

The Localization of Interaction and Learning in the Repeated Prisoner's Dilemma

Robert Hoffmann
Nigel Waring

SFI WORKING PAPER: 1996-08-064

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu

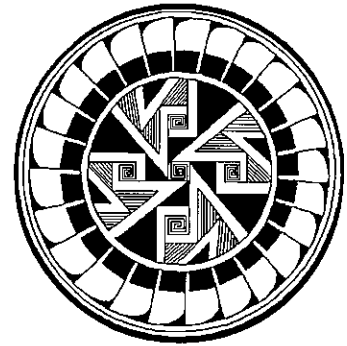


SANTA FE INSTITUTE

The Localization of Interaction and Learning in the Repeated Prisoner's Dilemma

Robert Hoffmann and Nigel Waring

96-08-064



To appear as "The Simulation of Localized Interaction and Learning in Artificial Adaptive Agents." In *Evolutionary Computing*, edited by T. C. Fogarty. Springer-Verlag, 1996.

A Working Paper for the SFI Economics Research Program. This research was undertaken at the 1995 SFI Graduate Workshop in Computational Economics.

The Localisation of Interaction and Learning in the Repeated Prisoner's Dilemma¹

Robert Hoffmann² and Nigel Waring³

February 15, 1996

Abstract

The evolution of cooperation in the repeated prisoner's dilemma depends on the conditions under which the game is played. The results of a series of computer simulations show that the emergence of cooperative play in the game is strongly affected by the localisations of both interaction and learning. While the localisation of learning promotes mutual cooperation, the localisation of interaction has an ambiguous effect on it. Moreover, the localisation of learning seems to promote disequilibrium behaviour in the system.

1 Introduction

Despite four decades of intensive research, the growth in the literature on the repeated prisoner's dilemma game (RPD) is showing no signs of abating. While many social scientists still perceive a need to complete their understanding of the evolution of cooperation among selfish and rational individuals, other disciplines are increasingly jumping on the bandwagon for their own reasons. In biology, the game has been used as a representation of the problem of reciprocal altruism between animals and organisms (see Axelrod and Hamilton [7], Maynard Smith [25], Trivers [41]). The computer sciences are testing novel techniques to model human decision-making and interaction, such as

¹The authors wish to thank Shaun Hargreaves Heap, John Miller, George Smith, Bob Sugden as well as participants of the SFI 1995 Graduate Workshop in Computational Economics.

²School of Management and Finance, University of Nottingham, UK. Email: lizrh2@lin1.nott.ac.uk

³School of Information Systems, University of East Anglia, UK. Email: ncw@sys.uea.ac.uk

neural nets and classifier systems, through their application to this well-studied scenario. The exchange between the different disciplines involved in researching the RPD can be highly rewarding. Although the present study falls into the first category, it has evolved out of the collaboration between an economist and a computer scientist.

	c	d
c	R,R	S,T
d	T,S	P,P

Figure 1: The Prisoner's Dilemma

The prisoner's dilemma (figure 1) is a symmetric pairwise game in which each player has the option to cooperate (c) or to defect (d). The four payoffs of the game obey the simple conditions $T > R > P > S$ and $2R > T + S$. For games in which the last round is common knowledge, dd in every round is the only Nash equilibrium. The conundrum of the game is the fact that the resulting equilibrium of mutual defection (dd) is pareto-dominated by mutual cooperation (cc). Despite this compelling logical finding, mutual cooperation in the game has been observed both in laboratory experiments (Dawes and Thaler [11]; for surveys of experimental activity, see Rapoport and Orwant [33], Roth [34] and Poundstone [32]) and in many social and economic settings described by the game (see Dawes and Thaler [11], Axelrod [3]). Hence, the research agenda in this field is to identify the circumstances under which mutual cooperation can be sustained in the game. Although early results established that mutual cooperation is possible in the infinitely repeated game (see Fudenberg and Maskin [14]) as well as in the finitely repeated game with incomplete information (Kreps *et al.* [21]), there is no consensus on *how* this cooperation can be ensured. Specifically, these results stipulate cooperation as only one possible of a host of multiple equilibria.

One source of this difficulty has been identified in the conventional assumption of instrumental rationality. It is held by some authors that cognitive limitations prevent real agents to gather, process and store information in order to be able to behave rationally in terms of the theory (Simon [38, 39]). Instead, agents are seen to display 'bounded rationality', that is, employ decision-making routines updated by emulative and experimental learning on the basis of their performance. In game theory, routines are modelled as set strategies that are adopted on the basis of the payoffs they generate.

This type of agent learning is suitably modelled in an evolutionary context (Maynard Smith [25], Mailath [23], Hodgson [16]), where alternative strategies diffuse in populations of players according to the principles of (artificial) selection and survival of the fittest. Evolutionary game theory analyses this type of interaction and learning among populations of boundedly rational agents paired successively to play repeated games.

Work in this area has also encountered the problems of demonstrating the existence and uniqueness of cooperative equilibria (Boyd and Lorberbaum [10]). In response, some authors have abandoned the equilibrium analysis of the game in favour of an investigation of the *process* of play leading towards stability. However, progress has been hampered by problems of computation in the analysis of the models. Evolutionary models as dynamic systems can exhibit complexity in the guise of lock-in, path dependence, multiple equilibria (Arthur [1]) or even chaotic behaviour (Nowak and Sigmund [29]). In addition, experimental learning of boundedly rational agents may introduce stochasticity to the model that is difficult to capture by theoretical means (Axelrod [4]).

As a result, much of the recent work on the RPD has adopted the computer simulation technique to simulate the behaviour of interacting and learning populations of boundedly rational agents. The computer permits analysis in the presence of both high model and high dynamic complexity. Moreover, computer simulations are ideally suited to monitor complex processes as they unfold. The pioneering work in this area is due to Axelrod [3], who was the first to use the simulation technique to demonstrate the individual viability, the spread and the subsequent stability of conditionally cooperative behaviour among populations of players. However, his seminal results are in the process of re-examination with the realisation that the inductive nature of the computer simulation technique generates a crucial dependence of the results on the parametric assumptions of the model. In particular, seven key simulation parameters which have a strong impact on the outcome of the simulation have been identified (Axelrod and Dion [6], Hoffmann [17]): they are the composition of the initial population, the relative size of the RPD payoffs within the game's conditions, the number of rounds played per interaction, the presence of stochastic shocks or agent 'mistakes', the type of agent strategies, the definition of the agent learning dynamic and the structure of agent interaction and learning. While Axelrod's results remain unchallenged with respect to the specific settings for the parameters he employed, much research is being carried out to investigate the effect of varying the values of individual simulation parameters on the evolution of cooperation. This research aims to establish a comprehensive list of the factors that promote cooperative behaviour in the game.

Work to shed light on the effect of varying the more fundamental of these parameters has progressed steadily over recent years. The effect of varying the relative size of the

game’s payoffs within its conditions (Mueller [27], Nachbar [28], Hoffmann [17]) and of noisy strategy transmission (Hirshleifer and Martinez Coll [15], Mueller [27], Miller [26], Bendor, Kramer and Stout [8], Hoffmann [17]) is now well-understood. However, little is yet known about what impact the use of alternative population structures has on simulation results. The present study reports results from a series of simulations designed to investigate this matter systematically.

2 Population Structure

The process matching agents to play the RPD and learn from one another, or in Axelrod and Dion’s [6] words, the ‘population structure’ of the simulation, has a profound impact on the simulation results (Axelrod and Dion [6], Hirshleifer and Martinez Coll, Mueller [15], Sigmund [37]). However, most of the early contributions sidestepped the complex issue of examining the precise effect of population structure by employing either exhaustive or randomised matching processes that avoid asymmetric effects. Nonetheless, Axelrod acknowledged at the very outset of the RPD-simulation literature that much of the behaviour the literature is trying to capture involves biased matching processes (Axelrod [3]). In Axelrod’s view, agents mainly interact with and learn from certain subsets of the total population.

Recently, alternative and more realistic population structures of this sort have been explored in the literature. Vanberg and Congleton [42] as well as Stanley, Ashlock and Tesfatsion [40] investigated games where players have an option to exit once selected to play. However, this violates the conditions of the game as defined by Axelrod [3] and arguably defeats its paradoxical character. Axelrod [3] and Holland [18] consider agents which carry social labels or tags on the basis of which they are matched to play the game. This area of research is promising but yet at an early stage. A third strand of work is concerned with introducing a territorial structure to interaction and learning (Axelrod [3], Nowak and May [30, 31], Ellison [12], Feldman and Nagel [13], Routledge [35], Kirchkamp [20]). The motivation for this approach is that much of the interaction and learning that computer simulations aim to model takes place within ‘neighbourhoods’ of players. For instance, firms tend to interact with and learn from other firms either in geographical or in more abstract proximity (for example in terms of vertical or horizontal integration). In computer simulations, this kind of interaction can be captured by placing agents into territorial structures and restricting them to interact and learn within geographical neighbourhoods. For interaction, a neighbourhood determines the subset of the simulation pool with which a player is matched to play the repeated game. Localising learning means restricting the subset of the pool from which players

can learn better-performing strategies. Nowak and May [30, 31] study a population of agents distributed on squares on a torus which are capable only of the Always Defect (AD) and Always Cooperate (AC) strategies. Each agent interacts with the agents on all eight adjacent squares and imitates the strategy of any better-performing one. Nowak and May find that the distribution of strategies on the torus depends on the relative size of the RPD-payoffs. Cooperative behaviour can be sustained in clusters of agents that insulate cooperators from hostile ADs under certain RPD payoffs.

Both Routledge [35] and Kirchkamp [20] consider localised interaction between agents on a torus employing finite automata to play the game. While these authors detect the effects of varying the payoffs, the learning assumption as well as the amount of noise in the system on simulation results, they do not compare global and local population structures.

Hence, despite the obvious promise of analysing the effect of the localisation of both learning and interaction, further research is clearly required. Specifically, two questions arise from the work reported to date. First, in order to ascertain the effect of localising interaction and learning, it is necessary to compare the conventional global model with models incorporating a localised population structure. In the literature to date, little comparison of this sort is made. The current work is intended to address this issue. The comparison is most suitably made by investigating the localisation as a variable of the model. In this sense, the current work reports on the results of increasingly localising interaction and learning starting with global population structures.

Secondly, both Axelrod and the subsequent RPD-literature treat the neighbourhoods for interaction and learning as identical. The argument is that restricting players to interact within a neighbourhood also impairs their ability to observe and thus learn on a global scale. However, it is argued here that the localisation of learning and interaction are two separate topics which deserve individual attention. In many instances it is not clear why the localisation of both learning and interaction should coincide. It is easy to imagine situations where individuals interact locally while being able to observe what individuals outside their interaction-neighbourhood are doing. For example, while the inhabitants of a small village may rarely have interactions outside it, they may nevertheless be able to observe the behaviour of outsiders via access to the mass media. Conversely, individuals in large urban areas may be able to interact with a larger proportion of their environment, but cognitive constraints may prevent them from observing and remembering the behaviour of everyone they have encountered. As a result, in the following, a model will be presented in which the localisations of interaction and learning are separate parameters.

3 The Model

In order to investigate these issues, the following model was employed. Following recent literature on bounded rationality (Aumann [2], Rubinstein [36], Binmore and Samuelson [9]), the agents in this model select *Moore machines* to play the RPD on their behalf. A Moore machine is a finite automaton which generates an output contingent upon its own current state and its current input. Finite automata such as Moore machines are well-suited to represent the game strategies of boundedly-rational players. For the RPD, the possible states, inputs and outputs of the machine are elements of the set $\{d, c\}$. The inputs into the machine are the moves of the opponents, and the output the machine's own moves. The automaton's states map its input onto its output. Moore machines can have any finite number of states. One-state Moore machines have either a d-state or a c-state and are equivalent to AD- and AC-automata respectively. Two-state Moore machines have one d-state and one c-state. These machines generate moves on the basis of the outcome of the previous round of play in a game. There are thus twenty-six possible two-state Moore machines, a complete list of which is provided by Binmore and Samuelson [9] as well as Linster [22]. The present work examines the effect of independently localising agent interaction and learning with agents using both one- and two-state Moore machines. Two-state Moore machines can exhibit behaviour considerably more complex than that of their more forgetful colleagues and hence the effect of the localisations may differ. Moreover, the use of these two conventional machine types allows comparison with previous work on both automata types.

In the current model, populations of n such agents are located on the periphery of one-dimensional rings. The advantage of rings (or tori, the two-dimensional version of the ring structure) compared with other spatial models such as chessboards is that boundary effects, where players have an unequal number of neighbours, are avoided. The players on the ring are restricted to interact with their immediate i neighbours on either side of the ring. For example, with an interaction neighbourhood of size $i = 2$, each player would interact with one player from either side. Further, players are assigned to learning neighbourhoods of size l , which define the number of neighbours learning agents can emulate. Hence the population of players on the ring is structured in a web of overlapping interaction as well as learning neighbourhoods. The increasing localisation of interaction and learning is analysed with simulations using alternative values for l and i respectively subject to $2 \leq i \leq n - 1$ and $2 \leq l \leq n - 1$ and otherwise identical parametrisations. Series of simulations were carried out sampling possible combinations of i and l .

Learning in this model is implemented using a Genetic Algorithm (GA). The GA is a population-based heuristic search technique. Its popularity derives from its proven

robustness and success at finding near-optimal solutions to a wide range of problems involving complex search spaces. Potential solutions in the GA are encoded as strings of binary digits analogous in function to the chromosome of biological systems. An initial randomly generated population of bit-strings are manipulated by the GA using idealised genetic operators. The first, *crossover*, allows for recombination of parent bit-strings in a process functionally analogous to sexual reproduction; the second, *mutation*, causes random perturbations of individual digits in the resulting offspring strings. Evolutionary pressures are applied to the population by only allowing the most ‘fit’ individuals to mate, and using the resulting offspring to replace weak population members. Fitness in this context refers to the efficacy of the individual solutions when applied to a given problem. By promoting the propagation of useful genetic material at the expense of less useful material, the population eventually converges on near-optimal solutions.

The GA’s use of evolutionary processes in optimisation is a popular means of modelling learning under bounded rationality (see Axelrod [5], Marks [24], Miller [26], Routledge [35]). Learning boundedly rational agents, much like the GA, display optimising behaviour on the aggregate despite their individual myopia. Each individual agent merely seeks to improve their performance by selecting amongst alternative decision-making rules using imitation and trial-and-error learning. Individual rules are thus subject to artificial evolutionary processes generated by agents’ selection between them. Thus, on the aggregate, successful rules can spread amongst pools of players much in the same way as good solutions are propagated by the GA.

The GA is well-suited for incorporation into the simulation of interacting boundedly rational agents. Agent strategies in the guise of binary strings representing Moore machines can be used as the possible GA-solutions. Finite automata lend themselves to the binary representation in the computer program and thus can be conveniently subjected to the GA’s artificial evolution (see Hoffmann [17] for an exploration of finite automata and their genetic representation as well as manipulation). The evaluation of the automata’s fitness consists of their average payoffs generated by their interaction with other machines.

We depart from the conventional approach to GA implementation in three ways. First, in the canonical GA there is no spatial dimension which allows the formation of a relationship between a strategy and its location. Here we adopt a cellular implementation of a GA (see Kapsalis *et al.* [19]) to enable the imposition of spatial properties on the model. One consequence of this approach can be seen in the type of mating process employed. Traditionally, the selection of partners for mating is unrestricted. Here selection is limited to individuals belonging to the same learning neighbourhood. Next, the traditional approach to integrating new members into the population has been abandoned in favour of one more suited to the cellular implementation of a GA.

Of the two offspring produced each mating, only one, selected randomly, is allowed to survive. This is used to replace the strategy of the first parent automaton. The second parent's strategy remains intact. The first parent strategy cannot be selected again during the same learning period; the second can while it remains unmodified. The final departure from the GA as conventionally used is in the way fitness scores are derived. Traditionally, the GA evaluates alternative solutions to a static problem. Here fitness is a measure of a strategy's performance in a mutually adapting environment of competing players. This form of evolutionary adaption is called *bootstrapping* or *coevolution* (Axelrod [5]).

Each simulation thus proceeds as follows. The n players on the ring are allocated a random automaton of the type concerned. Each agent then interacts with all their i neighbours and obtains an average payoff accordingly. After the interactions have been completed, half the population are chosen at random for a strategy update. This update involves selecting an agent from the learning neighbourhood of the player concerned using a stochastic process that discriminates from a probability distribution favouring the better neighbours. Crossover using a randomly-selected single crossover point occurs with probability 0.6, otherwise the learning player's automaton remains intact. If crossover occurs, one resulting offspring is chosen at random to replace the learning player's strategy. Subsequently, the player's strategy is subjected to a mutation rate of 0.001. After this process is repeated for all $n/2$ learning players, the next generation of the simulation commences and proceeds as the first.

4 The Results

The work reported here explores the effect of independently localising the interaction and learning of populations of boundedly rational agents playing the RPD. The agents are distributed on rings and their strategies are represented by Moore machines. Agent co-adaptation is modelled by the GA. Two series of simulations of this model were carried out. In series one, twenty-one players were placed on the ring and allocated one-state Moore machines at random. The parameters i and l were varied between two and twenty at intervals of two with otherwise identical parametrisations. For each i/l -coordinate, one hundred simulations were carried out and the results aggregated respectively. One round per game was played. In series two, thirty-one players were placed on a ring and randomly allocated two-state Moore machines. The variables i and l were independently varied between two and thirty at intervals of four. For each localisation coordinate, twenty five simulations of five hundred generations were run and the results aggregated. Fifty rounds per game were played. For both series, the

payoffs used throughout were $T = 1$, $R = 0.6$, $P = 0.2$ and $S = 0$.

The GA generates an evolutionary learning process which manipulates the strategic composition of the population on the basis of individual strategies' fitnesses. A strategy's fitness in turn depends on its adapting strategic environment. The resultant dynamic system thus converges when the individual fitnesses in the population are equal. If the convergence is stable, the system has reached an attractor. If multiple attractors exist, the selection between them depends on the parametrisation of the system which defines its initial conditions. Subsequently, the mutation operator of the GA causes temporary or permanent deviations from the attractor(s). In the following, we report results from the observation of such dynamical systems under alternative localisations. The analysis will focus on the issue of how the selection of the cooperative attractor is affected by the change of localisation parameter. The next two sections address this issue for one-state and two-state machines respectively.

4.1 One-state Moore Machines

The systems comprising one-state automata displayed convergence on two attractors, mutual defection and mutual cooperation respectively. While all systems generally converged towards either of the two, temporary departures from as well as periodic shifts between them were in evidence. The selection between the attractors was strongly influenced by the two localisations. In general, both localisations promoted cooperative behaviour in the populations. This effect can be analysed using data on the frequency distributions of all rounds played in the simulations over the RPD's four outcomes dd, dc, cd and cc⁴.

The simulations displayed changes in behaviour in terms of these frequencies under varying values of i and l (see figures 2(a) to 2(c)). Most of the interactions of the observed simulation were characterised by mutual defection (figure 2(a)). However, high levels of localisation generated some amount of mutual cooperation (figure 2(c)).

The diagrams suggest that mutual cooperation was aided by the localisations. For $i = 2$, $l = 2$, 67% of rounds were characterised by mutual cooperation. About 20% of rounds played with $i = 2$, $l = 4$ were cc. As the low ridge in figure 2(c) suggests, $l = 2$ generated a small amount (about 1%) of mutually cooperative behaviour for all values of i . All other simulations displayed no significant amount of mutual cooperation. This result corresponds to Nowak and May's [30, 31] finding. However, we can confirm that only extreme levels of both localisations can produce this interesting effect.

The theoretical explanation of this result is that under the localisations, cooperative

⁴Since the incidence of dc and cd in a simulation must be identical, we report only the former.

players can persist in clusters which afford some degree of insulation from defecting players elsewhere. The localisation of learning generates monomorphic strategy clusters on the ring since the imitative learning implicit in the GA propagates fit strategies locally. Since cooperative players benefit from playing fellow cooperators while defecting players suffer in the presence of defectors, cooperation is aided by the localisations.

This phenomenon can be explained by example. Consider the section of a hypothetical ring depicted in figure 3. The column of digits should be imagined to fold back into itself. AD-players are represented by zeros, and AC-players by ones. Assume

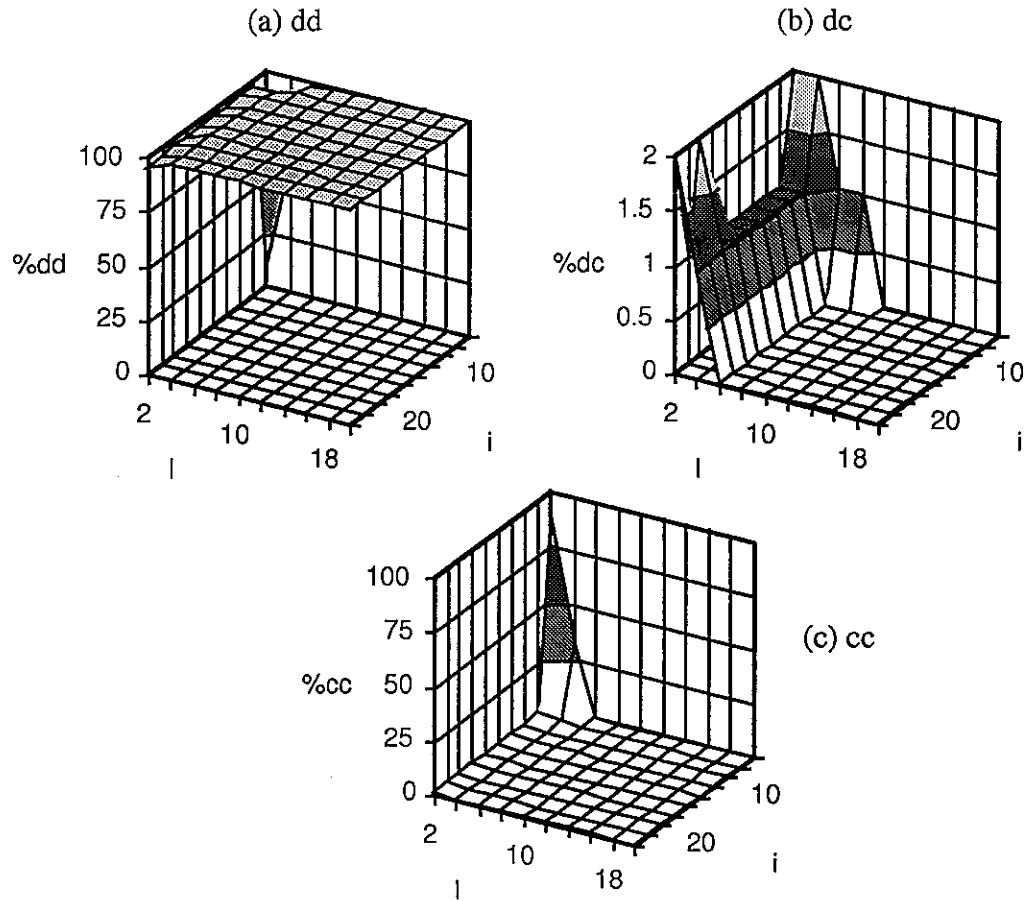


Figure 2: The relative frequencies of the RPD outcomes for one-state machines for (a) dd (b) dc (c) cc.

$i = l = 2$. The average score per generation for an AC-player in the middle of the cluster is $2R$, the border-ACs score $R + S$, the border-AD players obtain $T + P$, while the ADs in the middle of the cluster get $2P$. Changes in the cluster sizes can only arise by border individuals adopting a new strategy. Assuming deterministic imitation of the best performing neighbour, the AC-player on the border would swap to AD only if $T + P > 2R$. The AD-player on the border adopts the AC-strategy if $2P > R + S$. Under the present payoffs, the AC-player on the border is equally likely to learn from either neighbouring player, while the border AD-agent adopts the AC-automaton. In general, the spread of the strategies depends on the definition of the learning dynamic, on the relative size of the RPD payoffs and on the degree of localisation.

$$\begin{array}{c}
 1 \\
 1 \\
 1 \\
 0 \\
 0 \\
 0 \\
 0 \\
 1 \\
 1 \\
 1
 \end{array}
 \left. \begin{array}{l}
 \\
 \\
 \\
 \\
 \\
 \\
 \\
 \\
 \\
 \end{array} \right\}
 \begin{array}{l}
 2R \\
 R+S \\
 T+P \\
 T+P \\
 R+S \\
 2R
 \end{array}$$

Figure 3: Payoffs on a hypothetical ring

Figures 4(a) and 4(b) contain snapshots from the simulation of the interaction on a ring of size $n = 10$ with $i = l = 2$. The figures indicate how the two respective equilibria become established in practice. Each vertical column in the figure represents the population on the ring in a given generation. The horizontal axis shows the strategy histories of individual players over the entire simulation. The light/dark areas represent the spread of the AC-/AD-automaton respectively. The intermediate areas indicate cooperators on the border of defecting clusters. The postulated cluster formation on the ring is apparent in both figures. The evolution of cooperation and defection can be seen in the growth and decay of strategy clusters.

Apart from affecting attractor selection, the localisations had a second impact on the behaviour of the simulations. This concerns the stability of the attractor reached. Figure 2(b) reveals that the level of disequilibrium in terms of the frequency of the asymmetric outcomes was positively affected by the localisation of learning. The in-

cidence of dc constitutes departures from either of the two attractors. However, it is difficult to define unambiguous measures of the level and amount of disequilibrium in a system. Specifically, the dc-frequency hides whether the disequilibrium constitutes a temporary departure from one attractor or the convergence towards another. Further work in this direction is forthcoming.

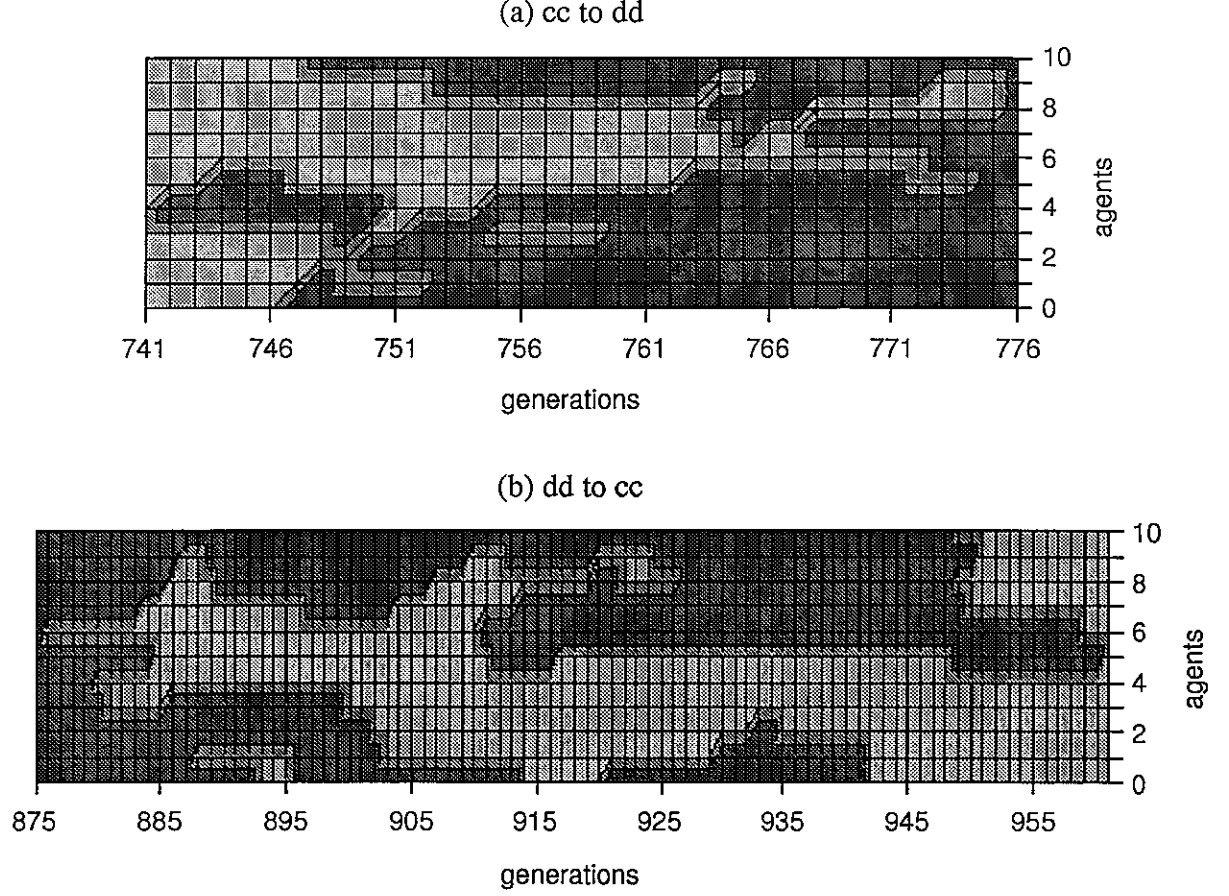


Figure 4: Snapshots of the history of a ring, size $n = 10$, of one-state machines showing (a) shift from cc to dd (b) shift back again to cc.

4.2 Two-state Moore Machines

The effects of the two localisations observed on systems comprising one-state machines were also evident in systems with the more complex two-state machines. Figures 5(a) to 5(d) display four typical simulation histories associated with extreme values for i and l . In the figures, the population average payoffs are plotted against the number of generations. The average payoff path for the population as a whole reveals much about the behaviour in the simulation. In general, it is evident that the systems spend most of the time at, or close to, two attractors, dd and cc (associated with the payoffs $R = 0.6$ and $P = 0.2$) respectively.

However, the systems displayed a greater amount of cooperative behaviour than the systems comprising one-state automata. The explanation for this phenomenon is that two-state Moore machines are capable of the type of conditionally cooperative behaviour that Axelrod [3] identified to be responsible for the evolution of cooperation. Close inspection of the emergence of cooperation in the present systems has revealed the following. In general, mutual cooperation in the systems is generated by the emergence of conditionally cooperative strategies such as TRIGGER⁵ especially and to a lesser extent TIT-FOR-TAT⁶ (TFT). These strategies discriminate between cooperative and defecting strategies and reciprocate their behaviour. Since these strategies score high payoffs against other cooperators while resisting exploitation by defecting strategies, they can score high payoffs in variegated strategy environments. As these strategies spread, the systems eventually converge on cc.

Figures 5(a) to 5(d) also reveal that shifts between the two attractors occur periodically. These are caused by the phenomenon of ‘genetic drift’, which arises through the GA’s mutation operator. Drift can allow a system to migrate between two basins of attraction in the following way. For example, once the cooperative attractor has been reached by the spread of conditionally cooperative strategies, the equilibrium is subsequently supported by a phenotypically identical polymorphism of nice strategies⁷. The equal fitness of the polymorphism allows the population to drift towards the more exploitable members of the nice class, such as AC, until the circumstances become viable for defecting strategies, such as AD, to re-emerge through mutation. The arrival and subsequent spread of defectors can then shift the system towards the attractor of dd. In convergence, a polymorphous mix of nasty⁸ strategies may drift the population

⁵The strategy that cooperates until a single defection of its opponent triggers its own continuous defection.

⁶The strategy that repeats its opponent’s previous move.

⁷In Axelrod’s term a strategy that never defects first.

⁸Continuing Axelrod’s terminology, we define nasty strategies as strategies which never cooperate

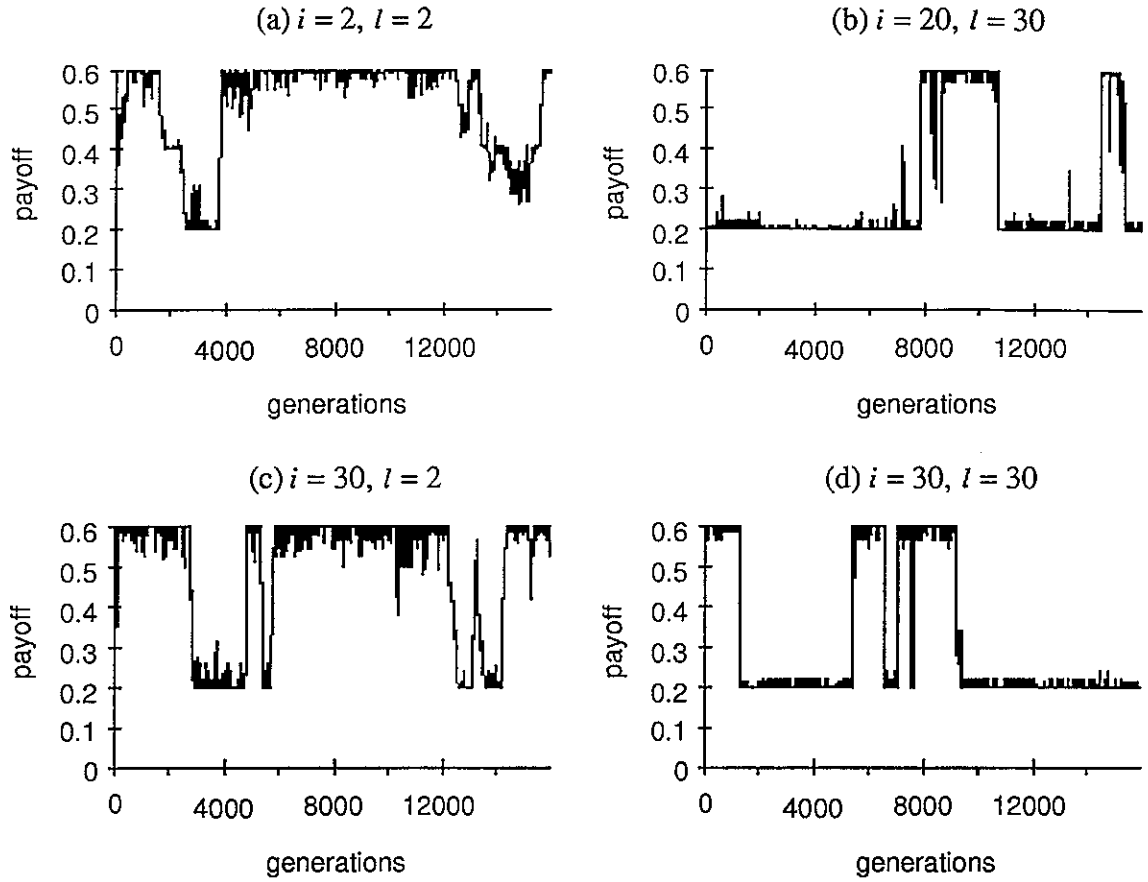


Figure 5: Four simulation histories with extreme values for i and l . (a) $i = 2, l = 2$; (b) $i = 20, l = 30$; (c) $i = 30, l = 2$; and (d) $i = 30, l = 30$.

towards nasty potential cooperators such as NASTY TRIGGER⁹, which again make the evolution of cooperation with the arrival of individual conditional cooperators feasible. Specifically, TRIGGER can exploit NASTY TRIGGER on every round of the game but the first. Thus, the system can oscillate between the two attractors.

The system's propensity to move between the two attractors in this way was again influenced by the variables i and l . Figures 6(a) to 6(c) contain the average relative first.

⁹The strategy that defects until a single cooperation of its opposition triggers its own continuous cooperation.

frequency distributions of the four outcomes against i and l for all the simulations. The inspection of figure 6(c) reveals the following about the relationship between mutual cooperation and the two localisations. Again, the highest level of mutual cooperation is associated with the lowest values for i and l . The localisation of learning has a positive effect on the frequency of cc for all values of i . However, the localisation of interaction seems to have an ambiguous effect on mutual cooperation. For low l , falling i promotes cooperative behaviour. For high l , the opposite is true. For $l = 2$, i seems to have an strong inverse relationship with cc, evident in the steep curve in the figure. As l rises, the curve becomes flatter until it reverses its slope. For $l = 30$, the positive relationship between i and cc becomes apparent. This phenomenon may point to a dynamic relationship between the two variables.

In order to gain some amount of understanding of this phenomenon, consider the simplified combinations of the two neighbourhood size parameters represented by the four areas depicted in figure 7. Although the complex nature of systems conventionally analysed by computational means renders the explanation of results difficult, some amount of speculation may prove insightful. Area B on the diagram represents global learning and interaction as conventionally studied in the literature. All agents obtain payoffs based on their interaction with all others, and can imitate any other agent in the population. In analysing the increasing localisation of both interaction and learning, we have to assess the effect of the three possible departures from global interaction and learning represented by the areas A , C and D .

Area D in the diagram involves systems in which learning remains global while interaction becomes localised. The interaction neighbourhood in evolutionary simulations constitutes the environment within which the fitness of an agent is established. Under the localisation of interaction, agent fitness thus becomes dependent on the composition of smaller neighbourhoods. Depending on the neighbourhoods, even globally poor strategies can thus prosper on the local level. The simulation results indicate that area D is characterised by the lowest incidence of mutual cooperation of the four areas. There are two factors that may contribute towards this effect. First, area D gives rise to the possibility that players imitate strategies which are wholly inappropriate in their own interaction neighbourhood. Although players can observe well-performing strategies elsewhere, the success of these may be limited to their respective interaction neighbourhoods. This effect could harm cooperative players more than defecting ones, maybe due to the robustness of defecting strategies. Furthermore, the localisation of interaction can cause genetic drift before convergence if clusters exist. For example, AC-players may spread by drift in a cluster of TRIGGER-players since the localised interaction prevents the payoffs between them to differ. In this sense, drift may then affect cooperative strategies more than defecting ones. However these observations

remain speculative.

The departure from area B to area A involves the localisation of learning with global interaction. In an evolutionary sense, the learning neighbourhood generates the environment within which a given player's strategy can spread by being imitated. Thus, localised learning gives rise to the formation of monomorphic strategy clusters by local imitation. The localisation of learning makes it more difficult for globally fit genetic material to disseminate in the population. Although the global interaction of area D allows global fitnesses to be attributed to all players, the learning localisation makes

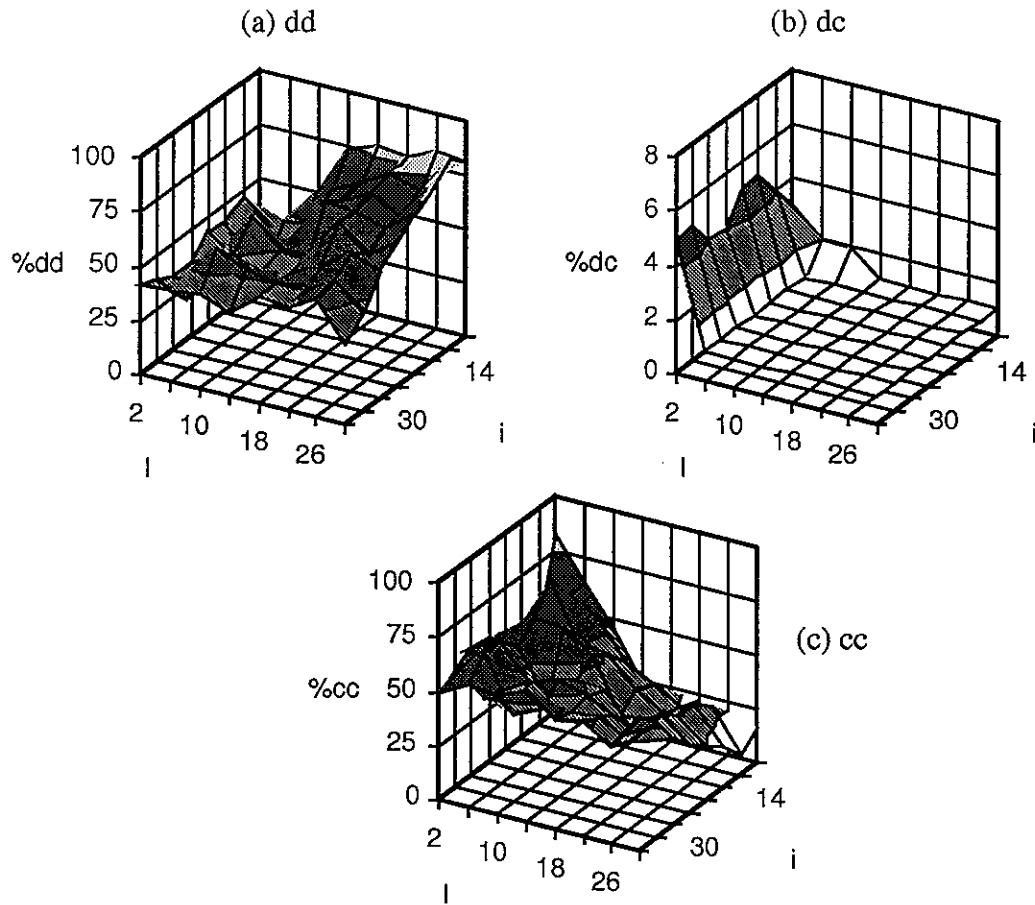


Figure 6: The relative frequencies of the RPD outcomes for two-state machines for (a) dd (b) dc (c) cc.

it less likely that this player can be observed by a given individual. In a sense, the localisation renders the learning procedure of each player less efficient by reducing the sampling of other players' strategies. Players are only able to learn from the most successful player in their neighbourhood.

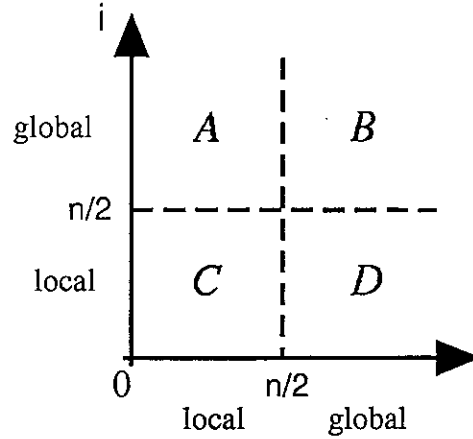


Figure 7: Simplified combinations of i and l .

Area A in the diagram is only second to area C in promoting cooperative behaviour. The localisation of learning generates the formation of clusters which aids cooperative behaviour and harms defecting players. Cooperators score high payoffs against each other while defectors perform poorly against their own kind.

The evolution of cooperation is promoted most when both interaction and learning are localised (area C). Cooperative players are isolated from defectors in clusters, and obtain high local payoffs which ensure individuals on the border to defecting clusters learn from other cooperators. Under these conditions, the localisation of interaction has a positive effect on cooperation which could be produced by a dynamic effect between the two variables. For example, the learning of inappropriate strategies with low i and high l becomes offset with falling l . This may neutralise the harmful effect of low i on cooperation.

Apart from the issue of attractor selection, the localisation of learning seemed to affect the incidence of disequilibrium behaviour. Again, the proportion of asymmetric outcomes in the simulations were strongly inversely related to l (figure 6(b)). Small learning neighbourhoods caused large amounts of dc-behaviour. Moreover, figures 5(a) to 5(d) confirm this finding. The average payoff paths of the populations look fuzzier

for small values of l . More research into this phenomenon is forthcoming.

5 Summary

The localisations of interaction as well as learning have a marked effect on the evolution of cooperation among agents using both one- and two-state machines. As far as systems with one-state machines are concerned, only very high levels of both types of localisation can generate some amount of mutually cooperative behaviour amongst agents. A large amount of insulation is required to protect AC-automata from defecting ones. Moreover, it seems that the localisation of learning has a more pronounced impact than the localisation of interaction. Learning rather than interacting in neighbourhoods generates the clusters protecting cooperation.

For two-state machines, the respective impacts of the two localisations are as follows. The localisation of learning again fosters cooperative behaviour while localised interaction has an ambiguous effect on it. Localised interaction promotes cooperation under localised learning, but harms it under global learning. In the presence of strategy clusters associated with small l , localised interaction promotes the growth of cooperative clusters by generating high payoffs for insulated cooperators. When no clusters are present with large l , the local payoffs generated by the localisation of interaction seem to harm cooperative players. The largest amount of mutual cooperation is produced when both localisations are at their maximum.

Lastly, the localisation of learning appeared to promote disequilibrium behaviour in the form of asymmetric outcomes in systems of both one-state and two-state automata. Presumably, the generation and persistence of strategy clusters hampered the convergence of the systems on either of the two attractors.

References

- [1] W.B. Arthur. Self-reinforcing mechanisms in economics. In P.W. Anderson, K.J. Arrow, and D. Pines, editors, *The Economy as a Complex Adaptive System*. Addison-Wesley, 1988.
- [2] R.J. Aumann. Survey of repeated games. In *Essays in Game Theory and Mathematical Economics*. Bibliographisches Institut Mannheim, 1981.
- [3] R. Axelrod. *The Evolution of Cooperation*. Penguin Books, 1984.

- [4] R. Axelrod. An evolutionary approach to norms. *American Political Science Review*, 80(4), 1986.
- [5] R. Axelrod. The evolution of strategies in the iterated prisoner's dilemma. In L. Davis, editor, *Genetic Algorithms and Simulated Annealing*. Pitman, 1987.
- [6] R. Axelrod and D. Dion. The further evolution of cooperation. *Science*, 242, 1988.
- [7] R. Axelrod and W.D. Hamilton. The evolution of cooperation. *Science*, 211:379–403, 1980.
- [8] J. Bendor, R.M. Kramer, and S. Stout. When in doubt: Cooperation in a noisy prisoner's dilemma. *Journal of Conflict Resolution*, 35(4), 1991.
- [9] K.G. Binmore and L. Samuelson. Evolutionary stability in repeated games played by finite automata. *Journal of Economic Theory*, 57, 1992.
- [10] R. Boyd and J.P. Lorberbaum. No pure strategy is evolutionary stable in the repeated prisoner's dilemma game. *Nature*, 327, 1987.
- [11] R.M. Dawes and R.H. Thaler. Anomalies: Cooperation. *Journal of Economic Perspectives*, 2(3), 1988.
- [12] G. Ellison. Learning, local interaction, and coordination. *Econometrica*, 61(5), 1993.
- [13] B. Feldman and K. Nagel. Lattice games with strategic takeover. In L. Nadel and D. Stein, editors, *1992 Lectures in Complex Systems*. Addison-Wesley, 1993.
- [14] D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3), 1986.
- [15] J. Hirshleifer and J. Martinez Coll. What strategies can support the evolutionary emergence of cooperation? *Journal of Conflict Resolution*, 32(2), 1988.
- [16] G.M. Hodgson. *Economics and Evolution*. Polity Press, 1993.
- [17] J.R. Hoffmann. The evolution of cooperation revisited. Technical Report 9507, The Economics Research Centre, University of East Anglia, 1995.
- [18] J. Holland. The effect of labels (tags) on social interactions. Discussion paper 93-10-064, Santa Fe Institute, 1993.

- [19] A. Kapsalis, G.D. Smith, and V.J. Rayward-Smith. A unified paradigm for parallel genetic algorithms. In T.C. Fogarty, editor, *Lecture Notes in Computer Science*, volume 865, pages 131–149. Springer Verlag, 1994.
- [20] O. Kirchkamp. Spatial evolution of automata in the prisoners' dilemma. University of Bonn, 1995.
- [21] D. M. Kreps, P. Milgrom, J. Roberts, and R. Wilson. Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27, 1982.
- [22] B.G. Linster. Evolutionary stability in the infinitely repeated prisoner's dilemma played by two-state moore machines. *Southern Economic Journal*, 58, 1992.
- [23] G.J. Mailath. Introduction: Symposium on evolutionary game theory. *Journal of Economic Theory*, 57, 1992.
- [24] R.E. Marks. Breeding hybrid strategies: Optimal behaviour for oligopolists. working paper 89-006, Australian Graduate School of Business, 1989.
- [25] J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.
- [26] J.H. Miller. The coevolution of automata in the repeated prisoner's dilemma. Technical Report 89-003, Santa Fe Institute, 1989.
- [27] U. Mueller. Optimal retaliation for optimal cooperation. *Journal of Conflict Resolution*, 31(4), 1988.
- [28] J. H. Nachbar. Evolution in the finitely repeated prisoner's dilemma. *Journal of Economic Behavior and Organization*, 19, 1992.
- [29] M. Nowak and K. Sigmund. Oscillations in the evolution of reciprocity. *Journal of Theoretical Biology*, 137, 1989.
- [30] M.A. Nowak and R.M. May. The spatial dilemmas of evolution. *International Journal of Bifurcation and Chaos*, 3(1), 1992a.
- [31] M.A. Nowak and R.M. May. Evolutionary games and spatial chaos. *Nature*, 359, 1992b.
- [32] W. Poundstone. *The Prisoner's Dilemma*. Oxford University Press, 1993.

- [33] A Rapoport and C. Orwant. Experimental games: A review. *Behavioral Sciences*, 7, 1962.
- [34] A.E. Roth. Laboratory experimentation in economics: A methodological overview. *The Economic Journal*, 98, 1988.
- [35] B.R. Routledge. Co-evolution and spatial interaction. University of British Columbia, 1993.
- [36] A. Rubinstein. Finite automata in the repeated prisoner's dilemma. *Journal of Economic Theory*, 39, 1986.
- [37] K. Sigmund. *Games of Life*. Oxford University Press, 1993.
- [38] H. A. Simon. *Models of Man*. John Wiley & Sons, 1957.
- [39] H. A. Simon. *Models of Bounded Rationality*. MIT Press, 1982.
- [40] E.A. Stanley, D. Ashlock, and L. Tesfatsion. Iterated prisoner's dilemma with choice and refusal of partners. In C.G. Langton, editor, *Artificial Life III*. Santa Fe Institute, Addison-Wesley, 1994.
- [41] R. Trivers. *Social Evolution*. Benjamin/Cummings Publishing, 1985.
- [42] V.J. Vanberg and R.D. Congleton. Rationality, morality and exit. *American Political Science Review*, 86(2), 1992.