Analyse & Kritik 22/2000 (© Lucius & Lucius, Stuttgart) p. 75–97

Rainer Hegselmann/Andreas Flache

Rational and Adaptive Playing

A Comparative Analysis for All Possible Prisoner's Dilemmas[*]

Abstract: In this paper we compare *two micro foundations* for modelling human behaviour and decision making. We focus on perfect strategic rationality on the one hand and a simple reinforcement mechanism on the other hand. *Iterated prisoner's dilemmas* serve as the play ground for the comparison. The main lesson of our analysis is that in the space of all possible 2×2 PDs different micro foundations do matter. This suggests that researchers can not safely rely on the assumption that implementing simple models of decision making will yield the same results that may be obtained when more sophisticated decision rules are built into the agents.

0. Introduction

In this paper we compare *two micro foundations* for modelling human behaviour and decision making. *Playing rationally* in a game theoretical sense will be the first micro foundation, a particular type of *adaptive learning* will be the second one. *Iterated prisoner's dilemmas* serve as the play ground for the comparison. We will *not* be able to compare *all* types of adaptive learning and *all* rational solutions for iterated PDs. However, for certain types of rational and adaptive solutions we will show how they behave *in all possible PDs.* In the *first* section we propose an easily applicable method to represent the set of all possible 2×2 PDs. The *second* section uses this method to describe and visualise for all possible iterated PDs the conditions under which players who are rational in a game theoretical sense can attain cooperative solutions. Our central approach is to analyse the effect of payoff parameters on the 'shadow of the future' that is required to make conditionally cooperative behaviour individually rational. More technically, we analyse how the payoff parameters shape the minimum probability of continuation of the game that is required for Trigger- and Tit-for-Tat equilibria in the PD-supergame. In the *third* section, we use a stochastic learning model of a simple adaptive agent to compare the corresponding results with those derived from the preceding game theoretical analysis. The third section provides both analytical results

[*] Both authors contributed equally. The order of names is chosen randomly.

on equilibria of the learning process and computer simulations that allow for a quantitative comparison of the 'shadow of the future' that is needed for cooperation between learning actors and rational actors, respectively. Section 4, finally, summarises and discusses results.

1. A Geometrical Characterisation of All Possible PDs

The general structure of a classical 2×2 PD is described by the matrix:

	cooperation[C]		defection[D]	
cooperation[C]	R_1	R_2	S_1	T_2
defection [D]	T_1	S_2	P_1	P_2

Table 1: PD in strategic form

A game with this general structure is a Prisoner's Dilemma (PD) if the following condition is satisfied:

$$T_i > R_i > P_i > S_i \qquad i = 1, 2. \tag{1}$$

Some scholars additionally require

$$T_i + S_i < 2R_i \qquad i = 1, 2. \tag{2}$$

The latter requirement guarantees that in an iterated PD players who take turns in cooperating *unilaterally*, like in the sequence

$$\ldots \langle D, C\rangle, \langle C, D\rangle, \langle C, D\rangle, \langle D, C\rangle \ldots,$$

can *never* be better off than players who mutually cooperate all the time. The *set of all possible* PDs simply consists of all payoff combinations that meet the requirements (1) and/or (2). Drawing on Harris (1969), we describe in this section an easy way to represent that set geometrically.

As a starting point it should be noted that the payoffs T_i, R_i, P_i, S_i are values of a *Neumann-Morgenstern* utility function. Such a cardinal utility function is unique up to transformations of the type

$$U^*(x) = m \cdot U(x) + n \qquad m > 0. \tag{3}$$

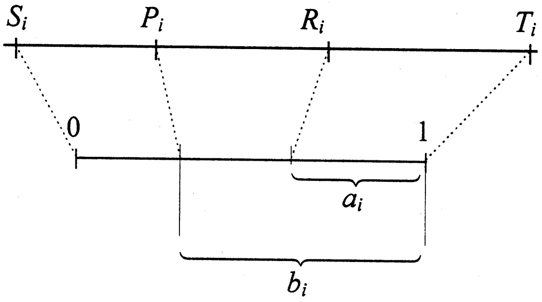


Figure 1: Normalised payoffs

Due to the uniqueness property we can always normalise PD payoffs such that we obtain $S_i = 0$ and $T_i = 1$. Figure 1 illustrates the normalisation procedure. Table 2 shows the new strategic form that we obtain after normalisation.

	cooperation[C]		defection[D]	
cooperation[C]	$1 - a_1$	$1 - a_2$	0	1
defection [D]	1	0	$1 - b_1$	$1 - b_2$

Table 2: PD in strategic form with normalised payoffs

With normalisation, the PD condition (1) turns into

$$1 > (1 - a_i) > (1 - b_i) > 0 \qquad i = 1, 2. \tag{4}$$

This requirement is equivalent to the two conditions given by (5).

$$\begin{array}{ll} 0 < a_i, b_i < 1 & \\ a_i < b_i & i = 1, 2. \end{array} \tag{5}$$

There is no agreement in the literature that Condition (2) is essential for a PD. However, for comparison note that with normalisation condition (2) turns into

$$a_i < 0.5 \tag{6}$$

The normalised PD condition (5) allows to characterise all possible *symmetric or asymmetric* PDs in terms of a pair of ordered pairs $\langle\langle a_1, b_1\rangle, \langle a_2, b_2\rangle\rangle$. In the case of *symmetric* PDs this reduces to only one pair $\langle a, b\rangle$ that fully describes the game.

Basically, the normalisation procedure 'reduces' four payoffs to two. The major advantage of the procedure is to allow for ***simple geometrical representations of all possible PDs as a special triangle of the unit square.*** Figure 2 shows the ensuing representation of the payoff space of all PDs.

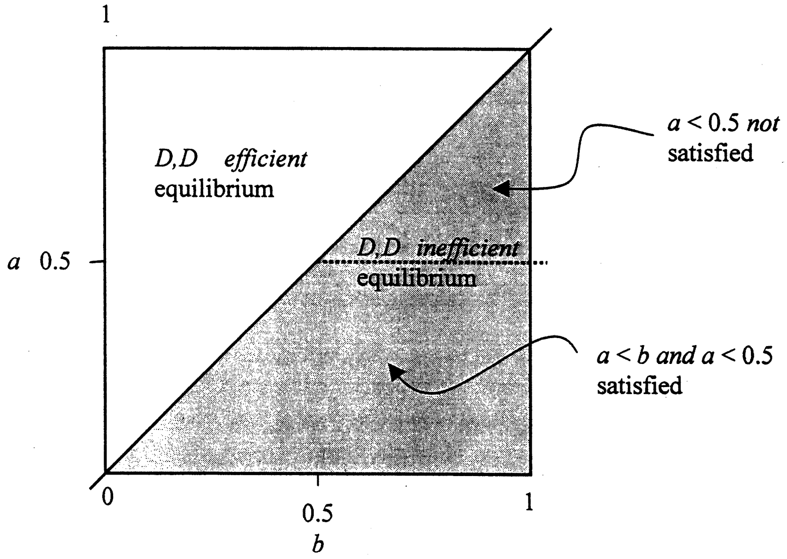


Figure 2: The PD unit square

The essential features of the unit square are:

- All points *within* the unit square satisfy the conditions $0 < a, b < 1$. Accordingly, these points represent games where a *dominant* strategy exists for both players. The inner unit square is the set of all symmetric 2×2 games with dominant strategies. The set of all ordered pairs of those points is the set of all symmetric *and asymmetric* games with dominant strategies for both players.

- The white triangle of the unit square (frame not included) consists of all points that satisfy $a \geq b$, which constitutes a *violation* of the essential PD condition (5) or (1), respectively. The diagonal *is part of* the white triangle. Hence, in this region we get games with dominant strategies, because the white triangle is part of the unit square. Playing dominant strategies *can* result in inefficient solutions, but this is *not necessarily* so. More precisely, all white points constitute the games for which mutual play of the dominant strategy is an *efficient equilibrium.*

- All points in the grey triangle (borders not included) are points which satisfy the essential PD condition $a < b$. Accordingly, those points constitute games in which mutual play of the dominant strategies is an *inefficient* solution. We refer to this area of the unit square as the *PD triangle.*

- In the upper part of the PD triangle, i.e. *above* the dotted line, we find all points that do *not* satisfy the second PD condition (6) or (2), respectively. The dotted line ($a = 0.5$) is part of that area.

In the following we will use the PD triangle and the unit square to compare conditions for cooperation in iterated PDs between rational players and adaptive players.

2. Trigger and Tit-For-Tat Supergame Equilibria for All Possible PDs

Axelrod's famous *The Evolution of Cooperation* (1984) made it widely known that under certain conditions perpetual cooperation can be sustained in iterated PDs, based on certain equilibria of supergame strategies. In other words: Players who are rational in a game theoretical sense can attain mutual ongoing cooperation, despite the fact that defection is the dominant strategy in the constituent PD. The key condition for this possibility of cooperation is that the probability α of continuation of the game for a further iteration exceeds a certain threshold value. The threshold condition depends on the supergame strategy and the payoffs involved. The PD triangle allows for straightforward visualisations of the corresponding threshold conditions over the set of all possible PDs. We start with an analysis of what is called the *Trigger strategy* (TR), i.e. the strategy that starts cooperatively but responds to the first defection of it's opponent with eternal own defection.

$\langle TR, TR \rangle$ is an equilibrium iff the (individual) probability for continuation of the game, α_i, and the payoffs of the game satisfy the following condition.[2]

$$\alpha_i \geq \frac{T_i - R_i}{T_i - P_i} = \alpha_i^* \qquad i = 1, 2 \tag{7}$$

After normalisation of all payoffs (7) turns into

$$\alpha_i \geq \frac{a_i}{b_i} = \alpha_i^* \qquad i = 1, 2 \tag{8}$$

The PD triangle provides an easy understanding of how equilibrium conditions behave within the class of all possible PDs. There is, *firstly*, a 2-dimensional way to represent (8). The underlying idea is to visualise all games, i.e. points

[2] For the (fairly simple) proof cf. Taylor 1976; 1987; Friedman 1977; 1986; Axelrod 1984.

in the unit square, that have the *same* threshold. In other words, we show all points $\langle b, a\rangle$ for which the ratio a/b is a constant. These points are located on a straight line with the slope a/b. The line is given by

$$y = \frac{a}{b}x \tag{9}$$

We refer to the lines defined by (9) as *iso-α^*-lines.* To visualise the *α^*-threshold* condition we draw the *iso-α^*-lines* across the PD-triangle. Figure 3 shows the result.

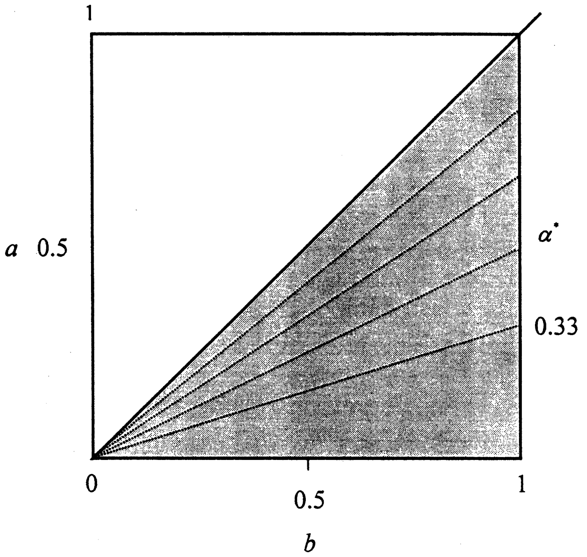


Figure 3: PD-triangle with iso-α^*-lines for TR strategies

In Figure 3, all dotted lines are *iso-α^*-lines.* The lines start next to the point $\langle 0,0\rangle$, because $a, b > 0$; and the lines end 'immediately' before they reach the right border of the PD triangle, because $a, b < 1$. Finally, the lines 'touch' the border at points with $b = 1$. For those points we have $\alpha^* = \frac{a}{b} = \frac{a}{1} = a$. Hence, we can interpret and use the right vertical border of the PD triangle as an axis indicating possible values for α^*. We refer to this right border as the α^*-axes. For all points in the PD-triangle, it holds that $0 < \alpha^* < 1$. All straight lines beginning next to the origin $\langle 0,0\rangle$ and touching the α^*-axes connect all possible points, i.e. PDs, with the *same* threshold value for α^*. The value where the *iso-α^*-line* touches the α^*-axes shows the level of α^* that applies for all points on the corresponding *iso-α^*-line.* All points/games on *and below* a certain *iso-α^*-line* can be solved cooperatively based on Trigger strategies if the probability α for another iteration to take place is at least α^*. For example: In Figure 3 all points on and below the line between the origin and the point 0.33 on the α^*-axes are PDs that can be solved cooperatively if the probability for another iteration is not less than 0.33.

Our analysis also shows that in principle *all* possible PDs can be solved cooperatively based on Trigger strategies—given that the probability α is high enough. The reason is that for a PD it holds by definition that $a < b$. As a consequence, the threshold condition (8) will always range between 0 and 1. This is a noteworthy and remarkable fact. In Figure 3 this result is reflected by the fact that for all points in the PD triangle we find an *iso-α^*-line* that touches the α^*-axes between 0 and 1.

There is, *secondly*, a *3-dimensional representation* of (8) for all possible PDs. The underlying idea is simple enough: we use the third dimension over the PD-triangle to visualise for every point in the triangle the corresponding threshold α^*. Figure 4 demonstrates the result. The figure shows that α^* becomes extremely high as a and b get closer to each other. Those points of the PD-triangle where the difference between a and b is small, represent games where mutual cooperation is only slightly more attractive as compared to mutual defection. Accordingly, in these games rational actors need to face a long 'shadow of the future', i.e. a large probability α of continuation of the game, for refraining from the temptation to defect.

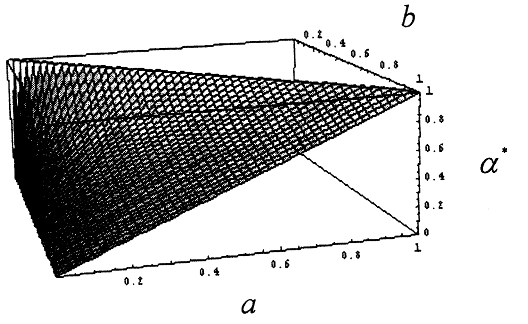


Figure 4: Trigger threshold α^* for all possible PDs.

The strategy combination Tit-for-Tat against Tit-for-Tat, $\langle TFT, TFT \rangle$, is in an equilibrium if condition (10) is satisfied.[3]

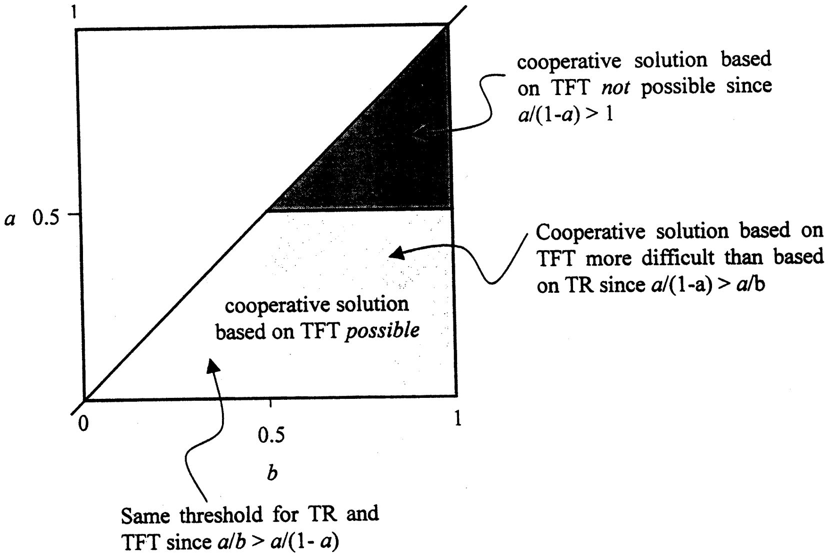
$$\alpha_i \geq max\{\frac{T_i - R_i}{T_i - P_i}, \frac{T_i - R_i}{R_i - S_i}\} = \alpha_i^{**} \qquad i = 1,2 \qquad (10)$$

After normalisation of all payoffs (10) turns into

$$\alpha_i \geq max\{\frac{a_i}{b_i}, \frac{a_i}{1 - a_i}\} = \alpha_i^{**} \qquad i = 1,2 \qquad (11)$$

[3] For a proof cf. Axelrod 1984, 207ff.; Taylor 1987, 60f.

The first element of (11) is the threshold condition for $\langle TR, TR\rangle$. We know (by definition) that this expression always is smaller than or equal to 1, because $a_i < b_i$ for every PD. However, the second threshold in (11) exceeds 1 if $a_i > 0,5$. That is the case for all points in the dark grey area of the PD triangle in Figure 5. Accordingly, all PDs in the dark grey area *can never be solved cooperatively using TFT-strategies*, because the probability of continuation can never be greater than 1. At the same time, conditional cooperation with TR strategies is still possible in this region of the PD-triangle.



cooperative solution based on TFT *not* possible since *a*/(1-*a*) > 1

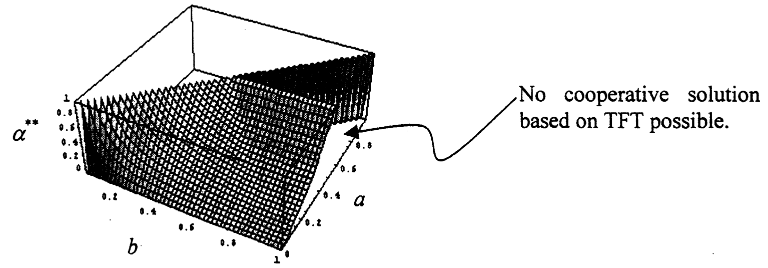
Cooperative solution based on TFT more difficult than based on TR since *a*/(1-a) > *a*/b

cooperative solution based on TFT *possible*

Same threshold for TR and TFT since *a*/*b* > *a*/(1- *a*)

Figure 5: PD triangle for TFT-strategies.

Figure 6 shows a 3-dimensional version of Figure 5. In addition to its two dimensional counterpart, the figure illustrates that the threshold value α^{**} dramatically increases in the area of the triangle in which $a/b < a(1-a)$. Moreover, for the one third of the PD triangle where $a(1-a) > 1$, there is not any possibility for a cooperative solution based on TFT strategies.



No cooperative solution based on TFT possible.

Figure 6: TFT threshold α^{**} for all possible PDs.

Our analysis has shown a straightforward geometrical representation of how threshold continuation probabilities for certain supergame equilibria are distributed over the PD triangle of all possible PDs. In the following, we use these results for a comparative analysis of rational and adaptive decision making in iterated PDs.

3. Equilibria of a Simple Learning Model for All Possible PDs

This section presents a simple model of learning behaviour in all iterated PDs and compares model results with the game theoretical analysis of the preceding section. Section 3.1 describes the model. Section 3.2 derives analytical conditions for equilibria of the learning process. Section 3.3, finally, presents computer simulations for quantitative comparisons of the conditions for cooperation generated by the two alternative micro foundations.

3.1 A Stochastic Learning Model

We apply a Bush-Mosteller stochastic learning model (Bush/Mosteller 1955) to formalise a simple reinforcement rule of learning by trial and error. In using this model we follow a line of work originating from the pioneering contribution of Rapoport and Chammah (1965). More recently, stochastic learning models have been applied in computer simulations of social dilemma behaviour by Macy (1989; 1991). Moreover, Roth and Erev (1995) used these models for prediction of experimental results. However, hitherto no systematic comparison is available of stochastic learning models with game theoretical models of rational behaviour in Prisoner's Dilemmas. In the following we provide this comparison for a particular stochastic learning model.

The model assumes that both players have an internal state, their propensity to cooperate, p, that changes over time on basis of the experiences players make. The cooperation propensity p varies between 0 and 1. In every iteration of the game both players simultaneously decide whether to cooperate or defect. In iteration t, player i cooperates with probability p_{ti} and defects with probability $1 - p_{ti}$.

Both players independently adapt their own cooperation propensity after decisions are taken. The adaptation of propensities reflects the trial and error mechanism. An actor's propensity to repeat his most recent decision increases when this decision was related to a satisfactory outcome. Conversely, the propensity of repetition declines, when the related outcome is deemed dissatisfactory. The degree of satisfaction, s, reflects the difference between an actor's expectation level e, and the payoff he attained, u. We assume that s varies between -1 and $+1$, where negative satisfaction expresses that the payoff falls below the expectation level. Positive satisfaction indicates that a payoff larger than the expectation level was attained. Equation (12) formalises the function that translates the payoff into the satisfaction level s.

$$s(u) = \begin{cases} \frac{u}{e} - 1, & \text{if} u \leq e \\ \\ \frac{u-e}{1-e}, & \text{if} u > e \end{cases} \tag{12}$$

In the normalised 2×2 PD, the payoff u varies between 0 and 1. Accordingly, we assume for (12) that $0 < e < 1$.

The satisfaction derived from the most recent payoff and the present propensity to cooperate, p, shape the change of the propensity, Δp . The magnitude of the 'raw' change in the propensity equals the satisfaction level. This change is multiplied with a term ensuring that the function flattens off when propensities approach the boundaries of the interval [0,1]. Finally, the change in the propensity is scaled with the parameter l, the learning rate $(l > 0)$. Table 3 shows how Δp is calculated.

		Satisfaction	
		$s(u) \geq 0$	$s(u) < 0$
Decision	C	$s(u)l(1-p)$	$s(u)lp$
taken	D	$-s(u)lp$	$-s(u)l(1-p)$

Table 3: Change in actor's propensity to cooperate, Δp, as function of own decision taken and satisfaction derived from the corresponding outcome, $s(u)$.

Table 3 indicates that a learning rate smaller than one corresponds to a relatively slow learning process. In this region of the parameter space, l dampens the effect of the satisfaction level on the change in propensity, as compared to a learning rate of *l=1*. By contrast, a learning rate larger than one entails a learning process that is faster than with *l=1*. At this level of l, the effect of satisfaction on the 'raw change' in propensity is amplified by the learning rate. However, this amplification makes it necessary to ensure that the learning dynamics still generate valid propensities. For this purpose, our model clips propensities at the boundaries of the interval [0,1]. Notice that this clipping rule is only needed for $l \geq 1$. Equation (13) formalises the clipping rule.

Equation (13) not only defines the clipping rule. The second purpose of (13) is to avoid the possibility that propensities never actually attain the extremes of *p=0* or *p=1*, even when propensities move ever closer to the extreme values. For this purpose, we assume that actors fully commit themselves to play a particular strategy in the subsequent iteration, if the difference between their propensity p and one of the two interval boundaries falls below a certain very small 'commitment threshold' ϵ. Hence, we assume for both players that the propensity to cooperate in iteration $t+1$, p_{t+1} , ensues from the propensity to cooperate in iteration t, p_t, as described by (13).

$$p_{t+1} = \begin{cases} p_t + l\Delta p_t, & \text{if } \epsilon < p_t + l\Delta p_t < 1 - \epsilon \\ 1, & \text{if } p_t + l\Delta p_t > 1 - \epsilon \\ 0, & \text{if } p_t + l\Delta p_t < \epsilon \end{cases} \tag{13}$$

3.2 Equilibria of the Learning Process: Analytical Results

To analyse the learning mechanism, we use the standard equilibrium concept of the theory of dynamic systems. Broadly, an equilibrium of the learning dynamics is a configuration of propensities in which both players show stable behaviour. More technically, we define an equilibrium as a configuration of propensities for which it is guaranteed that $\Delta p = 0$ for both players. Notice that this equilibrium concept is not to be confused with the Nash-equilibrium of the game theoretical analysis. There are equilibria of the learning dynamics that are not Nash-equilibria in the repeated game. There are also Nash-equilibria of the repeated game that are not equilibria of the learning process. We discuss examples further below.

Inspection of the learning dynamics allows to considerably narrow down the range of potential equilibria. A *necessary condition for an equilibrium* is that the probability of an *asymmetric outcome* (CD or DC) is zero. To explain, the learning dynamics imply that both players' propensities decline after an asymmetric outcome has occurred. The player who chose D attains a payoff *T=1*, yielding a satisfaction of $s(1) = 1$. Hence, his propensity to cooperate declines, because $l\Delta p = -lp$ (with exception of the extreme case *p=0*). Correspondingly, the player who chose C attains a satisfaction of $s(-1) = -1$, also resulting in $\Delta p = -lp$. Clearly, it is not guaranteed that $\Delta p = 0$, if the probability for an asymmetric outcome is larger than zero, because the asymmetric outcome can only arise in the first place, when at least one player's propensity was not zero before. This implies the following necessary condition for an equilibrium.

Necessary equilibrium condition. A configuration of propensities can only be in equilibrium if the probability for an asymmetric outcome (CD or DC) is zero. Hence, *the learning process is only in equilibrium if*

- *both* players' propensities to cooperate are zero, or
- *both* players' propensities to cooperate are one.

The necessary equilibrium condition immediately yields two equilibrium conditions that are both *necessary and sufficient.* These conditions correspond with the CC-equilibrium (p=*1* for both players) and the DD-equilibrium (p=*0* for both players), respectively.

CC-equilibrium condition. *Full cooperation*, i.e., *p=1 for both players*, is an equilibrium iff

$$a \leq (1-e). \tag{14}$$

DD-equilibrium condition. Full defection, i.e., *p=0 for both players*, is an equi- librium iff

$$b \leq (1-e) \tag{15}$$

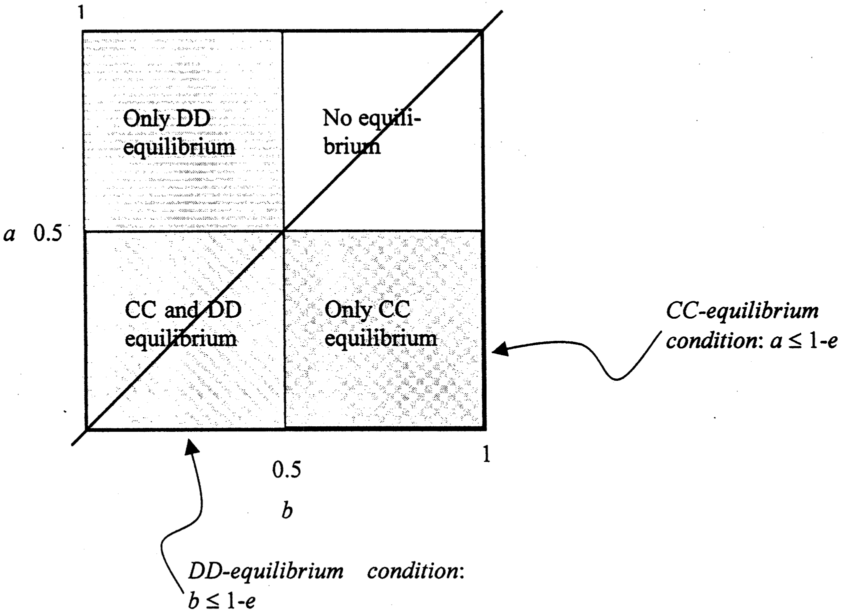


Figure 7: Unit-square representation of equilibrium conditions of the stochastic learning model ($e = 0.5$).

The CC-equilibrium condition and the DD-equilibrium condition ensue from the rules for the change of propensities in Table 3. To explain, in the CC-equilibrium both players always attain a payoff of u=*1-a*. Correspondingly, in the DD-equilibrium both players always attain a payoff of u=*1-b*. The definition of the satisfaction function, (12), ensures that the corresponding satisfaction levels are non-negative, iff $u \geq e$. This, in turn, is equivalent to the inequalities (14) and (15), respectively. The rules of Table 3 ensure that propensities never decline after a CC outcome, iff (14) is satisfied. Similarly, the rules of Table 3 ensure that propensities never increase after a DD outcome, iff (15) is satisfied. This implies the CC- and DD-equilibrium conditions, respectively.

The equilibrium conditions derived above allow to distinguish 4 regions of qualitatively different dynamics in the unit square. Figure 7 visualises the regions for an expectation level of *e=0.5*. Figure 7 demonstrates two ***qualitative differences between Nash-equilibrium conditions and the equilibrium conditions of the learning dynamics.***

- CC can be an equilibrium of the learning dynamics, even when it is an inefficient outcome. This occurs under the condition that both CC and DD yield satisfactory payoffs and the DD payoff exceeds the CC payoff ($a > b$). Notice that CC can never be a Nash-equilibrium under this condition.

- CC may *never* be an equilibrium of the learning dynamics, even when CC is an efficient outcome. This occurs under the condition that neither CC nor DD yield a satisfactory payoff and the CC payoff exceeds the DD payoff ($a < b$). Notice that under this condition there is always some continuation probability α that satisfies the Nash-equilibrium condition for Trigger strategies. However, cooperation between learning actors is impossible here, regardless of the continuation probability α.

3.3 Quantitative Comparison of Rational Actors and Learning Actors

The preceding analysis highlights qualitative discrepancies between the results of rational behaviour and learning behaviour. In the following, we turn to a quantitative comparison. As a starting point, we show in 3.3.1 that the existence of equilibria of the learning model implies that the learning process will converge on an equilibrium in a finite number of iterations. In 3.3.2 we use computer simulations to estimate the continuation probabilities required for attaining equilibria of the learning process. In 3.3.3, finally, we address effects of a central parameter of the learning model, the learning rate l, in order to assess the generalisability of results.

3.3.1 Convergence on the Equilibria

The CC- and DD-equilibrium conditions guarantee that there is a positive probability that the dynamics of the model end up in the corresponding equilibrium, if the game is played sufficiently long and both players start with a propensity to cooperate *between* zero and one. For the CC-equilibrium condition the proof can be sketched as follows. There is a certain probability that the outcome of a particular iteration t will be CC, as long as both players' propensities to cooperate are larger than zero. The CC-equilibrium condition implies that then both players attain positive satisfaction. Hence, both players increase their propensities to cooperate according to the rules of Table 3. This increases the probability that the outcome of the next iteration is CC again. If the outcome is in fact CC, this results in a further increase of the propensities, etc. After a sufficiently large number of consecutive occurrences of CC, both players' propensities exceed the threshold level $1 - \epsilon$.[4] At this point, both propensities are 'locked in' to full commitment to cooperation and the CC-equilibrium is attained. In terms of markov chain theory, bilateral full cooperation is an absorbing state of the stochastic learning process and this state is reachable from every other combination of players' propensities (with exception of $p=0$ for both players). For the DD-equilibrium a similar reasoning applies. There is always a positive probability that the outcome of a particular iteration t is an asymmetric outcome (CD or DC), as long as no equilibrium has been attained. After an asymmetric outcome both players' propensities decline, so that there is now certainly a positive probability for a subsequent DD-outcome. The DD-equilibrium condition guarantees that propensities to cooperate decline after every occurrence of DD. Hence, after a sufficient number of consecutive occurrences of DD the dynamics stabilise on the DD-equilibrium, if the corresponding DD-equilibrium condition is satisfied.

3.3.2 Quantitative Comparison of Conditions for Cooperation

In the game-theoretical analysis we derived a measure of the restrictiveness of conditions for cooperation in terms of the threshold continuation probability α that is required to make conditional cooperation individually rational. That analysis is deterministic in the sense that mutual cooperation is always individually rational, iff the continuation probability exceeds the critical threshold α^*. Unfortunately, it is less straightforward to obtain such a threshold for the learning model. The reason is that the learning model generates a stochastic process that may sometimes converge on equilibrium before the game ends and sometimes may fail to converge—under the same level of α. To attain at least an indication for the continuation probability required, we employed computer simulation of the stochastic learning process. With the simulation analysis, we estimate the threshold α that is required to ensure that learning actors attain mutual cooperation *with a certain probability q*. We use the

[4] More precisely, it is necessary to prove here that for every small number ϵ there is a number n of consecutive CC outcomes after which players' propensities exceed $1 - \epsilon$. The proof is straightforward and can be obtained from the authors on request.

symbol α^q to denote the threshold level of α corresponding to a probability of mutual cooperation of q.

The method to estimate α^q departs from a simulation of the distribution of the number of iterations that is required until a game attains lock-in on CC. We denote this number t^{CC}. To estimate t^{CC} for a particular parameter combination, we use 1.000 replications of the learning dynamics and we run every simulation for maximally 10.000 iterations. For illustration, Figure 8 shows the distribution that we obtain for the particular combination of *a=0.25* and *b=0.75*. Notice that this parameter combination represents a case where only the CC-equilibrium condition is satisfied. Furthermore, we used *l=1*, *e=0.5* and ϵ = 0.001. Finally, we assumed that both actors start with a propensity to cooperate of *p=0.5*.

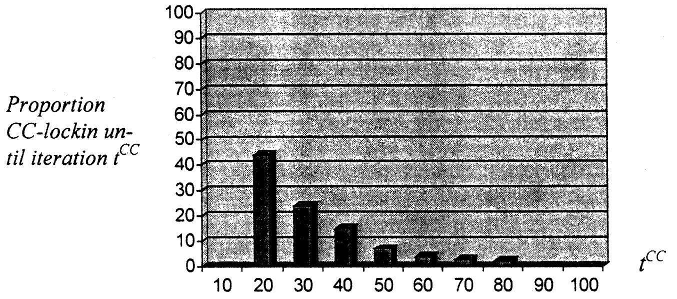


Figure 8: Simulated distribution of number iterations t^{CC} required for CC-lock-in (a=0.25, b=0.75, e=0.5, l=1, ϵ=0.001).

Figure 8 shows that the proportion of runs attaining CC-lock-in until iteration 10 is zero. Hence, at least 10 iterations are required before mutual cooperation can stabilise. About 44% of all runs attain CC-lock-in between iteration 10 and 20, another 24% settle on mutual cooperation between iteration 20 and 30, etc. We use statistics like this one to estimate α^q. For this purpose, we vary α between 0 and 1 in small steps, to estimate for every level of α the proportion of runs that attains CC-lock-in before the game ends. The smallest α for which this proportion is equal to or larger than q is the estimated value α^q. To estimate the proportion of CC-lock-in for a *particular* α, we sum over all iterations t the estimated probability that 1) lock-in occurs in this particular iteration *and* 2) the game will continue at least until this iteration. The latter event occurs with probability t^{α}, whereas the probability for event 1) is derived from the statistic computed by simulation (see Figure 8). The joint probability of both events is then obtained by multiplication of the separate probabilities, because the events 'lock-in in iteration t' and 'continuation of the game until iteration t' are statistically independent.

We used the method described above to estimate for all possible games in the unit square that satisfy the CC-equilibrium condition the indicators $\alpha^{0.5}$ and $\alpha^{0.9}$, i.e. the minimum continuation probabilities required for a chance of 50% and 90% of stabilisation on mutual cooperation before the game ends, respectively. In the simulations, we varied both a and b between 0 and 1 in steps of 0.02. Furthermore, we used 1.000 replications and maximally 10.000 iterations per replication. Again, we employ *l=1*, *e=0.5*, ϵ = 0.001 and an initial propensity to cooperate of *p=0.5* for both players. Figure 9 shows the results. Notice that the white regions in Figure 9 represent the parts of the unit square where the CC- equilibrium is not satisfied.

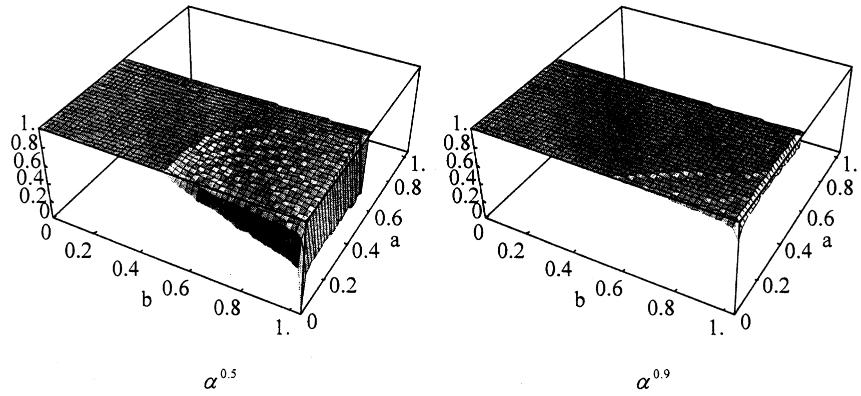


Figure 9: Estimated threshold continuation probability for 50% chance on CC-lock-in ($\alpha^{0.5}$) and 90% chance on CC-lock-in ($\alpha^{0.9}$). (Initial propensities p=0.5, l=1, e=0.5, ϵ=0.001). White regions: CC-equilibrium condition not satisfied.

Comparison of the results of Figure 9 with the game theoretical analysis of Section 2 reveals three main results

- For the learning model, quantitative effects of the payoff parameters a and b on $\alpha^{0.5}$ and $\alpha^{0.9}$ arise *only* for the area where ***both*** equilibrium conditions are met. By contrast, in the game theoretical analysis we found effects of a and b throughout the entire PD-triangle ($a < b$).

- Within the region where both equilibrium conditions are satisfied, the qualitative effects of the parameters a and b predicted by the learning model are consistent with results of the game theoretical analysis (Figures 3-6). Higher levels of a make the conditions for mutual cooperation more restrictive, i.e. the level of α required for a certain probability q

for the CC-equilibrium increases. Conversely, higher levels of b facilitate mutual cooperation.

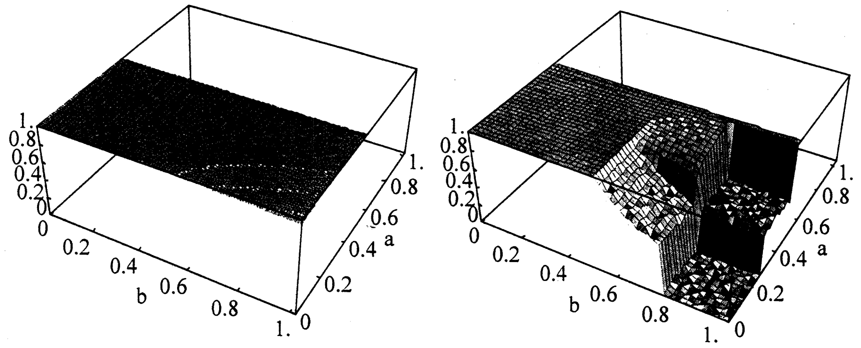
- Comparison of the left part of Figure 9 ($\alpha^{0.5}$) with Figure 4 above clearly indicates that cooperation is harder to attain for learning actors as compared to rational players. More precisely, throughout the region where $a < b$, the continuation probability required for a 90% *chance* of mutual cooperation between learning actors, $\alpha^{0.9}$, is larger or equal to the continuation probability α^* that *guarantees* the individual rationality of mutual cooperation on basis of Trigger strategies. To preview, we show below that learning rates higher than *l=1* may partially change this result.

It is immediately clear that quantitative effects of a and b on any α^q arise only in the region where the CC-equilibrium condition is satisfied. The reason is that otherwise lock-in on CC can *never* occur, regardless of the number of iterations for which the game is continued. However, Figure 9 also shows that quantitative effects arise only in the region where the *DD-equilibrium condition* is *not* satisfied. To understand this result of Figure 9, consider a parameter combination that satisfies both equilibrium conditions and assume a learning rate of *l=1*. For this parameter combination *more than* 50% of all runs will necessarily end up in the DD-equilibrium. To explain, with initial propensities of 0.5, about 50% of the simulation runs start in one of the states CD or DC. In these runs, both players' propensity to cooperate immediately drops to zero. As a consequence, the dynamics of these runs immediately converge on the DD-equilibrium. In addition, a certain fraction of the runs that start with CC or DD will also converge on this equilibrium, because at some point they enter one of the asymmetric states. For example, with *a=0.25* and *b=0.25* both players' propensity to cooperate is 0.75 in the iteration subsequent to an initial outcome of CC. This leaves a probability of 0.75 (1-0.75) = 0.1875 for the outcome CD and the same probability for the outcome DC. Hence, with a probability of 0.375 the outcome of the second iteration is CD or DC. Accordingly, about 37.5% of the runs that start with CC converge on the DD-equilibrium in the third iteration. Continuation of this reasoning shows that additional 22% of the runs starting with CC converge on DD in the fourth iteration, about 12% end up there in the fifth iteration, and so on. This example shows that the probability for the DD-equilibrium within the first four iterations always exceeds 50% if the DD-equilibrium condition is satisfied. This explains why our simulations reveal no quantitative effects on any α^q for $q \geq 50\%$. The reason is that the proportion of runs attaining CC-lock-in before the game ends always remains below 50%.

3.3.3 Effects of the Learning Rate

The preceding analyses revealed that for a certain combination of parameters of the learning model, the conditions for cooperation between learning actors are more restrictive as compared to conditions for rational players using Trigger strategies. However, our study leaves open whether this result generalises to other parameter combinations than the one we actually simulated. In explorative simulations we found that neither variation in the expectation level e, nor variation in the initial propensities for cooperation, p_{i0}, change the main results of the model comparison. At the same time, we found that the learning rate l has a profound effect on the prospects for cooperation between adaptive actors. Accordingly, this section addresses effects of variation in the learning rate.

To assess results of the learning rate, we repeated the simulation series of Figure 9 with a relatively low learning rate, l=*0.5*, and a comparatively high learning rate of l=*2*. Figure 10 shows the results for the estimated continuation probability required for a 50% chance of stable mutual cooperation, $\alpha^{0.5}$.



l=0.5 l=2

Figure 10: Estimated minimum continuation probability for 50% chance on CC-lock-in ($\alpha^{0.5}$) for two different learning rates l. (Initial propensities p=0.5, e=0.5, ϵ=0.001). White regions: CC-equilibrium condition not satisfied.

Comparison of the two parts of Figure 10 and the left part of Figure 9 ($\alpha^{0.5}$) suggest two main observations.

- The learning rate l affects the continuation probability required to attain a chance of 50% mutual cooperation only in the region of the parameter space where exclusively the CC-equilibrium condition is satisfied. In particular, for l=*0.5* the number of iterations required for CC-lockin is very large, as exemplified by the fact that for this level of l, $\alpha^{0.5}$ is almost equal to 1 throughout this region of the parameter space.

- If exclusively the CC-equilibrium condition is satisfied, we find that the higher the learning rate l, the smaller is the continuation probability required to attain a probability of 50% for mutual cooperation.

Furthermore, comparison of the right part of Figure 10 with Figure 4 above suggests the following.

- In a part of the region of the parameter space where exclusively the CC-equilibrium condition is satisfied, high learning rates make conditions for mutual cooperation between learning actors *less* restrictive as compared to rational players. In particular, we find for *l=2* that 50% or more of all runs for learning actors attain CC-equilibrium even *without any repetition* ($\alpha^{0.5} = 0$), when a falls below 0.25 and b exceeds 0.75. By contrast, for rational actors, a threshold level of $\alpha^* = 0$ can only occur for the extreme case of *a=0*.

The explanation of why effects of the learning rate only occur in the CC-equilibrium region follows the line sketched for the case of *l=1*. Again, the central reason is that propensities to cooperate drop considerably after an asymmetric outcome (CD or DC) occurred at some point in the game. For learning rates smaller than one, DD-equilibrium does not immediately obtain after an asymmetric outcome, because l dampens the reduction of propensities. However, with *l=0.5* initial propensities still drop enough to make it considerably likely that a sequence of consecutive DD-outcomes arises which is long enough to drive propensities eventually into the DD-equilibrium. Moreover, with l smaller than one, the number of consecutive CC-iterations is relatively large that is required before propensities can stabilise on the CC-equilibrium. As a consequence, in the region of the parameter space where both equilibrium conditions are satisfied, less than 50% of all runs attain CC-equilibrium at all, regardless whether l is smaller or larger than one.

The striking feature of the learning rate of *l=2* is that about 50% of all runs attain cooperation without any repetition. The explanation of this phenomenon follows from the updating rules for propensities described above. More in particular, we show, *firstly*, that every run that starts with CC immediately attains CC-lock-in, when exclusively the CC-equilibrium condition is satisfied and a falls below a certain threshold. Moreover, we show, *secondly*, that every run that starts with DD immediately attains CC-lock-in, when b exceeds a certain threshold and exclusively the CC-equilibrium condition is satisfied. As to the first assertion, the reasoning is that the reinforcement following an initial CC outcome will always drive both players' propensities into CC-lock-in as soon as the term $l\Delta p$ is large enough to let propensities become equal to or larger than one. For *l=2*, this is the case when $a \leq 0.25$. More in general, Table 3 above implies the condition $a \leq (l-1)/2l$ (assuming *e=0.5* and initial propensities of *p=0.5*). Hence, when this condition is satisfied about 25% of all runs attain CC-lock-in without any repetition. The reasoning for the second assertion follows along the same line. The negative reinforcement of defection that follows an initial DD outcome will always drive both players' propensities into CC-lock-in when the term $l\Delta p$ exceeds 0.5. For *l=2*, this is the case when $b \geq 0.75$. More in general, Table 3 implies the condition $b \geq (l+1)/2l$ (assuming *e=0.5* and initial propensities of *p=0.5*). Hence, this condition guarantees that another 25% of all runs attain CC-lock-in without any repetition. To conclude, when both conditions are satisfied, a learning rate of *l=2* guarantees that about 50% of the runs immediately attain CC-lock-in. Finally, the above reasoning also explains the discontinuities in the left part of Figure 10. These discontinuities arise exactly at the thresholds where *a* falls below 0.25 and *b* exceeds 0.75, each condition adding about 25% to the chance that CC-lock-in occurs immediately after iteration 1.

4. Discussion

Axelrod's famous *The Evolution of Cooperation* (1984) has popularised the notion that cooperation is feasible even in Prisoner's Dilemma Situations. The central mechanism Axelrod proposed is conditional cooperation in repeated games on basis of Tit-for-Tat strategies. However, following Axelrod's work analysts often tend to overlook that the particular payoff parameters Axelrod has chosen represent only one special case of a Prisoner's dilemma situation (for a similar criticism cf. Binmore 1998). Accordingly, our paper analyses conditions for mutual cooperation in *all possible symmetrical 2×2 Prisoner's Dilemmas.* Moreover, we compare two different micro foundations of individual decision making that yield conditions for cooperation in iterated games. The first micro foundation is the strategically rational actor of game theory, the second micro foundation is a simple model of adaptive learning decision making.

We represent the payoff space of 2×2 PDs as a unit square spanned by two parameters ranging between zero and one. This allows for an easily accessible representation of some well-known game theoretical results about the conditions for mutual cooperation. The game theoretical analysis demonstrates that sustained conditional cooperation is in principle feasible for *every* game in the PD-triangle of the unit square. However, this does not hold for Tit-for-Tat, it only holds for an extremely 'intolerant' form of conditional cooperation, Trigger strategies. Moreover, our analysis illustrates that the smaller the payoff of mutual cooperation and the larger the payoff of mutual defection, the higher is the probability for continuation of the game (Axelrod's 'shadow of the future') that is required to make conditional cooperation an individually rational behaviour. Finally, we show that conditional cooperation between Tit-for-Tat strategies is *never* individually rational in one third of the payoff space, i.e. in those games where the payoff of mutual cooperation is relatively small and only slightly exceeds the payoff of mutual defection.

The unit square also proved to be a useful instrument for comparison of micro foundations. We used a Bush-Mosteller stochastic learning model to formalise a simple trial-and-error learning mechanism. Analysis of the equilibrium conditions of the learning process shows two main qualitative differences between the micro foundations. First, in contrast with rational players, learning actors may attain mutual cooperation even when this is *inefficient* from both players' point of view. Second, for a considerable fraction of the payoff space, learning actors can *never attain sustained mutual cooperation* even when this is *efficient* for both players.

We used computer simulation to obtain a quantitative comparison of the learning mechanism with game theoretical results. Broadly, we found that for a large range of parameter combinations of the learning model, the conditions for mutual cooperation are more restrictive for learning actors than they are for rational players. More technically, in this region of the parameter space the probability of continuation of the game that is required for a chance of at least 50% of sustained mutual cooperation between learning actors is considerably larger than the continuation probability that suffices to *guarantee* the individual rationality of mutual cooperation on basis of Trigger strategies. However, further analysis revealed conditions under which cooperation between learning players may be easier to attain as compared to cooperation between rational actors. This is the case when the learning process is relatively fast and the payoffs of mutual defection are small in comparison with the payoffs of mutual cooperation. In this case, there is a considerable chance that learning players attain sustained mutual cooperation without any 'shadow of the future', a result that is clearly inconsistent with rational decision making.

Clearly, our analysis illustrates that the importance of Tit-for-Tat strategies for mutual cooperation needs to be put into perspective. At the same time, we are aware of a number of restrictions underlying our study. As to the game theoretical analysis, we employ at least two strong idealisations. The first idealisation is to assume that actors possess perfect information on every aspect of the game and their opponents' past behaviour. However, previous game theoretical work suggests that our analysis can be extended to games with imperfect information. For example, Hegselmann and Flache (1998) derived equilibrium conditions for a PD game where actors observe only some of the past moves of their opponents (for similar analyses cf. Green/Porter 1984; Bendor/Mookherjee 1987). The second idealisation in the game theoretical analysis is that we consider only two types of equilibria, Trigger strategy equilibria and Tit-for-Tat equilibria. It is well known that in iterated PD-games a multitude of different equilibria with other strategies than TR or TFT are feasible. However, TR and TFT equilibria are probably the equilibria that are best analysed in the game theoretical literature (e.g., Friedman 1977, 1986; Axelrod 1984; Taylor 1987) and we were mainly interested in visualisation of and comparison with previous results. Moreover, the conditions for TR equilibria in particular may be considered as necessary conditions for the possibility of any form of conditional cooperation, because TR is the strategy that imposes the severest sanction available for conditional cooperaters, eternal defection (cf. Myerson 1991).

As to the comparison of micro foundations, we focused on two simple and extreme representations of decision making, perfect strategic rationality on the one hand and a simple reinforcement mechanism on the other hand. Clearly, more 'realistic' models of decision making might be employed that combine elements of rational 'forward-looking' and learning 'backward-looking' decision making (cf. Fudenberg/Levine 1998). However, we believe that restriction to our simple models is useful as a first step, because the two micro foundations we use may be considered as idealised counterparts in terms of the degree of cognitive sophistication (or: lack thereof) attributed to the players. With this comparison we could systematically assess whether "bounded rationality" (Simon 1982) does at all affect outcomes of iterated 2×2 Prisoner's Dilemmas—and it does. An important lesson of our analysis is that in the space of all possible 2×2 PDs different micro foundations do matter. This suggests that researchers can not safely rely on the assumption that implementing simple models of decision making will yield the same results that may be obtained when more sophisticated decision rules are built into the agents.

Bibliography

Axelrod, R. (1984), *The Evolution of Cooperation*, New York; dt.: *Die Evolution der Kooperation*, München 1987

Bendor, J./D. Mookherjee (1987), Institutional Structure and the Logic of Ongoing Collective Action, in: *American Political Science Review 81*, 129–154

Binmore, K. (1998), Review of Axelrod, R. 1997. The Complexity of Cooperation, in: *Journal of Artificial Societies and Social Simulation 1*, http://www.soc.surrey.ac.uk/JASSS/1/1/review1.html

Bush, R. R./F. Mosteller (1955), *Stochastic Models for Learning*, New York

Friedman, J. W. (1977), *Oligopoly and the Theory of Games*, Amsterdam

— (1986), *Game Theory with Applications to Economics*, Oxford

Fudenberg, D./D. K. Levine (1998), *The Theory of Learning in Games*, Cambridge

Green, E. J./R. H. Porter (1984), Noncooperative Collusion under Imperfect Price Information, in: *Econometrica 52*, 87–100

Harris, R. J. (1969), A Geometric Classification System For 2×2 Interval-Symmetric Games, in: *Behavioral Science 14*, 138–146

Hegselmann, R./A. Flache (1998), Understanding Complex Social Dynamics: A Plea For Cellular Automata Based Modelling, in: *Journal of Artificial Societies and Social Simulation*, http://www.soc.surrey.ac.uk/JASSS/1/3/1.html

Macy, M. W. (1989), Walking out of Social Traps. A Stochastic Learning Model for the Prisoner's Dilemma, in: *Rationality and Society 2*, 197–219

— (1991), Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange, in: *American Journal of Sociology 97*, 808–843

Myerson, R. B. (1991), *Game Theory. Analysis of Conflict*, Cambridge

Rapoport, A./A. M. Chammah (1965), *Prisoner's Dilemma. A Study of Conflict and Cooperation*, Ann Arbor

Roth, A. E./I. Erev. (1995), Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term, in: *Games and Economic Behavior 8*, 164–212

Simon, H. A. (1982), A Behavioral Model of Rational Choice, in: H. A. Simon (eds.), *Models of Bounded Rationality: Behavioral Economics and Business Organization 2*, (first published in 1955), Cambridge

Taylor, M. (1987), *The Possibility of Cooperation*, Cambridge