Cooperation in an Unpredictable Environment

Anders Eriksson and Kristian Lindgren
Department of Physical Resource Theory — Complex Systems Group
Chalmers and Göteborg University
SE-412 96 Göteborg, Sweden.

Abstract

A framework for studying the evolution of cooperative behaviour in a random environment, using evolution of finite state strategies, is presented. The interaction between agents is modelled by a repeated game with random observable payoffs. The agents are thus faced with a more complex situation, compared to the Prisoner's Dilemma that has been widely used for investigating the conditions for cooperation in evolving populations (Matsuo 1985; Axelrod 1987; Miller 1989; Lindgren 1992; Ikegami 1994; Lindgren & Nordahl 1994; Lindgren 1997). Still, there are robust cooperating strategies that usually evolve in a population of agents. In the cooperative mode, these strategies selects an action that allows for maximizing the payoff sum of both players in each round, regardless of the own payoff. Two such players maximize the expected total long-term payoff. If the opponent deviates from this scheme, the strategy invokes a punishment action, which aims to lower the opponent's score for the rest of the (possibly infinitely) repeated game. The introduction of mistakes to the game actually pushes evolution towards more cooperative strategies even though the game becomes more difficult.

Cooperation can be characterised as an interaction where agents refrain from taking short-term profit and instead act in a way that allows for a larger long-term gain. In this sense, cooperative behaviour may be seen at several levels in natural systems. As discussed by Maynard-Smith and Szathmáry (1995), cooperation may be one crucial factor for the major transitions that has occurred in the evolution of life on Earth — the creation of the more complex eukaryotic cell, the formation of multi-cellular organisms, the appearance of social animals, etc. In these transitions, previously separately reproducing components form a new self-reproducing entity, in which the different parts may have specialised roles.

In the field of Artificial Life, we have the opportunity to create models that allow for such major evolutionary transitions, but so far they seem to be missing in the models studied. One approach to get an increased knowledge about how such transitions may appear is to investigate under what circumstances cooperative behaviour may be advantageous for a replicating entity, and how cooperation may be achieved. During the past two decades there have been a large number of papers discussing the evolution of cooperation in the perspective of the Prisoner's Dilemma game. The work by Axelrod (1984) was a starting point for a series of papers putting an evolutionary perspective to how cooperation is established. The main conclusion from the initial work was that cooperation could be established if interactions between individuals are repeated. A large number of modifications and extensions of the PD was tried to firmly establish the fact that cooperation is possible under a wide variety of circumstances (Matsuo 1985; Molander 1985; Axelrod 1987; Boyd & Lorberbaum 1987; Boyd 1989; Miller 1989; Lindgren 1992; Nowak & May 1993; Stanley & Tesfatsion 1993; Ikegami 1994; Lindgren & Nordahl 1994; Nowak & El-Sedy 1995; Wu & Axelrod 1995; Lindgren 1997).

The possibility for a cooperative equilibrium to be established in a repeated game has long since been well known in game theory. The Folk Theorem states that in a repeated game, with sufficiently low probability for the game to end, any possible score above the min-max payoff can be supported at equilibrium by some strategy (Fudenberg & Tirole 1991; Binmore 1994). Such an equilibrium is kept by punishing those who deviate from the equilibrium strategy. One example of punishment is to minimise the possible score for the opponent in the present round.

One of the main limitations with the Prisoner's Dilemma game and many of the similar games studied is the static character of the interaction situation. Each time two individuals encounter each other, the situation is identical to the previous one, i.e., the payoff values for the different choices are unchanged. In a real situation, it may be much more common that they meet each other in different situations most of the time; some situations may be of the Prisoner's Dilemma type, other situations may be easy, so short-sighted profit maximisation coincides with the cooperative maximal payoff.

The advantage with the PD game is its simplicity, allowing for various extensions that investigate different

characteristics of cooperative behaviour in simple models. Our aim with the current research is to keep the simplicity at the same time as the situation in which agents meet are less static. Therefore we generate a completely new random payoff matrix for each round in a repeated two-player game, but the players still have full information on the current payoff values for the different actions and players.

The questions are: How can cooperation emerge? What will it look like? Under what circumstances is it stable? In order to investigate possible answers to these questions, we have constructed an evolutionary model in which strategies replicate and mutate, with a replication rate depending on the score achieved in the game. Two types of dynamics are studied. First, in the mixed population, all interact with all and second, in the spatially extended model, interactions are restricted to the four nearest neighbours. We also study the effect of mistakes, as it is a complication for cooperation, and to what extent strategies may evolve to take care of this.

The random payoff game

The repeated random payoff game used in this study is a two-person game in which each round is characterized by two possible actions per player and a randomly generated but observable payoff matrix. The payoff matrix elements are random, independent, and uniformly distributed real numbers between zero and one. New payoffs are drawn every round of the game.

In the single round game a player performs a certain action that may depend on the observed payoffs. Therefore, the single round game can be characterised by its Nash equilibria (NE), i.e., a pair of actions such that if only one of the players switches action that player will reduce her payoff[1].

There are a number of simple single round (elementary) strategies that are of interest in characterising the possible types of behaviour. Assume first that there is exactly one NE, and that rational (single round) players play the corresponding actions. The payoff in this case is $\max(x, h)$, where x and h are independent uniformly distributed stochastic variables between zero and one, and this results in an expectation value of $2/3 \approx 0.667$.

Let us define a strategy "NashSeek" as follows. If there is only one NE in the current payoff matrix, one chooses the corresponding action. If there are two NE, one aim for the one that has the highest sum of the two players payoffs, while if there is no NE, one optimistically chooses the action that could possibly lead to the highest own payoff.

A second strategy, "MaxCoop", aims for the highest sum of both players' payoffs. If two such players meet, they score $\max(x_1+h_1, x_2+h_2, x_3+h_3, x_4+h_4)$ together, where x_i and h_i are independent uniformly distributed stochastic variables between zero and one, and this results in an expectation value of $s_C = 3589/5040 \approx 0.712$.

A third more optimistic and greedy strategy is "MaxMax", which selects the action that makes it possible to get the highest score provided that the opponent acts accordingly. Finally, we have also chosen to include a strategy that is punishing the opponent. The strategy Punish selects the action that minimizes the opponent's maximum payoff.

In an evolutionary model based on a round robin tournament for determining the fitness (based on the score of all pair-wise single round games) it is clear that NashSeek will take over sooner or later. If we extend the game, so that there is a high probability that players will encounter each other repeatedly in new situations (new random payoffs), other possibilities may appear. This requires, though, that the players may use composite strategies that in some way take into account what has happened before in the interaction with the same opponent.

A straightforward approach is to define composite strategies represented by deterministic finite state automata (FSA) (Miller 1989; Nowak & El-Sedy 1995; Lindgren 1997), in which a state corresponds to a certain elementary strategy as described above. From one round of the game to the next, a player may switch to a different internal state (and different type of behaviour) depending on what happened in the previous round. The opponent's action is classified in terms of the basic strategies, and it is checked whether the action is consistent with any of the elementary strategies. Different composite strategies will then react in different ways to the interpreted behaviour of the opponent. See Figure 1 for a sample strategy.

Mixed population dynamics

We consider a population of N agents, competing for the same resources. The population is at the limit of the carrying capacity level of the environment, so the number of agents N is fixed. In each generation, the agents play the repeated stochastic payoff game with the other agents, and reproduce to form the next generation. The population evolves according to the ordinary replicator dynamics (Taylor & Jonker 1978): the score for every composite strategy is compared to the average score of the population, and those above average will get more offspring, and thus a larger share in the next generation.

In the simulations of the model, in a few thousand generations, the average score of the population increases to a level corresponding to the single round Nash equilibrium. The evolution later on continues with a fast transition to the level of the MaxCoop score, indicating that

[1] Since agents cannot randomise their actions, we only consider pure strategy Nash equilibria. Depending on the payoffs there are either zero, one, or two pure NEs in this game, appearing with probabilities 1/8, 3/4, and 1/8, respectively.

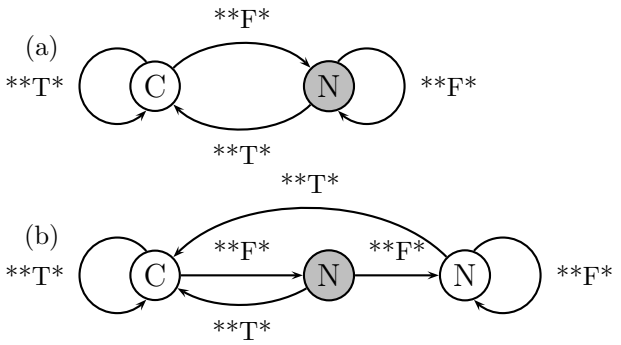

Figure 1: Mutation of a composite strategy into an equivalent strategy with an extra internal state is shown, from the parent (a) to the child (b). The letters in the nodes denote the elementary strategies MaxCoop (C) and NashSeek (N), and the black circle indicates the initial state. Depending on the interpretation of the opponent's previous action, the composite strategy may switch internal state. The transition **T* applies only when the opponent's action is consistent with MaxCoop (the third elementary strategy), but does not care about the other elementary strategies. The transition rule **F* is applied when the opponent does not follow the MaxCoop strategy. This composed strategy is similar to the Tit-for-tat strategy discussed in the context of the iterated Prisoner's Dilemma in that it cooperates by playing MaxCoop if the opponent played according to MaxCoop in the last round, otherwise it 'defects' by playing NashSeek. There are several different ways to change a strategy by mutations: by adding or removing states, changing the connectivity of the states, or by changing the actions in the states. Shown here is addition of a new state, by duplicating the grey node in such a way that the strategy's play is unchanged.

a mechanism for avoiding the shortsighted single round Nash seeking behaviour has evolved. As is seen in Figure 2, showing a typical evolutionary pattern after the initial transient, the population does not seem to be able to stabilize on the cooperative level, but there are several sharp drops in the score when mutant strategies manage to take a more substantial part of the population. In some cases the score is brought down to the level of the Nash seeking strategy again.

When the population switches to the level of the MaxCoop score it is dominated by cooperating strategies that punish deviators like the Nash seekers. This results in the near extinction of the Nash seeking strategies, which means that the punishment mechanism is not needed any longer. Therefore, mutants that are cooperative but have lost the capability to punish may survive and may in some cases even have an advantage, since it may be costly to punish, depending on the exact mechanism. In Figure 2 it is clearly seen that the fraction of composite strategies that keep the punishment mechanism oscillates irregularly. If the fraction of punishers is too low the population becomes vulnerable to invasion by Nash seekers. If Nash seekers become common, cooperating strategies with punishment again has an advantage and the process repeats.

Under what circumstances could the cooperative level be stabilised? If the situation is changed so that mutants that have lost their punishment mechanism have a disadvantage, it should be possible to achieve a stable cooperating population. One situation in which this may happen is when noise in the form of mistakes is present in the repeated game.

We model mistakes as a probability ρ of taking the action opposite to the intended one. Since we consider infinite games, even a very small chance of mistakes has a huge impact on the overall payoffs and how the strategies evolve, even when the effect on the expected payoffs in the single round games are negligible (this effect is approximately proportional to ρ). Strategies that punish deviations from a mutual elementary strategy must now take the possibility of mistakes into account (Lindgren 1992). A pair of strategy like the one in Figure 1a, for example, will perform much worse since it will play NashSeek much more often. Especially when playing against itself, it will inevitably end up with long sequences of mutual "defects" when both players play NashSeek, and they will only get back on the cooperating track again (playing MaxCoop) if new mistakes occurs. It is thus essential that the players have some way of recovering from unintentional actions, from both parties. To this end, the players now monitor their own actions, as well as the actions of the opponents.

A player still cannot detect when the opponent makes a mistake, but by "apologising" after a mistake in some appropriate way (e.g. by playing MaxCoop) the long run cooperation might be preserved. This mechanism may leave the player open for greedy players to make use of it, though, so it is more complicated to find an appropriate mechanism.

It is clearly seen in Figure 3, that in the case of mistakes the population stabilises in the cooperative mode. Since cooperation now requires both a punishment mechanism to avoid exploiting strategies and a mechanism for restoring cooperation after a mistake, it takes a longer time before cooperation is achieved. It should be noted that there is no single simple mechanism present in the cooperative population, but the dominating strategies can be characterised as forgiving punishers. There is neither a synchronisation nor a handshaking mechanism involved, as has been previously observed in the case of the Prisoner's Dilemma game with mistakes (Lindgren 1992; 1997). It should be noted that the mistake probability is very low (10^{-4}), and if the level is increased cooperation is much harder to achieve, in the population dynamics.

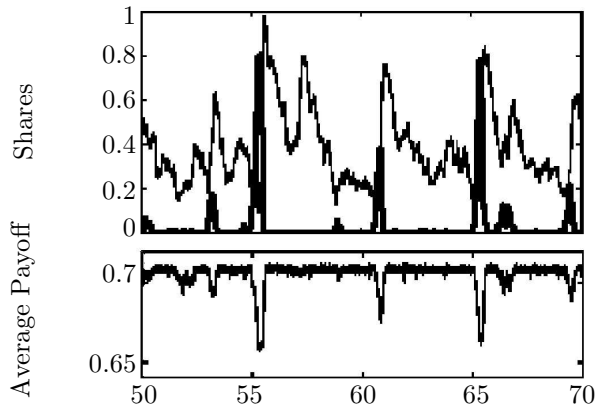

Figure 2: The figure shows a typical run of the model illustrating how non-punishing strategies take over from cooperating strategies with punishment, and the consequent collapses due to exploiting strategies. In the top is shown the share of players that play like MaxCoop when meeting others with the same behaviour but punish those that deviate from MaxCoop (thin line) and the share of exploiters (thick line). The bottom part shows the average payoff of the population. When the average payoff of the population is high, the share of strategies with a punishment mechanism decreases, due to an increase of mutant strategies without such a mechanism. Exploiting strategies are held back by the punishments. When the share of punishing strategies fall below a critical level, exploiters get payoffs above the population average and quickly come to dominate the population. This drastically lowers the population average payoff, which gives the cooperating-punishing strategies the edge, so that they can take over the population. This cycle then repeats.

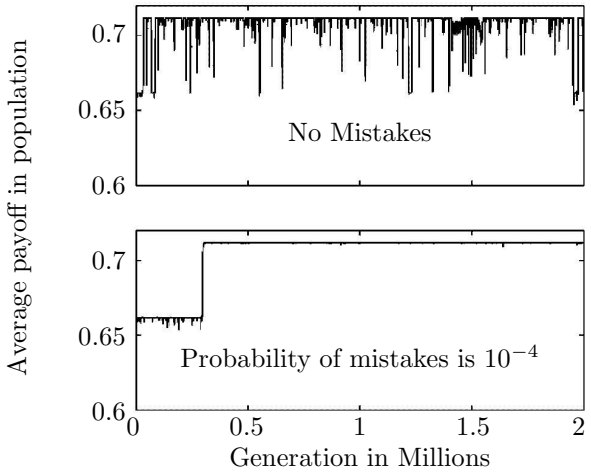

Figure 3: Typical runs without mistakes (top), and with mistakes (bottom). With mistakes, it is harder for the mutations to find a solution to the problem of cooperating by playing MaxCoop when possible, while evading exploiting strategies. The mutations of the resulting strategies are much less likely to be fit, and this imply that it is difficult for exploitable strategies to enter the population, as happens when there are no mistakes.

A more thorough investigation of the game with mistakes has been carried out and will be presented elsewhere.

Spatial population dynamics

In order to study the consequences of a small, local interaction neighbourhood on the evolution of strategies, we put each player on a site in a square lattice with periodic boundary conditions. The fitness of a player in a cell is given by the average of the expected payoffs from one by one games with its four neighbours. A cell is updated by replacing the player in the cell by the player with the highest fitness among the players in the cell and those in the neighbouring cells. In each generation, all cells are updated once, but the order is chosen randomly each time. In each generation, there is a probability of having mutations in the population, generated according to a Poisson process with a mean of m events per generation.

A typical run is shown in figure 4, with a mistake rate $\rho = 0.01$. The world is 200 by 200 cells, and the mutation rate is 20 events per generation. The maximum number of strategies simultaneously in the population is 600 strategies. In the first few generations, the population converges to the NashSeek level. About a thousand generations later we see the first emergence of cooperation, with a short transition.

However, the cooperating strategies are easily exploited and we see a steady decline in the average payoff. The decline is roughly linear; the rate is limited by the mutation rate and the maximum number of strategies. During this period, the number of strategies rises very quickly. The mutants enter the square in small lumps, and do not tend to increase in size but do not die out either.

At some stage, and before the payoff reaches the NashSeek level, the trend turns and the payoff starts to rise again. As opposed to the initial, fast increase, this climb is gradual and consists of a constant replacement of strategies; there is typically no single strategy that takes over a large fraction of the population.

After a few ups and downs with much smaller magnitude than during the transient phase, the average payoff settles on a high level of cooperation. The payoff p_{CC} for the pure MaxCoop strategy is 0.7065 when the mistake rate is $\rho = 0.01$, and this corresponds to the maximum average payoff that can be stable. Note that the payoff is between 0.705 and 0.706, so it is quite close to the maximum level.

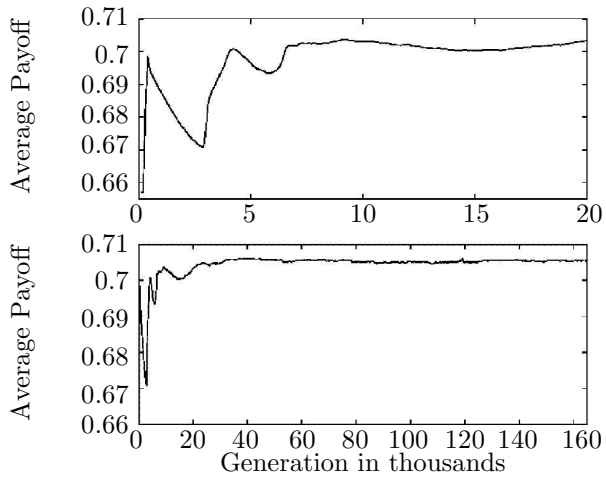

Figure 4: Time evolution of the average payoffs for a typical run. The top panel shows the initial transient, while the lower panel shows the convergence to a high, stable level of cooperation. The mistake rate is $\rho = 0.01$ and the mutation rate is 20 events per generation. Note that cooperation emerges readily on a level of mistakes where cooperation seems to be very hard to achieve in the mixed populations.

We have found that the populations evolve to high levels of cooperation, even for ρ up to and over 0.01. The level of cooperation is decreasing with increasing mistakes rate and discount rate, and the variation of this level between different runs is increasing with ρ. Also when ρ is zero, the level of cooperation is high and stable and we do not see the kind of fluctuations present in the simulations of the mixed population. This is consistent with results from the literature on spatial population dynamics in the Prisoner's Dilemma game (Lindgren & Nordahl 1994; Lindgren 1997).

Figure 5 shows a snapshot of the world at a given generation with different shades identifying different strategies. Figure 6 shows the fitness of the players encoded in a grey scale, given in the shaded bar to the right of the image. Note that the mapping is non-linear; the intensity of cell i is calculated as $g_i = 1/(p_{\mathrm{NC}} - f_i)$, then scaled to so that all intensities fit into the interval [0,1].

Summary

In summary, in the case without mistakes we find that there are several strategies that play on equal terms with a MaxCoop strategy that punishes exploiting strategies. For example, strategies that by mutation lose their punishment mechanism may enter and increase their fraction of the population by genetic drift. This in turn leads to a population that is vulnerable to mutants exploiting the cooperative behaviour, for example by strategies playing NashSeek. Thus, the Nash equilibrium that characterises the population dominated by MaxCoop-punishment is not an evolutionarily stable one, as can be seen in the simulations. The effect is the same as the one in the iterated Prisoner's Dilemma that makes Tit-for-tat an evolutionarily non-stable strategy (Boyd & Lorberbaum 1987; Boyd 1989).

This genetic drift is prevented by the introduction of mistakes in the actions of the elementary strategies. The strategies cooperating at a high payoff level have evolved mechanisms that can simultaneously protect the players from greedy strategies, and lead the game back to cooperation when a mistake occurs.

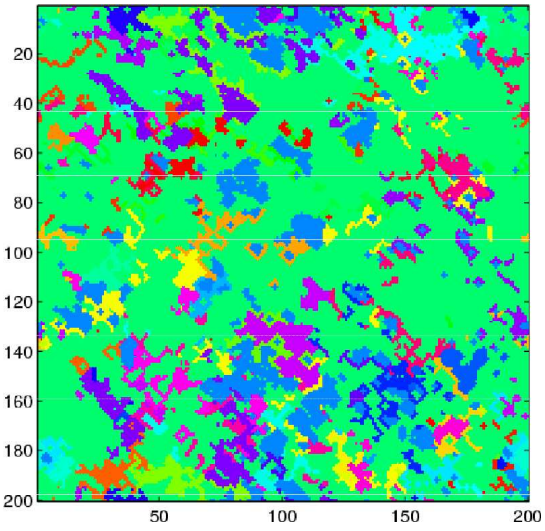

Figure 5: A snapshot of the distribution of strategies among the players at generation 162,000. The identity of the strategies are indicated by the different colours; there is no further significance to the choice of colours. Note the pattern of areas with the same strategy, interspersed by singlet mutations.

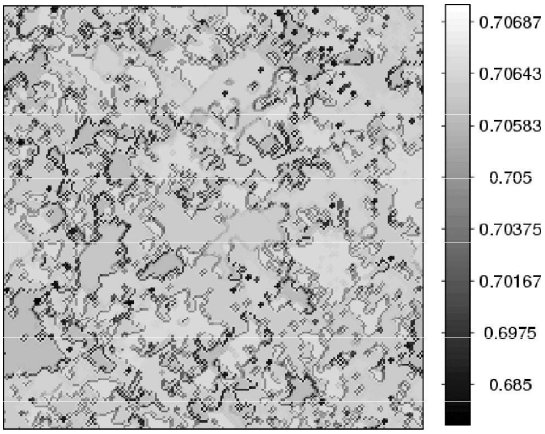

Figure 6: The fitness values of the players at generation 162,000, corresponding to figure 5. The fitness of the players is encoded in grey scale, given by the bar to the right of the image.

The evolution of a population dominated by MaxCoop (with some punishment mechanism) is not unexpected. A population of players (using identical strategies) can always enforce a certain score s^*, if this score is larger than (or equal to) the smallest punishment score s_P and smaller than (or equal to) the maximum cooperation score $s_C, s_P \leq s^* \leq s_C$. This means that the population can punish a new strategy (mutant) that enters, unless it adopts the same strategy and plays so that the score s is reached in the long run. This argument follows the idea behind the Folk Theorem that appears in various forms in the game-theoretic literature (Fudenberg & Tirole 1991; Dutta 1995). As has been seen also in experimental economics, punishment seems to be a key mechanism for humans in order to force selfish individuals to behave in a more cooperative way (Fehr & Gächter 2000b; 2000a; 2002).

In the spatial setting, the players are distributed on a lattice. We find that the possibility to reach cooperation is greatly enhanced, compared to the mixed population dynamics. As the mistake rate and the discount rate increases, the average level of cooperation is decreasing, while the variation in this level is increasing. Still, cooperation at fairly high levels has been observed for mistakes rates of 0.01.

When the mistake rate is zero, the spatial model does not exhibit the type of fluctuation present in the mixed population — in the spatial model, a high level of cooperation can be maintained. This difference has been observed also in studies of the PD game (Nowak & May 1993; Lindgren & Nordahl 1994; Lindgren 1997). A common pattern that appears is islands or regions of cooperating behaviour with edges or borders of non-cooperating strategies. When a strategy is replaced it is more often by a strategy from the interior, i.e., by a cooperating strategy.

The repeated game with stochastic observable payoffs offers a simple model world in which questions on the evolution of cooperation may be investigated. The model captures the uncertainty about the future situations we may find our opponents and ourselves in, and we think this may be a useful basis for further investigations of the circumstances under which cooperation may evolve.

References

Axelrod, R. 1984. *The Evolution of Cooperation.* New York: Basic Books.

Axelrod, R. 1987. The evolution of strategies in the iterated prisoner's dilemma. In *Genetic Algorithms and Simulated Annealing.* D. L. Los Altos, CA: Morgan Kaufmann. 32 – 41.

Binmore, K. 1994. *Playing Fair.* Cambridge, Mass.: MIT Press.

Boyd, R., and Lorberbaum, J. P. 1987. No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature* 327:58 – 59.

Boyd, R. 1989. Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. *Journal of Theoretical Biology* 136:47 – 56.

Dutta, P. K. 1995. A folk theorem for stochastic games. *Journal of Economic Theory* 66:1 – 32.

Fehr, E., and Gächter, S. 2000a. airness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14:159 –181.

Fehr, E., and Gächter, S. 2000b. Cooperation and punishment in public goods experiments. *American Economic Review* 90:980 – 994.

Fehr, E., and Gächter, S. 2002. Altruistic punishment in humans. *Nature* 415:137 – 140.

Fudenberg, D., and Tirole, J. 1991. *Game Theory.* Cambridge, MA: MIT Press.

Ikegami, T. 1994. From genetic evolution to emergence of game strategies. *Physica D* 75:310 – 327.

Lindgren, K., and Nordahl, M. G. 1994. Evolutionary dynamics of spatial games. *Physica D* 75:292 – 309.

Lindgren, K. 1992. Evolutionary phenomena in simple dynamics. In Chris Lanton, e. a., ed., *proceedings Artificial Life II*, 295 – 311. Addison-Wesley.

Lindgren, K. 1997. Evolutionary dynamics in game-theoretic models. In B. Arthur, S. D., and Lane, D., eds., *The economy as an evolving complex system II*, 337 – 367. Addison-Wesley.

Matsuo, K. 1985. Ecological characteristics of strategic groups in 'dilemmatic world'. In *proceedings IEEE International Conference on Systems and Cybernetics.*

Maynard-Smith, J., and Szathmáry, E. 1995. *The Major Transitions in Evolution.* Oxford: Oxford University Press.

Miller, J. H. 1989. The coevolution of automata in the repeated iterated prisoner's dilemma. Technical Report 89-103, Santa Fe Institute.

Molander, P. 1985. The optimal level of generosity in a selfish, uncertain environment. *Journal of Conflict Resolution* 29:611 – 618.

Nowak, M. A., K. S., and El-Sedy, E. 1995. Automata, repeated games and noise. *Journal of Mathematical Biology* 33:703 – 722.

Nowak, M. A., and May, R. M. 1993. Evolutionary games and spatial chaos. *Nature* 359:826 – 829.

Stanley, E. A., D. A., and Tesfatsion, L. 1993. Iterated prisoner's dilemma with choice and refusal of partners. In Langton, C. G., ed., *proceedings Artificial Life III*, 131–175. Reading, MA: Addison-Wesley.

Taylor, P., and Jonker, L. 1978. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences* 16:7683.

Wu, J., and Axelrod, R. 1995. How to cope with noise in the iterated prisoner's dilemma. *Journal of Conflict Resolution* 39:183 –189.