphs in Figure 5 show that the longer the game, the larger is the parameter region for T and P in which the fixed point is avoided and recurring periods of cooperation are sustained. It should be noted, though, that as the mutation rate decreases, the part of the cycle in which there is a significant level of cooperation decreases towards zero. This is due to the slow genetic drift that brings DC_k strategies back into the population, which eventually make it possible for the CC_k strategies to re-establish a significant level of cooperation.

3.4. Co-existence between Fixed Point Existence and Recurring Cooperation

The discussion in Section 3.2 left the question of whether stable fixed points are reached by population dynamics, and we found that it is not necessarily so in Section 3.3. This was illustrated for the extended strategy set for low ϵ . Combining the different findings by considering their boundaries in the parameter space, this suggests an additional property of the evolutionary dynamics. By considering the joint results of our findings (in Figures 4 and 5), one can note that there is a part of parameter space for which there is co-existence between a stable fixed point of defect strategies and a stable cycle with recurring cooperation.

¹An instance of the dynamics was counted as oscillating when the average payoff A repeatedly returns to at least 5% above full defection, *i.e.* $A > 1.05NP$. Frequently, it was the case that the oscillations had phases of cooperation well above the 5% threshold.

Figure 5. Parameter diagram showing which games have recurrent cooperation in the evolutionary dynamics from a starting point of initial cooperation. Top row: simple strategy set. Middle row: extended strategy set with Convincers and Followers. In the graphs, recurrent cooperation exists to the left of the line and a fixed point with defectors characterizes the behaviour to the right. For simple strategies, lowering ϵ steadily reduces the part of parameter space dominated by recurrent cooperation. In the bottom row, the delimiting line is shown for a variety of game lengths and for two levels of mutation (left and right), illustrating the fact that the longer the game the larger the parameter region supporting cooperative phases in the evolutionary dynamics.



4. Discussion and Conclusion

A key point in game theory is that a player's strategic choice must consider the strategic choice the opponent is making. For finitely repeated games, backward induction as a solution concept has become established by assuming player beliefs being based on common knowledge of rationality. However, this assumption says nothing about how players would react to a deviation from full defection—a deviation from the Nash equilibrium—since it a priori rules out actions and reactions that exemplify other ways of reasoning.

Motivated by the general importance of backward induction and what has been called its "paradox", we have introduced an evolutionary analysis of the interaction in a population of strategies that react differently to out-of-equilibrium play in the first round of the game. We have shown how extending a strategy set for this possibility, in the special case of the repeated Prisoners' Dilemma, allows for stable limit cycles in which cooperative players return after a period of defection. The introduction of Convincers and Followers, representing both strategies that try to establish cooperation and strategies that are capable of responding to that, are made in a way to preserve the structure of the selected strategy set so that elimination of weakly dominated strategies leads to full defection.

For the simple strategy set, as the mutation rate becomes sufficiently small, the cyclic behaviour disappears and the system is attracted to a stable fixed point. The stability of this fixed point was shown analytically for a sufficiently small mutation rate ϵ .

For the extended strategy set, for low levels of mutation, the numerical investigation of fixed point stability and oscillatory modes indicates that, for a certain part of payoff parameter space, the evolutionary dynamics does not reach a stable fixed point but stays in an oscillatory mode, unlike the case of simple strategy set. We characterise our results by a detailed quantitative analysis of where this occurs: showing how the length of the repeated game and the mutation rate affects the boundaries of this region.

One of the main results of the study is an affirmative answer to the question whether different responses to out-of-equilibrium play in the first round can make the dynamics avoid fixed points, and the corresponding Nash equilibrium. Additionally, the fixed point analysis showed the co-existence of a stable fixed point and stable oscillations with recurring phases of cooperation. This means that a system with different responses to out-of-equilibrium play may be found far from its possible stable fixed point. Taken together, this illustrates that the Nash equilibrium play can be unstable at the population level when mutations make explorations off the equilibrium path possible.

This paper contributes to the backward induction discussion in game theory, but more broadly to the study of repeated social and economic interaction. Many models, typically much larger and less transparent ones, of social and economic systems involve agents. If solving these systems means finding the Nash equilibria, then one may doubt whether that is a good representation of rational behaviour except under certain conditions as we have discussed in the paper. We have shown that strategies corresponding to the Nash equilibrium cannot be taken for granted when they interact and compete with strategies that act and respond differently to out-of-equilibrium play.

Acknowledgements

Financial support from the Swedish Energy Agency is gratefully acknowledged. We would also like to thank two anonymous reviewers for constructive criticism and for inspiring us to prove Proposition 1. We also thank David Bryngelsson for valuable comments on the introduction.

Appendix A. Stability of fixed point in the simple strategy set for small mutation rates

In order to show that the Nash equilibrium fixed point at zero mutation rate, $\epsilon = 0$, continuously translates into a stable fixed point as the mutation rate becomes positive, $\epsilon > 0$, we investigate the fixed point more thoroughly. First, we note that the stability of the fixed point without mutations ($\epsilon = 0$), characterised by $x_1 = 1$ (strategy S_0 dominating), can be determined by the largest eigenvalue of the Jacobian matrix $(\partial \dot{x}_i / \partial x_j)$ derived from the dynamics, Equation (3), where we use the notation $\dot{x}_i = dx_i/dt$. One finds that the largest eigenvalue is given by $\lambda_{\max} = -P < 0$, which shows that the fixed point is stable. This is also known from the stability analysis of the finitely iterated game by Cressman [19].

We will now proceed with reformulating the fixed point condition, for positive mutation rate $\epsilon > 0$, which is a set of n polynomial equations given by $\dot{x}_i = 0$ (with $i = 1, \dots, n$), into an equation of only one variable x_1 . Based on this we show that the Nash equilibrium fixed point for zero mutation rate $\epsilon = 0$, characterised by $x_1 = 1$, continuously moves into the interval $x_1 \in [0, 1]$, with retained stability.

We let index i denote strategy S_{i-1} in the simple strategy set ($i = 1, \dots, n$), where the number of strategies is $n = N + 1$. Due to the structure of the repeated game for the simple strategy set, determining $u(i, j)$, we can establish pair-wise relations between x_k and x_{k+1} for $k = 1, 2, \dots, n-1$, at any fixed point. We use the slightly more compact notation $s_{i,j} = u(i, j)$, for the score for a strategy i against j .

For x_1 and x_2 (strategies S_0 and S_1), using Equation (3), we have

$$\frac{dx_1}{dt} = x_1 (s_{1,1}x_1 + s_{1,2}(1 - x_1) - \bar{s} - \epsilon) + \frac{\epsilon}{n} = 0 \quad (4)$$

$$\frac{dx_2}{dt} = x_2 (s_{2,1}x_1 + s_{2,2}x_2 + s_{2,3}(1 - x_1 - x_2) - \bar{s} - \epsilon) + \frac{\epsilon}{n} = 0 \quad (5)$$

where we have used the fact that $s_{1,2} = s_{1,j}$ (for $j > 2$), and $s_{2,3} = s_{2,j}$ (for $j > 3$). At a fixed point, for $\epsilon > 0$, all strategies are present, $x_i > 0$, and we can divide these equations by x_1 and x_2 , respectively. Then, taking the difference between the equations gives us an equation for the relation between x_1 and x_2 ,

$$\frac{\epsilon}{n} \left(\frac{1}{x_1} - \frac{1}{x_2} \right) + x_1 + (T - P)x_2 - (1 - P) = 0 \quad (6)$$

where we have used that $s_{2,3} - s_{1,2} = 1 - P$, $s_{1,1} - s_{2,1} - s_{1,2} + s_{2,3} = 1$, and $s_{2,3} - s_{2,2} = T - P$. Solving this quadratic equation gives x_2 as a function of x_1 ,

$$x_2 = f_A(x_1) \quad (7)$$

with the function $f_A(x)$ defined by

$$f_A(x) = \frac{1}{2(T-P)} \left(x - (1-P) + \frac{\epsilon'}{x} \right) \left(-1 + \sqrt{1 + \frac{4(T-P)\epsilon'}{(x - (1-P) + \frac{\epsilon'}{x})^2}} \right) \quad (8)$$

Here we have introduced $\epsilon' = \epsilon/n$. In the same way, for x_{k-1} and x_k (for $k = 3, \dots, N-1$), the fixed point implies that

$$\frac{dx_{k-1}}{dt} = x_{k-1} \left(\sum_{j=1}^{k-2} s_{k-1,j} x_j + s_{k-1,k-1} x_{k-1} + s_{k-1,k} \left(1 - \sum_{j=1}^{k-1} x_j \right) - \bar{s} - \epsilon \right) + \frac{\epsilon}{n} = 0 \quad (9)$$

$$\frac{dx_k}{dt} = x_k \left(\sum_{j=1}^{k-2} s_{k,j} x_j + s_{k,k-1} x_{k-1} + s_{k,k} x_k + s_{k,k+1} \left(1 - \sum_{j=1}^k x_j \right) - \bar{s} - \epsilon \right) + \frac{\epsilon}{n} = 0 \quad (10)$$

where we have used the fact that $s_{k,j} = s_{k,k+1}$ for $j > k$. Again, the difference between these equations results in a relation between x_k and all x_j with $j < k$ (for $k = 3, \dots, N-1$). Since $s_{k-1,j} = s_{k,j}$ for $j < k-1$, the difference can be written

$$\frac{\epsilon}{n} \left(\frac{1}{x_{k-1}} - \frac{1}{x_k} \right) + P x_{k-1} + (T-P) x_k - (1-P) \left(1 - \sum_{j=1}^{k-1} x_j \right) = 0 \quad (11)$$

where the score differences, $s_{k-1,k-1} - s_{k,k-1} = P - S = P$, $s_{k,k+1} - s_{k,k} = T - P$ and, $s_{k-1,k} - s_{k,k+1} = P - R = P - 1$, have been used. This results in an expression for x_k in terms of all x_j (with $j < k$),

$$x_k = f_B \left(x_{k-1}, 1 - \sum_{j=1}^{k-1} x_j \right) \quad (12)$$

with the function $f_B(x, w)$ defined by

$$f_B(x, w) = \frac{1}{2(T-P)} \left(P x - (1-P) w + \frac{\epsilon'}{x} \right) \left(-1 + \sqrt{1 + \frac{4(T-P)\epsilon'}{(P x - (1-P) w + \frac{\epsilon'}{x})^2}} \right) \quad (13)$$

Finally, using the same approach for the last two variables, x_{n-1} and x_n , the equation for the relation between these can be written

$$\frac{\epsilon}{n} \left(\frac{1}{x_{n-1}} - \frac{1}{x_n} \right) + P x_{n-1} + (T-1) x_n = 0 \quad (14)$$

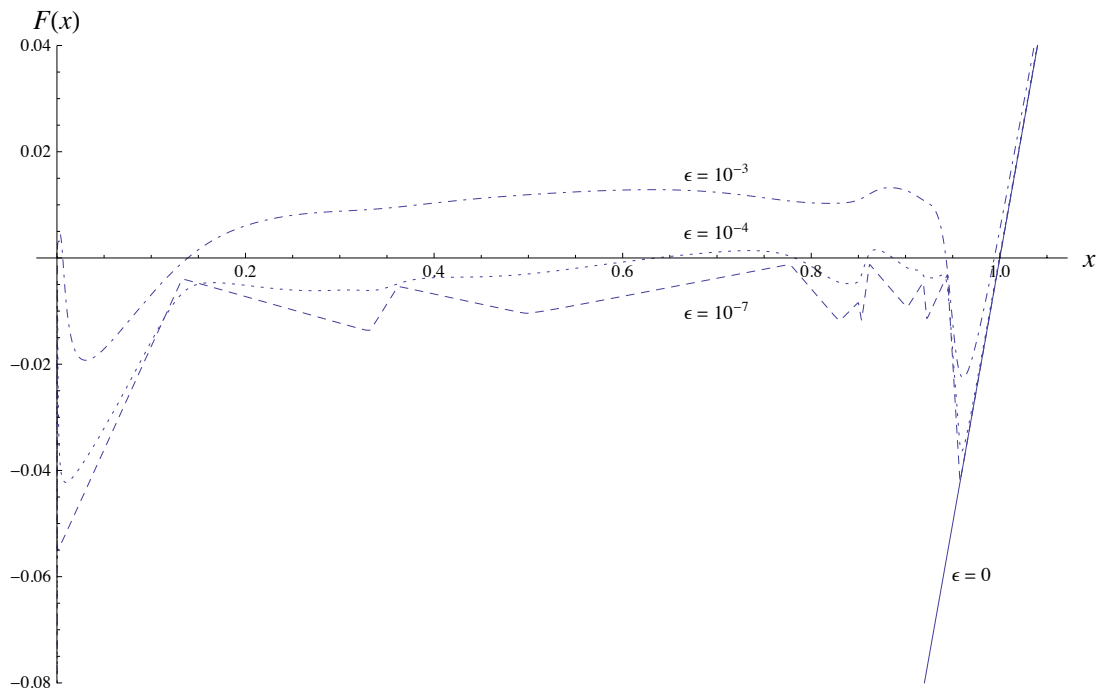
The last variable can then be expressed as

$$x_n = f_C(x_{n-1}) \quad (15)$$

with the function $f_C(x)$ defined by

$$f_C(x) = \frac{1}{2(T-1)} \left(P x + \frac{\epsilon'}{x} \right) \left(-1 + \sqrt{1 + \frac{4(T-1)\epsilon'}{(P x + \frac{\epsilon'}{x})^2}} \right) \quad (16)$$

Figure 6. The function $F(x)$ for the 5-round PD game with $T = 1.05$, $P = 0.15$, and four different mutation rates ϵ .



This means that we can recursively express the fixed point abundancies x_k of strategies $k = 2, \dots, n$ in terms of x_1 , using Eqs. (8), (13), and (16). Together with the normalisation constraint, this results in an equation with only one variable, x_1 , that determines the fixed points. In other words: Summation over all x_k and subtraction of 1 gives us a function of x_1 , $F(x_1)$, with zeroes at the fixed points,

$$F(x_1) = \sum_{k=1}^n x_k - 1 \quad (17)$$

Note that we can extend the function F also to the case $\epsilon = 0$, since all functions f_A , f_B and f_C are identical to zero in that case and we simply get $F(x_1) = x_1 - 1$, capturing the Nash equilibrium fixed point $x_1 = 1$. (Note, though, that in this extension none of the other fixed points at $\epsilon = 0$ are captured; any $x_k = 1$ defines a fixed point for zero mutation rate, but all except the first one are of no interest for us.)

All functions f_A , f_B and f_C are continuous and bounded, since $g(y) = y(1 - \sqrt{1 + 1/y^2})$ is. We also see that $g(y) \rightarrow 0$ as $|y| \rightarrow \infty$. Thus, $F(x)$ being composed of these functions is continuous in the interval $[0, 1]$.

As an example, Figure 6 illustrates $F(x)$ for $n = N + 1 = 6$, $T = 1.05$, $P = 0.15$ with four different choices of the mutation rate ϵ . The graph shows that as the mutation rate increases the fixed point at $x_1 = 1$ moves to the left and also that new fixed points may emerge. Most importantly, though, the graph indicates a discontinuous change of the function for x below about 0.95, when going from $\epsilon = 0$ to $\epsilon > 0$. In order to show that the fixed point always moves continuously from $x_1 = 1$, we need to be sure that any such a discontinuity is bounded away from $x_1 = 1$.

We already know that for $\epsilon = 0$ there is a zero in $x = (1, 0, \dots, 0)$. We want to show that as ϵ increases, this fixed point, characterised by $F(x_1) = 0$ for $x_1 = 1$, continuously moves into the unit

interval, $0 < x_1 < 1$. We accomplish this by performing a series expansion of $F(x)$ around $x = 1$ and $\epsilon = 0$, by showing that $F'(x_1) = 1$ for $x_1 = 1$, that F is increasing with ϵ in the neighbourhood of $x_1 = 1$, and that the coordinates x_k ($k > 1$) of the fixed point move continuously with x_1 when sufficiently close to 1.

The first part is straightforward: At zero mutation rate, $\epsilon = 0$, we have already noted that $F(x_1) = x_1 - 1$, and thus the derivative $F'(x_1) = 1$.

Next, we need to show that, at least sufficiently close to the fixed point and for sufficiently small ϵ , F increases with ϵ . We do this term by term in the sum of F . First assume that the point x_1 is close to fixed point value 1 at zero mutation rate, *i.e.*, $x_1 = 1 - \delta$, and that ϵ is small, so that $\epsilon \ll P$ and $\delta \ll P$. The first term in F is x_1 and does not depend on ϵ . The second term, x_2 , is given by f_A ,

$$\begin{aligned} x_2 = f_A(1 - \delta) &= \frac{1}{2(T - P)} \left(P - \delta + \frac{\epsilon'}{1 - \delta} \right) \left(-1 + \sqrt{1 + \frac{4(T - P)\epsilon'}{(P - \delta + \frac{\epsilon'}{1 - \delta})^2}} \right) = \\ &= \frac{\epsilon'}{P} \left(1 + \frac{\delta}{P} + O(\epsilon') \right) + O(\epsilon')^2 + O(\delta)^2 \end{aligned} \quad (18)$$

This term thus increases with ϵ' . For the next term x_3 , using Equation (13), we find that to first order in ϵ' and δ ,

$$\begin{aligned} x_3 = f_B(x_2, 1 - x_1 - x_2) &= f_B \left(\frac{\epsilon'}{P}, \delta - \frac{\epsilon'}{P} \left(1 + \frac{\delta}{P} \right) \right) = \\ &= \frac{\epsilon'}{P} \left(1 + \frac{2 - P}{P} \delta + O(\delta)^2 \right) + O(\epsilon')^2 + O(\delta)^2 \end{aligned} \quad (19)$$

More generally, for $2 < k < n$, repeatedly using Equation (13), we find

$$x_k = \frac{\epsilon'}{P} \left(1 + \frac{k - P(k - 1)}{P} \delta + O(\delta)^2 \right) + O(\epsilon')^2 + O(\delta)^2 \quad (20)$$

and finally, using Equation (16), that

$$x_n = \frac{\epsilon'}{P} \left(1 + \frac{(N - 1) - P(N - 2)}{P} \delta + O(\delta)^2 \right) + O(\epsilon')^2 + O(\delta)^2 \quad (21)$$

From the linearisation we can conclude that when x_1 is sufficiently close to 1, *i.e.*, $\delta \ll P$, the change in $F(x_1)$ from an increase of ϵ' is given by the $n - 1$ terms x_2, \dots, x_n , or $dF(x_1)/d\epsilon' \sim (n - 1)/P > 0$. Adding the fixed point constraint, $F(x_1) = 0$, to the linearisation determines x_1 and thus δ in terms of ϵ' . To first order in ϵ the fixed point is given by

$$x_1 \sim 1 - (n - 1) \frac{\epsilon'}{P} + O(\epsilon')^2 = 1 - (n - 1) \frac{\epsilon}{nP} + O(\epsilon)^2, \text{ and} \quad (22)$$

$$x_k \sim \frac{\epsilon}{nP} + O(\epsilon)^2 \text{ for } k = 2, \dots, n \quad (23)$$

where we have switched back to the original ϵ . This analysis shows that as ϵ increases from 0, the fixed point gradually moves into the unit interval.

Since the fixed point is changed continuously as the mutation rate increases from $\epsilon = 0$, the eigenvalues of the Jacobian also change continuously. The fixed point at $\epsilon = 0$ is stable, with the largest

eigenvalue being $\lambda_{\max} = -P < 0$, and we can conclude that for sufficiently small $\epsilon > 0$, the real part of the largest eigenvalue remains negative. From this we can conclude that *the fixed point associated with the Nash equilibrium in the finitely repeated Prisoner's Dilemma remains stable in the simple strategy set if the mutation rate is sufficiently small*. This concludes the proof of Proposition 1.

References

1. *Handbook of Computational Economics*, 1st ed.; Tesfatsion, L.S., Judd, K.L., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; Vol. 2.
2. Arthur, W.B.; Holland, J.H.; Lebaron, B.; Palmer, R.; Tayler, P. Asset pricing under endogenous expectations in an artificial stock market model. In *The Economy as an Evolving Complex System II*; Arthur, W.B., Durlauf, S.N., Lane, D.A., Eds.; Addison-Wesley: Boston, MA, USA, 1997; pp. 15–44.
3. Samanidou, E.; Zschischang, E.; Stauffer, D.; Lux, T. Agent-based models of financial markets. *Rep. Prog. Phys.* **2007**, *70*, 409–450.
4. Bosetti, V.; Carraro, C.; Galeotti, M.; Massetti, E.; Tavoni, M. A world induced technical change hybrid model. *The Energy Journal* **2006**, *0*, 13–38.
5. Parker, D.C.; Manson, S.M.; Janssen, M.A.; Hoffmann, M.J.; Deadman, P. Multi-agent systems for the simulation of land-use and land-cover change: A review. *Ann. Assoc. Am. Geogr.* **2003**, *93*, 314–337.
6. Simon, H.A. A Behavioral Model of Rational Choice. *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*; Wiley: New York, NY, USA, 1957; pp. 99–118.
7. Camerer, C. *Behavioral Game Theory: Experiments in Strategic Interaction*; The Roundtable series in behavioral economics; Russell Sage Foundation: New York, NY, USA, 2003.
8. Aumann, R.J. Rule-Rationality versus Act-Rationality. Discussion paper series; Center for Rationality and Interactive Decision Theory, Hebrew University: Jerusalem, Israel, 2008.
9. Binmore, K. Modeling rational players: Part I. *Economics and Philosophy* **1987**, *3*, 179–214.
10. Binmore, K. Modeling rational players: Part II. *Economics and Philosophy* **1988**, *4*, 9–55.
11. Nash, J.F. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America*, **1950**, *36*, 48–49.
12. Pettit, P.; Sugden, R. The backward induction paradox. *J. Phil.* **1989**, *86*, pp. 169–182.
13. Bosch-Domnech, A.; Montalvo, J.G.; Nagel, R.; Satorra, A. One, two, (three), infinity, ... : Newspaper and lab beauty-contest experiments. *Am. Econ. Rev.* **2002**, *92*, pp. 1687–1701.
14. Basu, K. The traveler's dilemma: Paradoxes of rationality in game theory. *Am. Econ. Rev.* **1994**, *84*, pp. 391–395.
15. Selten, R.; Stoecker, R. End behavior in sequences of finite Prisoner's Dilemma supergames A learning theory approach. *J. Econ. Behav. Organ.* **1986**, *7*, 47–70.
16. Schuster, P.; Sigmund, K. Replicator dynamics. *J. Theor. Biol.* **1983**, *100*, 533–538.
17. Hofbauer, J. The selection mutation equation. *J. Math. Biol.* **1985**, *23*, 41–53.
18. Nachbar, J.H. Evolution in the finitely repeated prisoner's dilemma. *J. Econ. Behav. Organ.* **1992**, *19*, 307–326.

19. Cressman, R. Evolutionary stability in the finitely repeated Prisoner's Dilemma game. *J. Econ. Theory* **1996**, *68*, 234–248.
20. Cressman, R. *Evolutionary Dynamics and Extensive Form Games*; MIT Press Series on Economic Learning and Social Evolution; MIT Press: Cambridge, MA, USA, 2003.
21. Noldeke, G.; Samuelson, L. An evolutionary analysis of backward and forward induction. *Games Econ. Behav.* **1993**, *5*, 425–454.
22. Cressman, R.; Schlag, K. The dynamic (in)stability of backwards induction. *J. Econ. Theory* **1998**, *83*, 260–285.
23. Binmore, K.; Samuelson, L. Evolutionary drift and equilibrium selection. *Rev. Econ. Stud.* **1999**, *66*, 363–93.
24. Hart, S. Evolutionary dynamics and backward induction. *Games Econ. Behav.* **2002**, *41*, 227–264.
25. Hofbauer, J.; Sandholm, W.H. Survival of dominated strategies under evolutionary dynamics. *Theor. Econ.* **2011**, *6*, 341–377.
26. Gintis, H.; Cressman, R.; Ruijgrok. Subgame perfection in evolutionary dynamics with recurrent perturbations. In *Handbook of Research on Complexity*; Barkley Rosser, J., Ed.; Edward Elgar Publishing: Northampton, MA, USA, 2009; pp. 353–368.
27. Ponti, G. Cycles of learning in the centipede game. *Games Econ. Behav.* **2000**, *30*, 115–141.
28. Rosenthal, R. Games of perfect information, predatory pricing and the chain-store paradox. *J. Econ. Theory* **1981**, *25*, 92–100.
29. Aumann, R.J. Backward induction and common knowledge of rationality. *Games Econ. Behav.* **1995**, *8*, 6–19.
30. Aumann, R.J. Reply to Binmore. *Games Econ. Behav.* **1996**, *17*, 138–146.
31. Binmore, K. Rationality and backward induction. *J. Econ. Methodol.* **1997**, *4*, 23–41.
32. Gintis, H. Towards a renaissance of economic theory. *J. Econ. Behav. Organ.* **2010**, *73*, 34–40.