Advances in Complex Systems, Vol. 8, No. 1 (2005) 87–116

World Scientific
www.worldscientific.com

HOW INDIVIDUALS LEARN TO TAKE TURNS: EMERGENCE OF ALTERNATING COOPERATION IN A CONGESTION GAME AND THE PRISONER'S DILEMMA

DIRK HELBING, MARTIN SCHÖNHOF and HANS-ULRICH STARK

*Institute for Transport & Economics, Dresden University of Technology,
Andreas-Schubert-Str. 23, 01062 Dresden, Germany*

JANUSZ A. HOŁYST

*Faculty of Physics and Center of Excellence for Complex Systems Research,
Warsaw University of Technology, Koszykowa 75, PL-00-662 Warsaw, Poland*

Received 16 March 2005

In many social dilemmas, individuals tend to generate a situation with low payoffs instead of a system optimum ("tragedy of the commons"). Is the routing of traffic a similar problem? In order to address this question, we present experimental results on humans playing a route choice game in a computer laboratory, which allow one to study decision behavior in repeated games beyond the Prisoner's Dilemma. We will focus on whether individuals manage to find a cooperative and fair solution compatible with the system-optimal road usage. We find that individuals tend towards a user equilibrium with equal travel times in the beginning. However, after many iterations, they often establish a coherent oscillatory behavior, as taking turns performs better than applying pure or mixed strategies. The resulting behavior is fair and compatible with system-optimal road usage. In spite of the complex dynamics leading to coordinated oscillations, we have identified mathematical relationships quantifying the observed transition process. Our main experimental discoveries for 2- and 4-person games can be explained with a novel reinforcement learning model for an arbitrary number of persons, which is based on past experience and trial-and-error behavior. Gains in the average payoff seem to be an important driving force for the innovation of time-dependent response patterns, i.e. the evolution of more complex strategies. Our findings are relevant for decision support systems and routing in traffic or data networks.

Keywords: Game theory; reinforcement learning; multi-agent simulation.

1. Introduction

Congestion is a burden of today's traffic systems, affecting the economic prosperity of modern societies. Yet, the optimal distribution of vehicles over alternative routes is still a challenging problem and uses scarce resources (street capacity) in an inefficient way. Route choice is based on interactive, but decentralized individual decisions, which cannot be well described by classical utility-based decision models [27]. Similarto the minority game [16, 39, 43], it is reasonable for different people to react to the same situation or information in *different* ways. As a consequence, individuals tend to develop characteristic response patterns or roles [26]. Thanks to this differentiation process, individuals learn to coordinate better in the course of time. However, according to current knowledge, selfish routing does not establish the system optimum of minimum overall travel times. It rather tends to establish the Wardrop equilibrium, a special user or Nash equilibrium characterized by equal travel times on all alternative routes chosen from a certain origin to a given destination (while routes with longer travel times are not taken) [71].

Since Pigou [53], it has been suggested to resolve the problem of inefficient road usage by congestion charges, but are they needed? Is the missing establishment of a system optimum just a problem of varying traffic conditions and changing origin-destination pairs, which make route-choice decisions comparable to one-shot games? Or would individuals in an *iterated* setting of a day-to-day route choice game with identical conditions spontaneously establish cooperation in order to increase their returns, as the folk theorem suggests [6]?

How would such a cooperation look? Taking turns could be a suitable solution [62]. While simple symmetrical cooperation is typically found for the repeated Prisoner's Dilemma [2, 3, 44–46, 49, 52, 55, 59, 64, 67, 69], emergent alternating reciprocity has been recently discovered for the games Leader and Battle of the Sexes [11].[a] Note that such coherent oscillations are a time-dependent but deterministic form of individual decision behavior, which can establish a persistent phase-coordination, while mixed strategies, i.e. statistically varying decisions, can establish cooperation only by chance or on statistical average. This difference is particularly important when the number of interacting persons is small, as in the particular route choice game discussed below.

Note that oscillatory behavior has been found in iterated games before:

- In the rock-paper-scissors game [67], cycles are predicted by the game-dynamical equations due to unstable stationary solutions [28].
- Oscillations can also result from coordination problems [1, 29, 31, 33], at the cost of reduced system performance.
- Moreover, blinker strategies may survive in repeated games played by a mixture of finite automata [5] or result through evolutionary strategies [11, 15, 16, 38, 39, 42, 43, 74].

However, these oscillation-generating mechanisms are clearly distinguishable from the establishment of phase-coordinated alternating reciprocity we are interested in (coherent oscillatory cooperation to reach the system optimum).

Our paper is organized as follows: In Sec. 2, we will formally introduce the route choice game for N players, including issues like the Wardrop equilibrium [71] and

[a] See Fig. 2 for a specification of these games.

the Braess paradox [10]. Section 3 will focus on the special case of the 2-person route choice game, compare it with the minority game [1, 15, 16, 38, 39, 42, 43, 74], and discuss its place in the classification scheme of symmetrical 2 × 2 games. This section will also reveal some apparent shortcomings of the previous game-theoretical literature:

- While it is commonly stated that among the 12 ordinally distinct, symmetrical 2 × 2 games [11, 57] only 4 archetypical 2 × 2 games describe a strategical conflict (the Prisoner's Dilemma, the Battle of the Sexes, Chicken, and Leader) [11, 18, 56], we will show that, for specific payoffs, the route choice game (besides Deadlock) also represents an interesting strategical conflict, at least for iterated games.
- The conclusion that conservative driver behavior is best, i.e. it does not pay off to change routes [7, 65, 66], is restricted to the special case of route-choice games with a system-optimal user equilibrium.
- It is only half the truth that cooperation in the iterated Prisoner's Dilemma is characterized by symmetrical behavior [11]. Phase-coordinated asymmetric reciprocity is possible as well, as in some other symmetrical 2 × 2 games [11].

New perspectives arise by less restricted specifications of the payoff values.

In Sec. 4, we will discuss empirical results of laboratory experiments with humans [12, 18, 32]. According to these, reaching a phase-coordinated alternating state is only one problem. Exploratory behavior and suitable punishment strategies are important to establish asymmetric oscillatory reciprocity as well [11, 20]. Moreover, we will discuss several coefficients characterizing individual behavior and chances for the establishment of cooperation. In Sec. 5, we will present multi-agent computer simulations of our observations, based on a novel win-stay, lose-shift [50, 54] strategy, which is a special kind of reinforcement learning strategy [40]. This approach is based on individual historical experience [13] and, thereby, clearly differs from the selection of the best-performing strategy in a set of hypothetical strategies as assumed in studies based on evolutionary or genetical algorithms [5, 11, 15, 16, 39, 42, 43]. The final section will summarize our results and discuss their relevance for game theory and possible applications such as data routing algorithms [35, 72], advanced driver information systems [8, 14, 30, 37, 41, 63, 70, 73], or road pricing [53].

2. The Route Choice Game

In the following, we will investigate a scenario with two alternative routes between a certain origin and a given destination, say, between two places or towns A and B (see Fig. 1). We are interested in the case where both routes have different capacities, say a freeway and a subordinate or side road. While the freeway is faster when it is empty, it may be reasonable to use the side road when the freeway is congested.

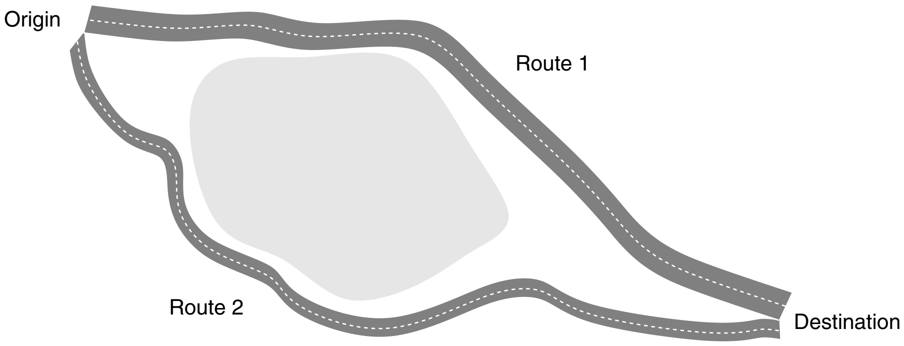


Fig. 1. Illustration of the investigated day-to-day route choice scenario. We study the dynamic decision behavior in a repeated route choice game, where a given destination can be reached from a given origin via two different routes, a freeway (route 1) and a side road (route 2).

The "success" of taking route i could be measured in terms of its inverse travel time $1/T_i(N_i) = V_i(N_i)/L_i$, where L_i is the length of route i and $V_i(N_i)$ the average velocity when N_i of the N drivers have selected route i. One may roughly approximate the average vehicle speed V_i on route i by the linear relationship [24]

$$V_i(N_i) = V_i^0\left(1 - \frac{N_i(t)}{N_i^{\max}}\right), \tag{1}$$

where V_i^0 denotes the maximum velocity (speed limit) and $N_i^{\max}$ the capacity, i.e. the maximum possible number of vehicles on route i. With $A_i = V_i^0/L_i$ and $B_i = V_i^0/(N_i^{\max} L_i)$, the inverse travel time then obeys the relationship

$$1/T(N_i) = A_i - B_i N_i, \tag{2}$$

which is linearly decreasing with the road occupancy N_i. Other monotonously falling relationships $V_i(N_i)$ would make the expression for the inverse travel times nonlinear, but they would probably not lead to qualitatively different conclusions.

The user equilibrium of equal travel times is found for a fraction

$$\frac{N_1^{\mathrm{e}}}{N} = \frac{B_2}{B_1 + B_2} + \frac{1}{N}\frac{A_1 - A_2}{B_1 + B_2} \tag{3}$$

of persons choosing route 1. In contrast, the system optimum corresponds to the maximum of the overall inverse travel times $N_1/T_1(N_1) + N_2/T_2(N_2)$ and is found for the fraction

$$\frac{N_1^{\mathrm{o}}}{N} = \frac{B_2}{B_1 + B_2} + \frac{1}{2N}\frac{A_1 - A_2}{B_1 + B_2} \tag{4}$$

of 1-decisions. The difference between both fractions vanishes in the limit $N \to \infty$. Therefore, only experiments with a few players allow one to find out whether the test persons adapt to the user equilibrium or to the system optimum. We will see that both cases have completely different dynamical implications: While the most successful strategy to establish the user equilibrium is to stick to the same decision in subsequent iterations [27, 65, 66], the system optimum can only be reached by a time-dependent strategy (at least, if no participant is ready to pay for the profits of others).

Note that alternative routes can reach comparable travel times only when the total number N of vehicles is large enough to fulfil the relationships $P_1(N) < P_2(0) = A_2$ and $P_2(N) < P_1(0) = A_1$. Our route choice game will address this traffic regime and additionally assume $N \le N_i^{\max}$. The case $N_i = N_i^{\max}$ corresponds to a complete gridlock on route i.

Finally, it may be interesting to connect the previous quantities with the vehicle densities ρ_i and the traffic flows Q_i: If route i consists of I_i lanes, the relation with the average vehicle density is $\rho_i(N_i) = N_i/(I_i L_i)$, and the relation with the traffic flow is $Q_i(N_i) = \rho_i V_i(N_i) = N_i/[I_i T_i(N_i)]$.

In the following, we will linearly transform the inverse travel time $1/T_i(N_i)$ in order to define the so-called payoff

$$P_i(N_i) = C_i - D_i N_i \tag{5}$$

for choosing route i. The payoff parameters C_i and D_i depend on the parameters A_i, B_i, and N, but will be taken as constant. We have scaled the parameters so that we have the payoff $P_i(N_i^{\mathrm{e}}) = 0$ (zero payoff points) in the user equilibrium and the payoff $N_1 P_1(N_1^{\mathrm{o}}) + N_2 P_2(N - N_1^{\mathrm{o}}) = 100\,N$ (an average of 100 payoff points) in the system optimum. This serves to reach generalizable results and to provide a better orientation to the test persons.

Note that the investigation of social (multi-person) games with linearly falling payoffs is not new [33]. For example, Schelling [62] has discussed situations with "conditional externality," where the outcome of a decision depends on the independent decisions of potentially many others [62]. Pigou has addressed this problem, which has been recently focused on by Schreckenberg and Selten's project SURVIVE [7, 65, 66] and others [8, 41, 58].

The route choice game is a special congestion game [22, 47, 60]. More precisely speaking, it is a multi-stage symmetrical N-person single commodity congestion game [68]. Congestion games belong to the class of "potential games" [48], for which many theorems are available. For example, it is known that there always exists a Wardrop equilibrium [71] with essentially unique Nash flows [4]. This is characterized by the property that no individual driver can decrease his or her travel time by a different route choice. If there are several alternative routes from a given origin to a given destination, the travel times on all used alternative routes in the Wardrop equilibrium are the same, while roads with longer travel times are not used. However, the Wardrop equilibrium as the expected outcome of selfish routing does not generally reach the system optimum, i.e. minimize the total travel times. Nash flows are often inefficient, and selfish behavior implies the possibility of decreased network performance.[b] This is particularly pronounced for

[b] For more details, see the work by T. Roughgarden.

the Braess paradox [10, 61], according to which additional streets may sometimes increase the overall travel time and reduce the throughput of a road network. The reason for this is the possible existence of badly performing Nash equilibria, in which no single person can improve his or her payoff by changing the decision behavior.

In fact, recent laboratory experiments indicate that, in a "day-to-day route choice scenario" based on selfish routing, the distribution of individuals over the alternative routes is fluctuating around the Wardrop equilibrium [27, 63]. Additional conclusions from the laboratory experiments by Schreckenberg, Selten *et al.* are as follows [65, 66]:

- Most people, who change their decision frequently, respond to their experience on the previous day (i.e. in the last iteration).
- There are only a few different behavioral patterns: direct responders (44%), contrarian responders (14%), and conservative persons, who do not respond to the previous outcome.
- It does not pay off to react to travel time information in a sensitive way, as conservative test persons reach the smallest travel times (the largest payoffs) on average.
- People's reactions to short-term travel forecasts can invalidate these. Nevertheless, travel time information helps to match the Wardrop equilibrium, so that excess travel times due to coordination problems are reduced.

A closer experimental analysis based on longer time series (i.e. more iterations) for smaller groups of test persons reveals a more detailed picture [26]:

- Individuals do not only show an adaptive behavior to the travel times on the previous day, but also change their response pattern in time [26, 34].
- In the course of time, one finds a differentiation process which leads to the development of characteristic, individual response patterns, which tend to be almost deterministic (in contrast to mixed strategies).
- While some test persons respond to small differences in travel times, others only react to medium-sized deviations, still others people respond to large deviations, etc. In this way, overreactions of the group to deviations from the Wardrop equilibrium are considerably reduced.

Note that the differentiation of individual behaviors is a way to resolve the coordination problem to match the Wardrop equilibrium exactly, i.e. which participant should change his or her decision in the next iteration in order to compensate for a deviation from it. This implies that the fractions of specific behavioral response patterns should depend on the parameters of the payoff function. A certain fraction of "stayers," who do not respond to travel time information, can improve the coordination in the group, i.e. the overall performance. However, stayers can also prevent the establishment of a system optimum, if alternating reciprocity is needed, see Eq. (14).

3. Classification of Symmetrical 2 × 2 Games

In contrast to previous laboratory experiments, we have studied the route choice game not only with a very high number of repetitions, but also with a small number $N \in \{2, 4\}$ of test persons, in order to see whether the system optimum or the Wardrop equilibrium is established. Therefore, let us shortly discuss how the two-person game relates to previous game-theoretical studies.

Iterated symmetrical two-person games have been intensively studied [12, 18], including Stag Hunt, the Battle of the Sexes, or the Chicken Game (see Fig. 2). They can all be represented by a payoff matrix of the form $\mathbf{P} = (P_{ij})$, where P_{ij} is the success ("payoff") of person 1 in a one-shot game when choosing strategy $i \in \{1, 2\}$ and meeting strategy $j \in \{1, 2\}$. The respective payoffs of the second person are given by the symmetrical values P_{ji}. Figure 2 shows a systematics of the previously mentioned and other kinds of symmetrical two-person games [21]. The relations

$$P_{21} > P_{11} > P_{22} > P_{12}, \tag{6}$$

for example, define a Prisoner's Dilemma. In this paper, however, we will mainly focus on the two-person route choice game defined by the conditions

$$P_{12} > P_{11} > P_{21} > P_{22} \tag{7}$$

(see Fig. 3). Despite some common properties, this game differs from the minority game [16, 39, 43] or El Farol bar problem [1] with $P_{12}, P_{21} > P_{11}, P_{22}$, as a minority decision for alternative 2 is less profitable than a majority decision for

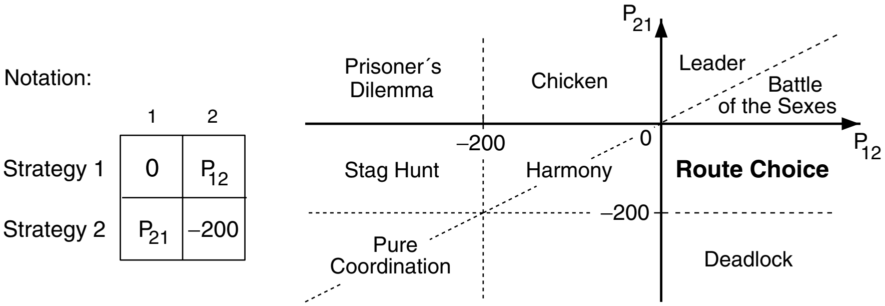
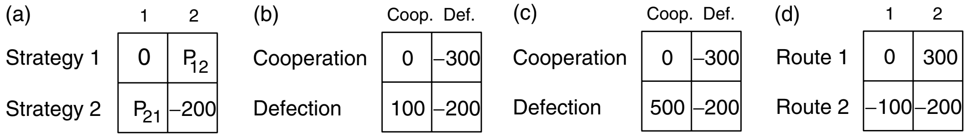


Fig. 2. Classification of symmetrical 2 × 2 games according to their payoffs P_{ij}. Two payoff values have been kept constant as payoffs may be linearly transformed and the two strategies of the one-shot game renumbered. Our choice of $P_{11} = 0$ and $P_{22} = -200$ was made to define a payoff of 0 points in the user equilibrium and an average payoff of 100 in the system optimum of our investigated route choice game with $P_{12} = 300$ and $P_{21} = -100$.



(a)

	1	2
Strategy 1	0	P_{12}
Strategy 2	P_{21}	–200

(b)

	Coop.	Def.
Cooperation	0	–300
Defection	100	–200

(c)

	Coop.	Def.
Cooperation	0	–300
Defection	500	–200

(d)

	1	2
Route 1	0	300
Route 2	–100	–200

Fig. 3. Payoff specifications of the symmetrical 2 × 2 games investigated in this paper. (a) General payoff matrix underlying the classification scheme of Fig. 2. (b) and (c) Two variants of the Prisoner's Dilemma. (d) Route choice game with a strategical conflict between the user equilibrium and the system optimum.

alternative 1. Although oscillatory behavior has been found in the minority game as well [9, 15, 16, 36, 43], an interesting feature of the route choice experiments discussed in the following is the regularity and phase-coordination (coherence) of the oscillations.

The two-person route choice game fits well into the classification scheme of symmetrical 2 × 2 games. In Rapoport and Guyer's taxonomy of 2 × 2 games [57], the two-person route choice game appears on page 211 as game number 7 together with four other games with strongly stable equilibria. Since then, the game has almost been forgotten and did not have a commonly known interpretation or name. Therefore, we suggest naming it the two-person "route choice game." Its place in the extended Eriksson–Lindgren scheme of symmetrical 2 × 2 games is graphically illustrated in Fig. 2.

According to the game-theoretical literature, there are 12 ordinally distinct, symmetric 2 × 2 games [57], but after excluding strategically trivial games in the sense of having equilibrium points that are uniquely Pareto-efficient, there remain four archetypical 2 × 2 games: the Prisoner's Dilemma, the Battle of the Sexes, Chicken (Hawk-Dove), and Leader [56]. However, this conclusion is only correct if the four payoff values P_{ij} are specified by the four values $\{1, 2, 3, 4\}$. Taking different values would lead to a different conclusion: If we name subscripts so that $P_{11} > P_{22}$, a strategical conflict between a user equilibrium and the system optimum results when

$$P_{12} + P_{21} > 2P_{11}. \tag{8}$$

Our conjecture is that players tend to develop alternating forms of reciprocity if this condition is fulfilled, while symmetric reciprocity is found otherwise. This has the following implications (see Fig. 2):

- If the 2 × 2 games Stag Hunt, Harmony, or Pure Coordination are repeated frequently enough, we always expect a symmetrical form of cooperation.
- For Leader and the Battle of the Sexes, we expect the establishment of asymmetric reciprocity, as has been found by Browning and Colman with a computer simulation based on a genetic algorithm incorporating mutation and crossing-over [11].

- For the games Route Choice, Deadlock, Chicken, and Prisoner's Dilemma both, symmetric (simultaneous) and asymmetric (alternating) forms of cooperation are possible, depending on whether condition (8) is fulfilled or not. Note that this condition cannot be met for some games, if one is restricted to ordinal payoff values $P_{ij} \in \{1, 2, 3, 4\}$ only. Therefore, this interesting problem has been largely neglected in the past (with a few exceptions, e.g. Ref. 51). In particular, convincing experimental evidence of alternating reciprocity is missing. The following sections of this paper will, therefore, not only propose a simulation model, but also focus on an experimental study of this problem, which promises interesting new results.

4. Experimental Results

Altogether we have carried out more than 80 route choice experiments with different experimental setups, all with different participants. In the 24 two-person (12 four-person) experiments evaluated here (see Figs. 4–15), test persons were instructed to choose between two possible routes between the same origin and destination. They

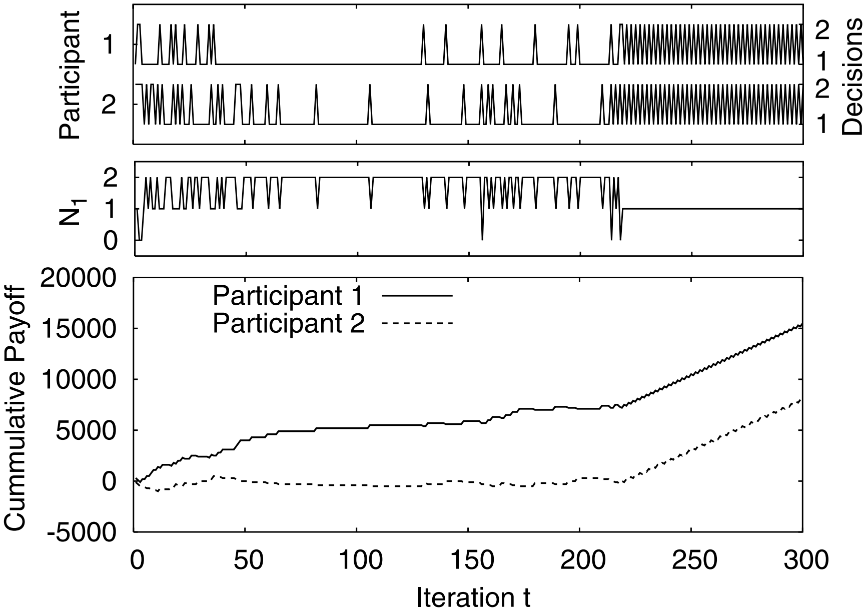


Fig. 4. Representative example for the emergence of coherent oscillations in a two-person route choice experiment with the parameters specified in Fig. 3(d). Top: Decisions of both participants over 300 iterations. Center: Number $N_1(t)$ of 1-decisions over time t. Note that $N_1 = 1$ corresponds to the system optimum, while $N_1 = 2$ corresponds to the user equilibrium of the one-shot game. Bottom: Cumulative payoff of both players in the course of time t (i.e. as a function of the number of iterations). Once the coherent oscillatory cooperation is established ($t > 220$), both individuals have high payoff gains on average.

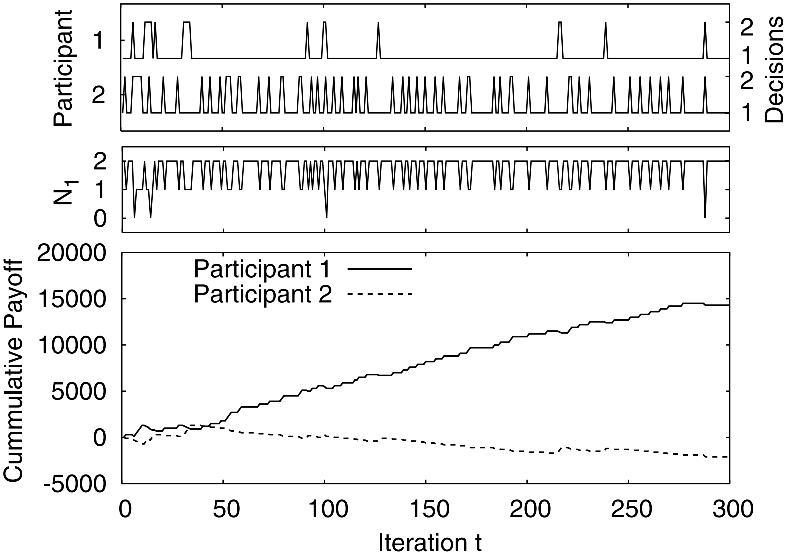


Fig. 5. Representative example for a two-person route choice experiment, in which no alternating cooperation was established. Due to the small changing frequency of participant 1, there were not enough cooperative episodes that could have initiated coherent oscillations. Top: Decisions of both participants over 300 iterations. Center: Number $N_1(t)$ of 1-decisions over time t. Bottom: The cumulative payoff of both players in the course of time t shows that the individual with the smaller changing frequency has higher profits.

knew that route 1 corresponds to a "freeway" (which may be fast or congested), while route 2 represents an alternative route (a "side road"). Test persons were also informed that, if two [three] participants chose route 1, everyone would receive 0 points, while if half of the participants chose route 1, they would receive the maximum average amount of 100 points, but 1-choosers would profit at the cost of 2-choosers. Finally, participants were told that everyone could reach an average of 100 points per round with variable, situation-dependent decisions, and that the (additional) individual payment after the experiment would depend on their cumulative payoff points reached in at least 300 rounds (100 points = 0.01 EUR).

Let us first focus on the two-person route-choice game with the payoffs $P_{11} = P_1(2) = 0, P_{12} = P_1(1) = 300, P_{21} = P_2(1) = -100$, and $P_{22} = P_2(2) = -200$ (see Fig. 3(d)), corresponding to $C_1 = 600, D_1 = 300, C_2 = 0$, and $D_2 = 100$. For this choice of parameters, the best individual payoff in each iteration is obtained by choosing route 1 (the "freeway") and have the co-player(s) choose route 2. Choosing route 1 is the dominant strategy of the one-shot game, and players are tempted to use it. This produces an initial tendency towards the "strongly stable" user equilibrium [57] with 0 points for everyone. However, this decision behavior is not Pareto efficient in the repeated game. Therefore, after many iterations, the players often learn to establish the Pareto optimum of the multi-stage supergame by selecting

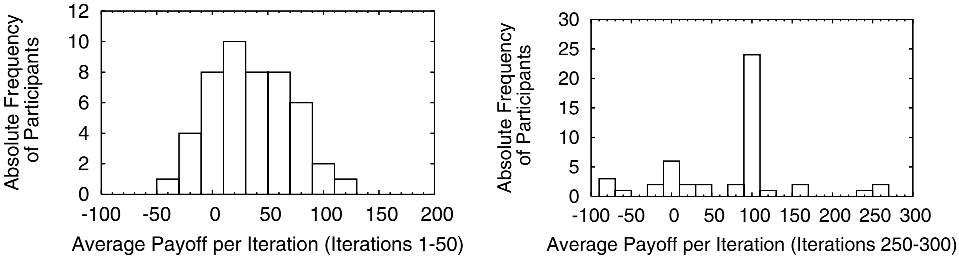


Fig. 6. Frequency distributions of the average payoffs of the 48 players participating in our 24 two-person route choice experiments. Left: Distribution during the first 50 iterations. Right: Distribution between iterations 250 and 300. The initial distribution with a maximum close to 0 points (left) indicates a tendency towards the user equilibrium corresponding to the dominant strategy of the one-shot game. However, after many iterations, many individuals learn to establish the system optimum with a payoff of 100 points (right).

route 1 in turns (see Fig. 4). As a consequence, the experimental payoff distribution shows a maximum close to 0 points in the beginning and a peak at 100 points after many iterations (see Fig. 6), which clearly confirms that the choice behavior of test persons tends to change over time. Nevertheless, in 7 out of 24 two-person experiments, persistent cooperation did not emerge during the experiment. Later on, we will identify reasons for this.

4.1. *Emergence of cooperation and punishment*

In order to reach the system optimum of $(-100+300)/2 = 100$ points per iteration, one individual has to leave the freeway for one iteration, which yields a reduced payoff of -100 in favor of a high payoff of $+300$ for the other individual. To be profitable also for the first individual, the other one should reciprocate this "offer" by switching to route 2, while the first individual returns to route 1. Establishing this oscillatory cooperative behavior yields 100 extra points on average. If the other individual is not cooperative, both will be back to the user equilibrium of 0 points only, and the uncooperative individual has temporarily profited from the offer by the other individual. This makes "offers" for cooperation and, therefore, the establishment of the system optimum unlikely.

Hence, the innovation of oscillatory behavior requires intentional or random changes ("trial-and-error behavior"). Moreover, the consideration of multi-period decisions is helpful. Instead of just two one-stage (i.e. one-period) alternative decisions 1 and 2, there are 2^n different n-stage (n-period) decisions. Such multi-stage strategies can be used to define higher-order games and particular kinds of supergame strategies. In the two-person second-order route choice game, for example, an encounter of the two-stage decision 12 with 21 establishes the system optimum and yields equal payoffs for everyone (see Fig. 8). Such an optimal and fair solution is not possible for one-stage decisions. Yet, the encounter of 12 with

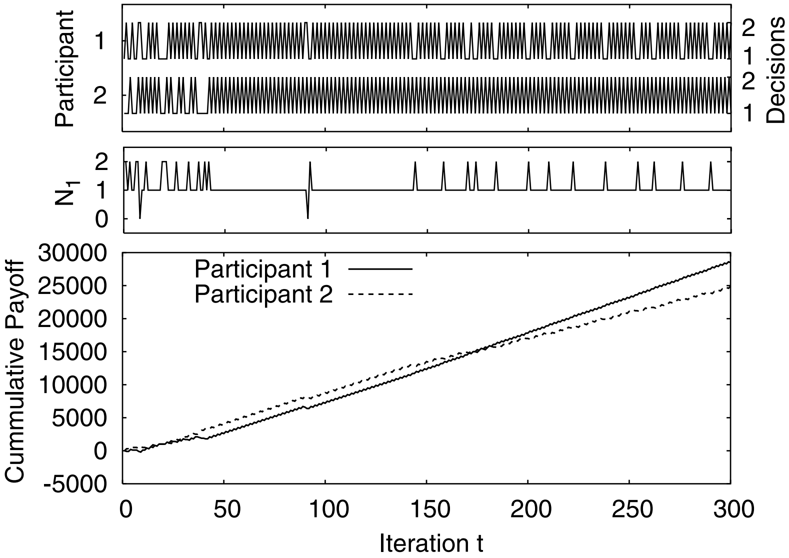
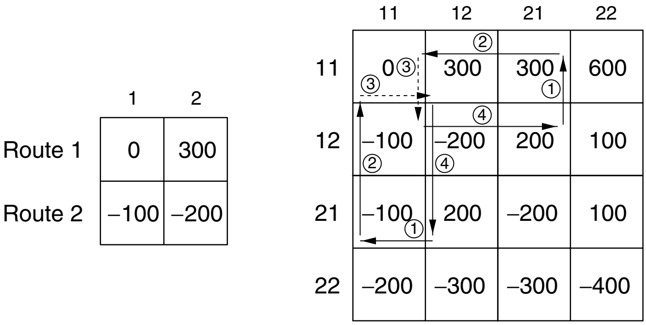


Fig. 7. Representative example for a two-person route choice experiment, in which participant 1 leaves the pattern of oscillatory cooperation temporarily in order to make additional profits. Note that participant 2 does not "punish" this selfish behavior, but continues to take routes in an alternating way. Top: Decisions of both participants over 300 iterations. Center: Number $N_1(t)$ of 1-decisions over time t. Bottom: Cumulative payoff of both players as a function of the number of iterations. The different slopes indicate an unfair outcome despite the high average payoffs of both players.

21 ("cooperative episode") is not a Nash equilibrium of the two-stage game, as an individual can increase his or her own payoff by selecting 11 (see Fig. 8). Probably for this reason, the first cooperative episodes in a repeated route choice game (i.e. encounters of 12-decisions with 21-decisions in two subsequent iterations) do often not persist (see Fig. 9). Another possible reason is that cooperative episodes may be overlooked. This problem, however, can be reduced by a feedback signal that indicates when the system optimum has been reached. For example, we have experimented with a green background color. In this setup, a cooperative episode could be recognized by a green background that appeared in two successive iterations together with two different payoff values.

The strategy of taking route 1 does not only dominate on the first day (in the first iteration). Even if a cooperative oscillatory behavior has been established, there is a temptation to leave this state, i.e. to choose route 1 several times, as this yields more than 100 points on average for the uncooperative individual at the cost of the participant continuing an alternating choice behavior (see Figs. 7 and 8). That is, the conditional changing probability $p_l(2|1, N_1 = 1; t)$ of individuals l from route 1 to route 2, when the system optimum in the previous iteration was established (i.e. $N_1 = 1$) tends to be small initially. However, oscillatory cooperation of period 2 needs $p_l(2|1, N_1 = 1; t) = 1$. The required transition in the decision behavior



	1	2
Route 1	0	300
Route 2	–100	–200

	11	12	21	22
11	0	300	300	600
12	–100	–200	200	100
21	–100	200	–200	100
22	–200	–300	–300	–400

Fig. 8. Illustration of the concept of higher-order games defined by n-stage strategies. Left: Payoff matrix $\mathbf{P} = (P_{ij})$ of the one-shot 2×2 route choice game. Right: Payoff matrix $(P^{(2)}_{(i_1 i_2),(j_1 j_2)}) = (P_{i_1 j_1} + P_{i_2 j_2})$ of the second-order route choice game defined by two-stage decisions (right). The analysis of the one-shot game (left) predicts that the user equilibrium (with both persons choosing route 1) will establish and that no single player could increase the payoff by another decision. For two-period decisions (right), the system optimum (strategy 12 meeting strategy 21) corresponds to a fair solution, but one person can increase the payoff at the cost of the other (see arrow 1), if the game is repeated. A change of the other person's decision can reduce losses and punish this egoistic behavior (arrow 2), which is likely to establish the user equilibrium with payoff 0. In order to leave this state again in favor of the system optimum, one person will have to make an "offer" at the cost of a reduced payoff (arrow 3). This offer may be due to a random or intentional change of decision. If the other person reciprocates the offer (arrow 4), the system optimum is established again. The time-averaged payoff of this cycle lies below the system optimum.

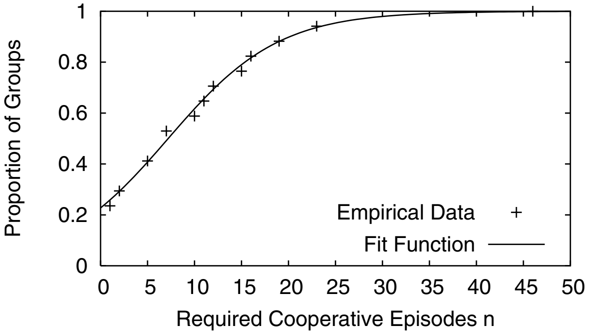


Fig. 9. Cumulative distribution of required cooperative episodes until persistent cooperation was established, given that cooperation occured during the duration of the game as in 17 out of 24 two-person experiments. The experimental data are well approximated by the logistic curve (9) with the fit parameters $c_2 = 3.4$ and $d_2 = 0.17$.

can actually be observed in our experimental data (see Fig. 10, left). With this transition, the average frequency of 1-decisions goes down to 1/2 (see Fig. 10, right). Note, however, that alternating reciprocity does not necessarily require oscillations of period 2. Longer periods are possible as well (see Fig. 11), but have occured only in a few cases (namely, 3 out of 24 cases).

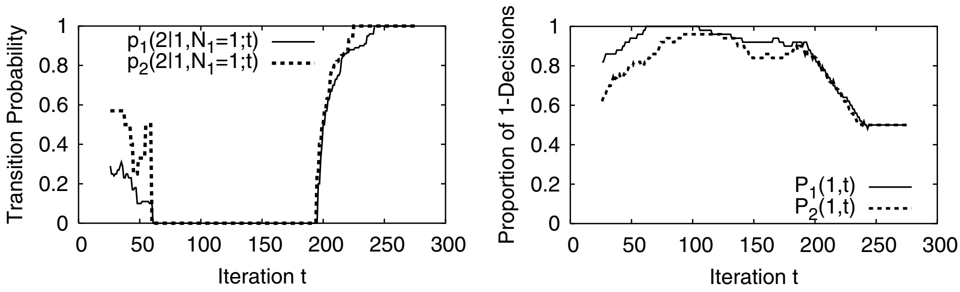


Fig. 10. Left: Conditional changing probability $p_l(2|1, N_1 = 1; t)$ of person l from route 1 (the "freeway") to route 2, when the other person has chosen route 2, averaged over a time window of 50 iterations. The transition from initially small values to 1 (for $t > 240$) is characteristic and illustrates the learning of cooperative behavior. In this particular group (cf. Fig. 4) the values started even at zero, after a transient time period of $t < 60$. Right: Proportion $P_l(1, t)$ of 1-decisions of both participants l in the two-person route choice experiment displayed in Fig. 4. While the initial proportion is often close to 1 (the user equilibrium), it reaches the value 1/2 when persistent oscillatory cooperation (the system optimum) is established.

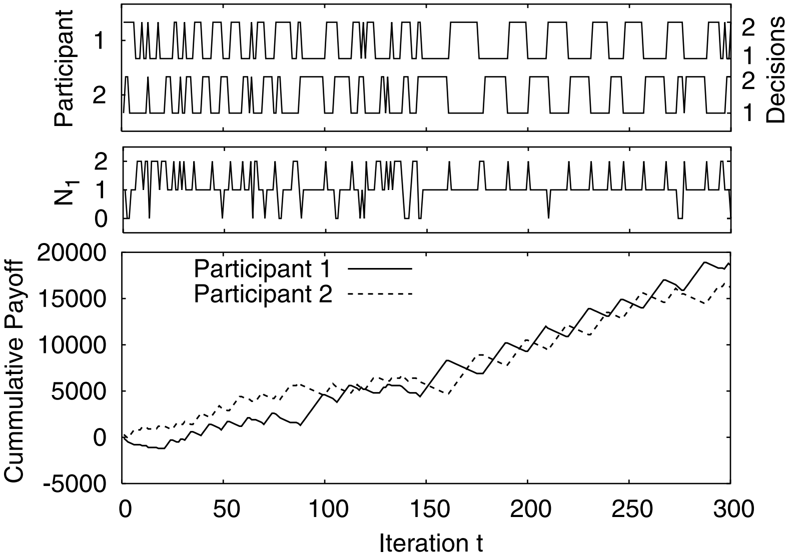


Fig. 11. Representative example for a two-person route choice experiment with phase-coordinated oscillations of long (and varying) time periods larger than 2. Top: Decisions of both participants over 300 iterations. Center: Number $N_1(t)$ of 1-decisions over time t. Bottom: Cumulative payoff of both players as a function of the number of iterations. The sawtooth-like increase in the cumulative payoff indicates gains by phase-coordinated alternations with long oscillation periods.

How does the transition to oscillatory cooperation come about? The establishment of alternating reciprocity can be supported by a suitable punishment strategy: If the other player should have selected route 2, but has chosen route 1 instead, he or she can be punished by changing to route 1 as well, since this causes an average payoff of less than 100 points for the other person (see Fig. 8). Repeated punishment of uncooperative behavior can, therefore, reinforce cooperative oscillatory behavior. However, the establishment of oscillations also requires costly "offers" by switching to route 2, which only pay back in the case of alternating reciprocity. It does not matter whether these "offers" are intentional or due to exploratory trial-and-error behavior.

Due to punishment strategies and similar reasons, persistent cooperation is often established after a number n of cooperative episodes. In the 17 of our 24 two-person experiments in which persistent cooperation was established, the cumulative distribution of required cooperative episodes could be mathematically described by the logistic curve

$$F_N(n) = 1/[1 + c_N \exp(-d_N n)] \tag{9}$$

(see Fig. 9). Note that, while we expect that this relationship is generally valid, the fit parameters c_N and d_N may depend on factors like the distribution of participant intelligence, as oscillatory behavior is apparently difficult to establish (see below).

4.2. *Preconditions for cooperation*

Let us focus on the time period before persistent oscillatory cooperation is established and denote the occurrence probability that individual l chooses alternative $i \in \{1, 2\}$ by $P_l(i)$. The quantity $p_l(j|i)$ shall represent the conditional probability of choosing j in the next iteration, if i was chosen by person l in the present one. Assuming stationarity for reasons of simplicity, we expect the relationship

$$p_l(2|1)P_l(1) = p_l(1|2)P_l(2), \tag{10}$$

i.e. the (unconditional) occurrence probability $P_l(1, 2) = p_l(2|1)P_l(1)$ of having alternative 1 in one iteration and 2 in the next agrees with the joint occurrence probability $P_l(2, 1) = p_l(1|2)P_l(2)$ of finding the opposite sequence 21 of decisions:

$$P_l(1, 2) = P_l(2, 1). \tag{11}$$

Moreover, if r_l denotes the average changing frequency of person l until persistent cooperation is established, we have the relation

$$r_l = P_l(1, 2) + P_l(2, 1). \tag{12}$$

Therefore, the probability that all N players simultaneously change their decision from one iteration to the next is $\prod_{l=1}^{N} r_l$. Note that there are 2^N such realizations of N decision changes 12 or 21, which have all the same occurrence probability because of Eq. (11). Among these, only the ones where $N/2$ players change from 1 to 2 and the other $N/2$ participants change from 2 to 1 establish cooperative episodes, given that the system optimum corresponds to an equal distribution over both alternatives. Considering that the number of different possibilities of selecting $N/2$ out of N persons is given by the binomial coefficient, the occurrence probability of cooperative events is

$$P_{\rm c} = \frac{1}{2^N} \binom{N}{N/2} \prod_{l=1}^{N} r_l \tag{13}$$

(at least in the ensemble average). Since the expected time period T until the cooperative state incidentally occurs equals the inverse of $P_{\rm c}$, we finally find the formula

$$T = \frac{1}{P_{\rm c}} = 2^N \frac{(N/2)!^2}{N!} \prod_{l=1}^{N} \frac{1}{r_l} . \tag{14}$$

This formula is well confirmed by our two-person experiments (see Fig. 12). It gives the lower bound for the expected value of the minimum number of required iterations until persistent cooperation can spontaneously emerge (if already the first cooperative episode is continued forever).

Obviously, the occurrence of oscillatory cooperation is expected to take much longer for a large number N of participants. This tendency is confirmed by our four-person experiments compared to our two-person experiments. It is also in agreement with intuition, as coordination of more people is more difficult. (Note that mean first passage or transition times in statistical phyisics tend to grow exponentially in the number N of particles as well.)

Besides the number N of participants, another critical factor for the cooperation probability are the changing frequencies r_l; they are needed for the exploration of innovative strategies, coordination and cooperation. Although the instruction of test persons would have allowed them to conclude that taking turns would be a

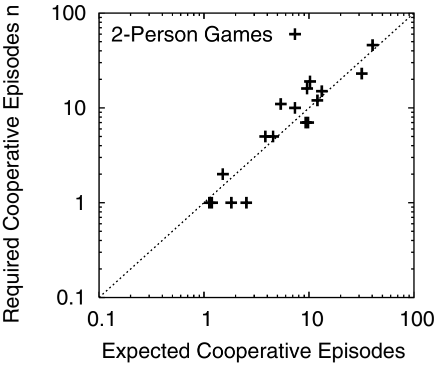


Fig. 12. Comparison of the required number of cooperative episodes y with the expected number x of cooperative episodes (approximated as occurrence time of persistent cooperation, divided by the expected time interval T until a cooperative episode occurs by chance). Note that the data points support the relationship $y = x$ and, thereby, formula (14).

good strategy, the changing frequencies r_l of some individuals was so small that cooperation within the duration of the respective experiment did not occur, in accordance with formula (14). The unwillingness of some individuals to vary their decisions is sometimes called "conservative" [7, 65, 66] or "inertial behavior" [9]. Note that, if a player never reciprocates "offers" by other players, this may discourage further "offers" and reduce the changing frequency of the other player(s) as well (see the decisions 50 through 150 of player 2 in Fig. 4).

Our experimental time series show that most individuals initially did not know a periodic decision behavior would allow them to establish the system optimum. This indicates that the required depth of strategic reasoning [19] and the related complexity of the game for an average person are already quite high, so that intelligence may matter. Compared to control experiments, the hint that the maximum average payoff of 100 points per round could be reached "by variable, situation-dependent decisions," increased the average changing frequency (by 75 percent) and with this the occurrence frequency of cooperative events. Thereby, it also increased the chance that persistent cooperation established during the duration of the experiment.

Note that successful cooperation requires not only coordination [9], but also innovation; in their first route choice game, most test persons discover the oscillatory cooperation strategy only by chance in accordance with formula (14). The changing frequency is, therefore, critical for the establishment of innovative strategies; it determines the exploratory trial-and-error behavior. In contrast, cooperation is easy when test persons *know* that the oscillatory strategy is successful; when two teams, who had successfully cooperated in two-person games, had afterwards to play a four-person game, cooperation was *always* and quickly established (see Fig. 13). In contrast, unexperienced co-players suppressed the establishment of oscillatory cooperation in four-person route choice games.

4.3. *Strategy coefficients*

In order to characterize the strategic behavior of individuals and predict their chances of cooperation, we have introduced some strategy coefficients. For this, let us introduce the following quantities, which are determined from the iterations before persistent cooperation is established:

- c_l^k = relative frequency of a *changed* subsequent decision of individual l if the payoff was negative ($k = -$), zero ($k = 0$), or positive ($k = +$).
- s_l^k = relative frequency of individual l to *stay* with the previous decision if the payoff was negative ($k = -$), zero ($k = 0$), or positive ($k = +$).

The Yule coefficient

$$Q_l = \frac{c_l^- s_l^+ - c_l^+ s_l^-}{c_l^- s_l^+ + c_l^+ s_l^-} \qquad (15)$$

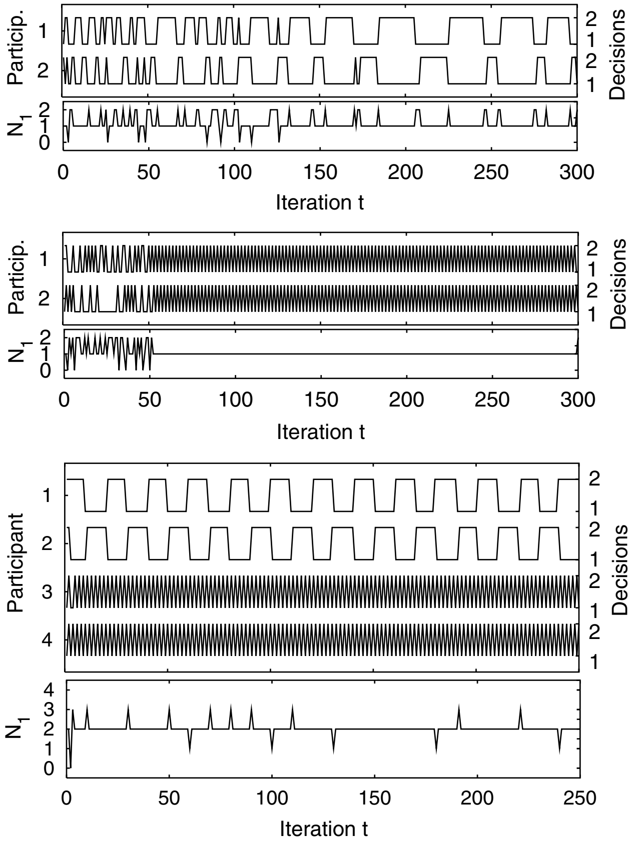


Fig. 13. Experimentally observed decision behavior when two groups involved in two-person route choice experiments afterwards played a four-person game with $C_1 = 900, D_1 = 300, C_2 = 100, D_2 = 100$. While oscillations of period 2 emerged in the second group (center), another alternating pattern corresponding to n-period decisions with $n > 2$ emerged in the first group (top). Bottom: After all persons had learnt oscillatory cooperative behavior, the four-person game just required coordination, but not the invention of a cooperative strategy. Therefore, persistent cooperation was quickly established (in contrast to four-person experiments with new participants). It is clearly visible that the test persons continued to apply similar decision strategies (bottom) as in the previous two-person experiments (top/center).

with $-1 \leq Q_l \leq 1$ was used by Schreckenberg, Selten *et al.* [65] to identify direct responders with $0.5 < Q_l \approx 1$ (who change their decision after a negative payoff and stay after a positive payoff), and contrarian responders with $-0.5 > Q_l \approx -1$ (who change their decision after a positive payoff and stay after a negative one). A random decision behavior would correspond to a value $Q_l \approx 0$. However, a problem arises if one of the variables c_l^-, s_l^+, c_l^+, or s_l^- assumes the value 0. Then, we have $Q_l \in \{-1, 1\}$, independently of the other three values. If two of the variables become zero, Q_l is sometimes even undefined. Moreover, if the values are small, the resulting conclusion is not reliable. Therefore, we prefer to use the percentage difference

$$S_l = \frac{c_l^-}{c_l^- + s_l^l} - \frac{c_l^+}{c_l^+ + s_l^+} \tag{16}$$

for the assessment of strategies. Again, we have $-1 \le S_l \le 1$. Direct responders correspond to $S_l > 0.25$ and contrarian responders to $S_l < -0.25$. For $-0.25 \le S_l \le 0.25$, the response to the previous payoff is rather random.

In addition, we have introduced the Z-coefficient

$$Z_l = \frac{c_l^0}{c_l^0 + s_l^0}, \tag{17}$$

for which we have $0 \le Z_l \le 1$. This coefficient describes the likely response of individual l to the user equilibrium. $Z_l = 0$ means that individual l does not change routes, if the user equilibrium was reached. $Z_l = 1$ implies that person l always changes, while $Z_l \approx 0.5$ indicates a random response.

Figure 14 shows the result of the two-person route choice experiments (cooperation or not) as a function of S_1 and S_2, and as a function of Z_1 and Z_2. Moreover, Figure 15 displays the result as a function of the average strategy coefficients

$$Z = \frac{1}{N} \sum_{l=1}^{N} Z_l \tag{18}$$

and

$$S = \frac{1}{N} \sum_{l=1}^{N} S_l. \tag{19}$$

Our experimental data indicate that the Z-coefficient is a good indicator for the establishment of cooperation, while the S-coefficient seems to be rather insignificant (which also applies to the Yule coefficient).

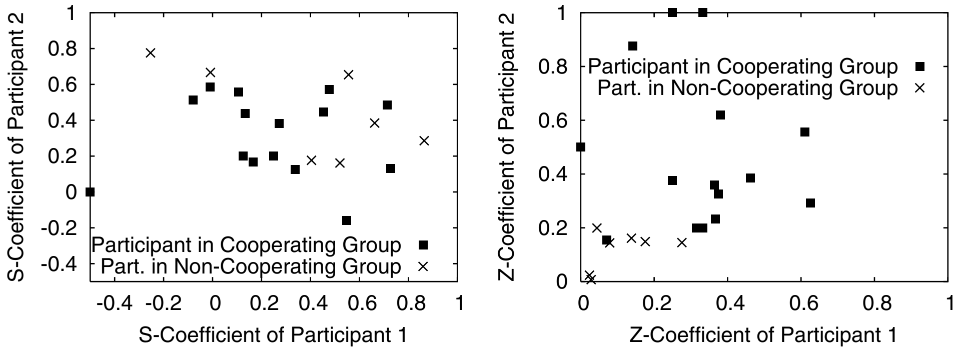


Fig. 14. Coefficients S_l and Z_l of both participants l in all 24 two-person route choice games. The values of the S-coefficients (i.e. the individual tendencies towards direct or contrarian responses) are not very significant for the establishment of persistent cooperation, while large enough values of the Z-coefficient stand for the emergence of oscillatory cooperation.

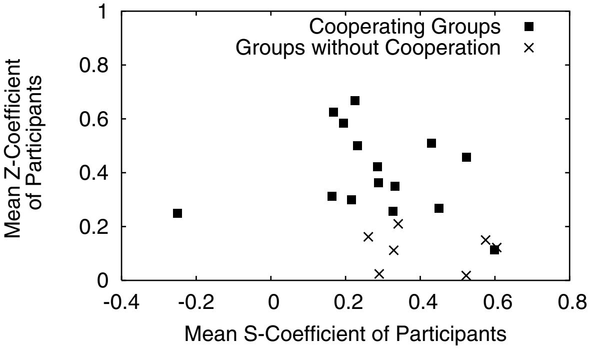


Fig. 15. S- and Z-coefficients averaged over both participants in all 24 two-person route choice games. The mainly small, but positive values of S indicate a slight tendency towards direct responses. However, the S-coefficient is barely significant for the emergence of persistent oscillations. A good indicator for their establishment is a sufficiently large Z-value.

5. Multi-Agent Simulation Model

In a first attempt, we have tried to reproduce the observed behavior in our two-person route choice experiments by game-dynamical equations [28]. We have applied these to the 2×2 route choice game and its corresponding two-, three- and four-stage higher-order games (see Sec. 4.1). Instead of describing patterns of alternating cooperation, however, the game dynamical equations predicted a preference for the dominant strategy of the one-shot game, i.e. a tendency towards choosing route 1.

The reason for this becomes understandable through Fig. 8. Selecting routes 2 and 1 in an alternating way is not a stable strategy, as the other player can get a higher payoff by choosing two times route 1 rather than responding with 1 and 2. Selecting route 1 all the time even guarantees that the own payoff is never below the one by the other player. However, when both players select route 1 and establish the related user equilibrium, no player can improve his or her payoff in the next iteration by changing the decision. Nevertheless, it is possible to improve the long-term outcome, if *both* players change their decisions, and if they do it in a coordinated way. Note, however, that a strict alternating behavior of period 2 is an optimal strategy only in infinitely repeated games, while it is unstable to perturbations in finite games.

It is known that cooperative behavior may be explained by a "shadow of the future" [2, 3], but it can also be established by a "shadow of the past" [40], i.e. experience-based learning. This will be the approach of the multi-agent simulation model proposed in this section. As indicated before, the emergence of phase-coordinated strategic alternation (rather than a statistically independent application of mixed strategies) requires an almost deterministic behavior (see Fig. 16). Nevertheless, some weak stochasticity is needed for the establishment of asymmetric cooperation, both for the exploration of innovative strategies and for phase coordination. Therefore, we propose the following reinforcement learning model, which could be called a generalized win-stay, lose-shift strategy [50, 54].

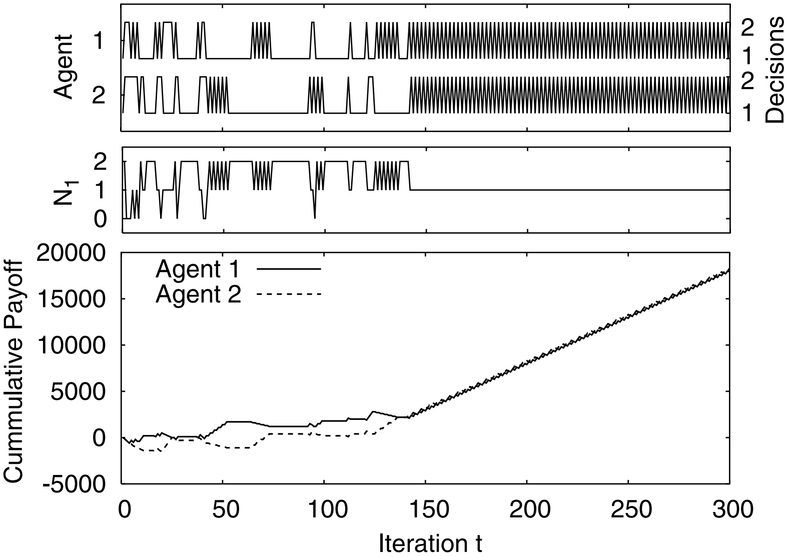


Fig. 16. Representative example for a two-person route choice simulation based on our proposed multi-agent reinforcement learning model with $P_{\text{av}}^{\text{max}} = 100$ and $P_{\text{av}}^{\text{min}} = -200$. The parameter ν_l^1 has been set to 0.25. The other model parameters are specified in the text. Top: Decisions of both agents over 300 iterations. Center: Number $N_1(t)$ of 1-decisions over time t. Bottom: Cumulative payoff of both agents as a function of the number of iterations. The emergence of oscillatory cooperation is comparable with the experimental data displayed in Fig. 4.

Let us presuppose that an individual approximately memorizes or has a good feeling of how well he or she has performed on average in the last n_l iterations and since he or she has last responded with decision j to the situation (i, N_1). In our success- and history-dependent model of individual decision behavior, $p_l(j|i, N_1; t)$ denotes agent l's conditional probability of taking decision j at time $t + 1$, when i was selected at time t and $N_1(t)$ agents had chosen alternative 1. Assuming that p_l is either 0 or 1, $p_l(j|i, N_1; t)$ has the meaning of a deterministic response strategy: $p_l(j|i, N_1; t) = 1$ implies that individual l will respond at time $t + 1$ with the decision j to the situation (i, N_1) at time t.

Our reinforcement learning strategy can be formulated as follows. The response strategy $p_l(j|i, N_1, t)$ is switched with probability $q_l > 0$, if the average individual payoff since the last comparable situation with $i(t') = i(t)$ and $N_1(t') = N_1(t)$ at time $t' < t$ is less than the average individual payoff $\bar{P}_l(t)$ during the last n_l iterations. In other words, if the time-dependent aspiration level $\bar{P}_l(t)$ [40, 54] is not reached by the agent's average payoff since his or her last comparable decision, the individual is assumed to substitute the response strategy $p_l(j|i, N_1; t)$ by

$$p_l(j|i, N_1; t + 1) = 1 - p_l(j|i, N_1; t) \tag{20}$$

with probability q_l. The replacement of dissatisfactory strategies orients at historical long-term profits (namely, during the time period $[t', t]$). Thereby, it avoids short-sighted changes after temporary losses. Moreover, it does not assume a comparison of the performance of the actually applied strategy with hypothetical ones as in most evolutionary models. A readiness for altruistic decisions is also not required, while exploratory behavior ("trial and error") is necessary. In order to reflect this, the decision behavior is randomly switched from $p_l(j|i, N_1; t+1)$ to $1 - p_l(j|i, N_1; t+1)$ with probability

$$\nu_l(t) = \max\left(\nu_l^0, \nu_l^1 \frac{P_{\text{av}}^{\max} - \bar{P}_l(t)}{P_{\text{av}}^{\max} - P_{\text{av}}^{\min}}\right) \ll 1. \tag{21}$$

Herein, $P_{\text{av}}^{\min}$ and $P_{\text{av}}^{\max}$ denote the minimum and maximum average payoff of all N agents (simulated players). The parameter ν_l^1 reflects the mutation frequency for $\bar{P}_l(t) = P_{\text{av}}^{\min}$, while the mutation frequency is assumed to be $\nu_l^0 \leq \nu_l^1$ when the time-averaged payoff $\bar{P}_l$ reaches the system optimum $\bar{P}_{\text{av}}^{\max}$.

In our simulations, no emergent cooperation is found for $\nu_l^0 = \nu_l^1 = 0$. $\nu_l^0 > 0$ or odd values of n_l may produce intermittent breakdowns of cooperation. A small, but finite value of ν_l^1 is important to find a transition to persistent cooperation. Therefore, we have used the parameter value $\nu_l^1 = 0.25$, while the simplest possible specification has been chosen for the other parameters, namely $\nu_l^0 = 0, q_l = 1$, and $n_l = 2$.

The initial conditions for the simulation of the route choice game were specified in accordance with the dominant strategy of the one-shot game, i.e. $P_l(1,0) = 1$ (everyone tends to choose the freeway initially), $p_l(2|1, N_1; 0) = 0$ (it is not attractive to change from the freeway to the side road) and $p_l(1|2, N_1; 0) = 1$ (it is tempting to change from the side road to the freeway). Interestingly enough, agents learnt to acquire the response strategy $p_l(2|1, N_1 = 1; t) = 1$ in the course of time, which established oscillatory cooperation with higher profits (see Figs. 16 and 17).

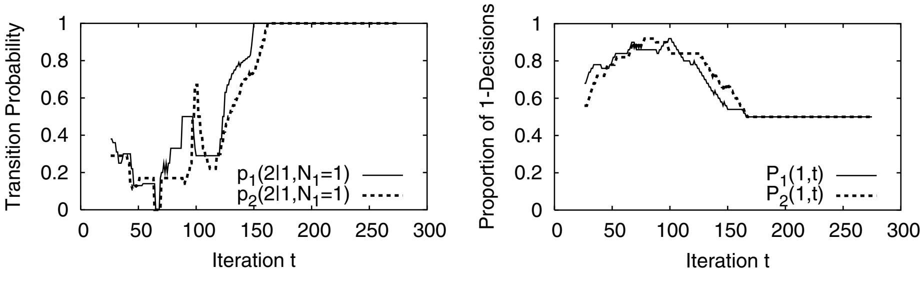


Fig. 17. Left: Conditional changing probability $p_l(2|1, N_1 = 1; t)$ of agent l from route 1 (the "freeway") to route 2, when the other agent has chosen route 2, averaged over a time window of 50 iterations. The transition from small values to 1 for the computer simulation displayed in Fig. 16 is characteristic and illustrates the learning of cooperative behavior. Right: Proportion $P_l(1,t)$ of 1-decisions of both participants l in the two-person route choice experiment displayed in Fig. 16. While the initial proportion is often close to 1 (the user equilibrium), it reaches the value 1/2 when persistent oscillatory cooperation (the system optimum) is established. The simulation results are compatible with the essential features of the experimental data (see, for example, Fig. 10).

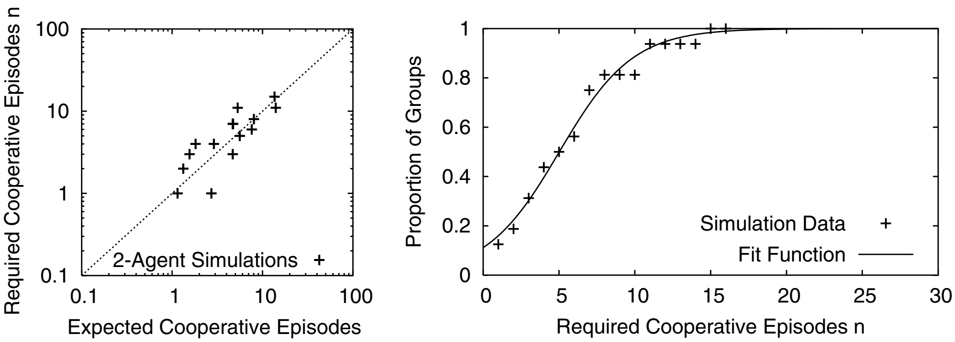


Fig. 18. Left: Comparison of the required number of cooperative episodes with the expected number of cooperative episodes in our multi-agent simulation of decisions in the route choice game. Note that the data points support formula (14). Right: Cumulative distribution of required cooperative episodes until persistent cooperation is established in our two-person route choice simulations, using the simplest specification of model parameters (not calibrated). The simulation data are well approximated by the logistic curve (9) with the fit parameters $c_2 = 7.9$ and $d_2 = 0.41$.

Note that the above described reinforcement learning model [40] responds only to the own previous experience [13]. Despite its simplicity (e.g. the neglect of more powerful, but probably less realistic k-move memories [11]), our "multi-agent" simulations reproduce the emergence of asymmetric reciprocity of two or more players, if an oscillatory strategy of period 2 can establish the system optimum. This raises the question why previous experiments of the N-person route choice game [27, 63] have observed a clear tendency towards the Wardrop equilibrium [71] with $P_1(N_1) = P_2(N_2)$ rather than phase-coordinated oscillations? It turns out that the payoff values must be suitably chosen [see Eq. (8)] and that several hundred repetitions are needed. In fact, the expected time interval T until a cooperative episode among $N = N_1 + N_2$ participants occurs in our simulations by chance is well described by formula (14); see Fig. 18. The empirically observed transition in the decision behavior displayed in Fig. 10 is qualitatively reproduced by our computer simulations as well (see Fig. 17). The same applies to the frequency distribution of the average payoff values (compare Fig. 19 with Fig. 6) or to the number of expected and required cooperative episodes (compare Fig. 18 with Figs. 9 and 12).

5.1. *Simultaneous and alternating cooperation in the Prisoner's Dilemma*

Let us finally simulate the dynamic behavior in the two different variants of the Prisoner's Dilemma indicated in Figs. 3(b) and (c) with the above experience-based reinforcement learning model. Again, we will assume $P_{11} = 0$ and $P_{22} = -200$. According to Eq. (8), a simultaneous, symmetrical form of cooperation is expected for $P_{12} = -300$ and $P_{21} = 100$, while an alternating, asymmetric cooperation is expected for $P_{12} = -300$ and $P_{21} = 500$. Figure 20 shows simulation results for the two different cases of the Prisoner's Dilemma and confirms the two predicted forms

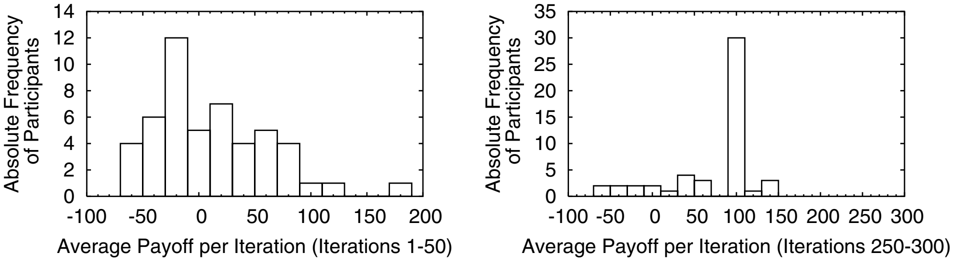


Fig. 19. Frequency distributions of the average payoffs in our computer simulations of the two-person route choice game. Left: Distribution during the first 50 iterations. Right: Distribution between iterations 250 and 300. Our simulation results are compatible with the experimental data displayed in Fig. 6.

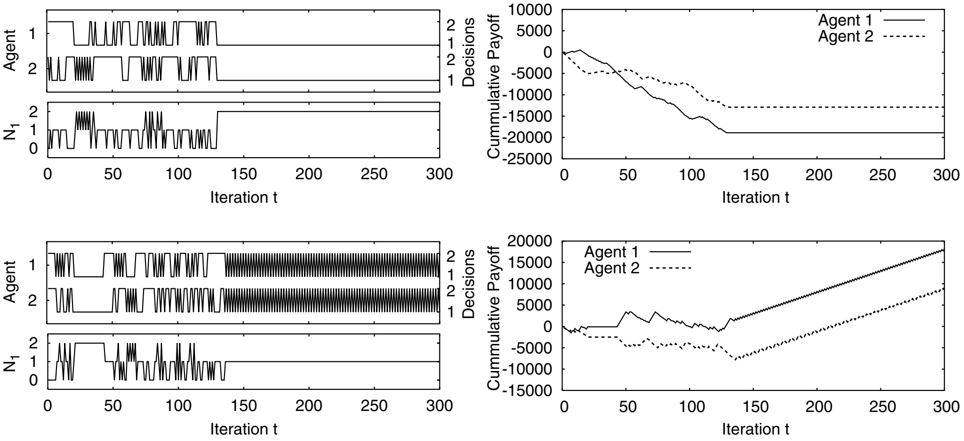


Fig. 20. Representative examples for computer simulations of the two different forms of the Prisoner's Dilemma specified in Figs. 3(b) and (c). The parameter ν_l^1 has been set to 0.25, while the other model parameters are specified in the text. Top: Emergence of simultaneous, symmetrical cooperation, where decision 2 corresponds to defection and decision 1 to cooperation. The system optimum corresponds to $P_{\rm av}^{\rm max} = 0$ payoff points, and the minimum payoff to $P_{\rm av}^{\rm min} = -200$. Bottom: Emergence of alternating, asymmetric cooperation with $P_{\rm av}^{\rm max} = 100$ and $P_{\rm av}^{\rm min} = -200$. Left: Time series of the agents' decisions and the number $N_1(t)$ of 1-decisions. Right: Cumulative payoffs as a function of time t.

of cooperation. Again, we varied only the parameter ν_l^1, while we chose the simplest possible specification of the other parameters $\nu_l^0 = 0, q_l = 1$, and $n_l = 2$. The initial conditions were specified in accordance with the expected non-cooperative outcome of the one-shot game, i.e. $P_l(1,0) = 0$ (everyone defects in the beginning), $p_l(2|2, N_1; 0) = 0$ (it is tempting to continue defecting), $p_l(1|1, N_1 = 1; 0) = 0$ (it is unfavorable to be the only cooperative player), and $p_l(1|1, N_1 = 2; 0) = 1$ (it is good to continue cooperating, if the other player cooperates). In the course of time, agents learn to acquire the response strategy $p_l(2|2, N_1 = 0; t) = 0$ when simultaneous cooperation evolves, but $p_l(2|2, N_1 = 1; t) = 0$ when alternating cooperation is established.

6. Summary, Discussion, and Outlook

In this paper, we have investigated the N-person day-to-day route-choice game. This special congestion game has not been thoroughly studied before in the case of small groups, where the system optimum can considerably differ from the user equilibrium. The two-person route choice game gives a meaning to a previously uncommon repeated symmetrical 2×2 game and shows a transition from the dominating strategy of the one-shot game to coherent oscillations, if $P_{12} + P_{21} > 2P_{11}$. However, a detailed analysis of laboratory experiments with humans reveals that the establishment of this phase-coordinated alternating reciprocity, which is expected to occur in other 2×2 games as well, is quite complex. It needs either strategic experience or the invention of a suitable strategy. Such an innovation is driven by the potential gains in the average payoffs of all participants and seems to be based on exploratory trial-and-error behavior. If the changing frequency of one or several players is too low, no cooperation is established in a long time. Moreover, the emergence of cooperation requires certain kinds of strategies, which can be characterized by the Z-coefficient (18). These strategies can be acquired by means of reinforcement learning, i.e. by keeping response patterns which have turned out to be better than average, while worse response patterns are being replaced. The punishment of uncooperative behavior can help to enforce cooperation. Note, however, that punishment in groups of $N > 2$ persons is difficult, as it is hard to target the uncooperative person, and punishment affects everyone. Nevertheless, computer simulations and additional experiments indicate that oscillatory cooperation can still emerge in route choice games with more than two players after a long time period (rarely within 300 iterations) (see Fig. 21).

Altogether, spontaneous cooperation takes a long time. It is, therefore, sensitive to changing conditions reflected by time-dependent payoff parameters. As a consequence, emergent cooperation is unlikely to appear in real traffic systems. This is the reason why the Wardrop equilibrium tends to occur. However, cooperation could be rapidly established by means of advanced traveller information systems (ATIS) [8, 14, 30, 37, 41, 63, 70, 73], which would avoid the slow learning process described by Eq. (14). Moreover, while we do not recommend conventional congestion charges, a charge for *unfair* usage patterns would support the compliance with individual route choice recommendations. It would supplement the inefficient individual punishment mechanism.

Different road pricing schemes have been proposed, each of which has its own advantages and disadvantages or side effects. Congestion charges, for example, could discourage the taking of congested routes, which is actually required to reach minimum *average* travel times. Conventional tolls and road pricing may reduce the trip frequency due to budget constraints, which potentially interferes with economic growth and fair chances for everyone's mobility.

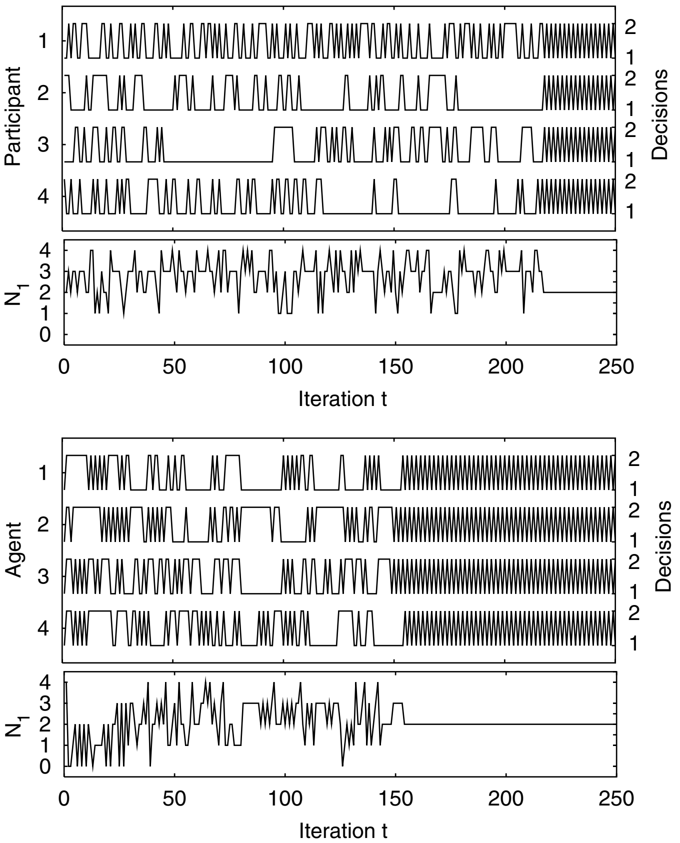


Fig. 21. Emergence of phase-coordinated oscillatory behavior in the four-person route choice game with the parameters specified in Fig. 13. Top: Experimental data of the decisions of four unexperienced participants over 250 iterations. Bottom: Computer simulation with the reinforcment learning model.

In order to activate capacity reserves, we therefore propose an automated route guidance system based on the following principles. After specification of their destination, drivers should get individual (and, on average, fair) route choice recommendations in agreement with the traffic situation and the route choice proportions required to reach the system optimum. If an individual selects a faster route instead of the recommended route that should be used, he/she will have to pay an amount proportional to the decrease in the overall inverse travel time compared to the system optimum. Moreover, drivers not in a hurry should be encouraged to take the slower route i by receiving the amount of money corresponding to the related increase in the overall inverse travel time. Altogether, such an ATIS could support the system optimum while allowing for some flexibility in route choice. Moreover, the fair usage pattern would be cost-neutral for everyone, i.e. traffic flows of potential economic relevance would not be suppressed by extra costs.

In systems with many similar routing decisions, a Pareto optimum characterized by asymmetric alternating cooperation may even emerge spontaneously. This could help to enhance the routing in data networks [72] and generally to resolve Braess-like paradoxes in networks [17].

Finally, it cannot be emphasized enough that taking turns is a promising strategy to distribute scarce resources in a fair and optimal way. It could be applied to a huge number of real-life situations due to the relevance for many strategical conflicts, including Leader, the Battle of the Sexes, and variants of Route Choice, Deadlock, Chicken, and the Prisoner's Dilemma. The same applies to their N-person generalizations, in particular social dilemmas [23, 25, 40]. It will also be interesting to find out whether and where metabolic pathways, biological supply networks, or information flows in neuronal and immune systems use alternating strategies to avoid the wasting of costly resources.

Acknowledgments

D. Helbing is grateful for the warm hospitality of the Santa Fe Institute, where the Social Scaling Working Group Meeting in August 2003 inspired many ideas in this paper. The results will be presented during the workshop on "Collectives Formation and Specialization in Biological and Social Systems" in Santa Fe (April 20–22, 2005).

References

[1] Arthur, W. B., Inductive reasoning and bounded rationality, *Am. Econ. Rev.* **84**, 406–411 (1994).

[2] Axelrod, R. and Dion, D., The further evolution of cooperation, *Science* **242**, 1385–1390 (1988).

[3] Axelrod, R. and Hamilton, W. D., The evolution of cooperation, *Science* **211**, 1390–1396 (1981).

[4] Beckmann, M., McGuire, C. B. and Winsten, C. B., *Studies in the Economics of Transportation* (Yale University Press, New Haven, 1956).

[5] Binmore, K. G., Evolutionary stability in repeated games played by finite automata, *J. Econ. Theory* **57**, 278–305 (1992).

[6] Binmore, K., *Fun and Games: A Text on Game Theory* (Heath, Lexington, MA, 1992), pp. 373–377.

[7] Bohnsack, U., Uni DuE: Studie SURVIVE gibt Einblicke in das Wesen des Autofahrers, Press release by *Informationsdienst Wissenschaft* (January 21, 2005).

[8] Bonsall, P. W. and Perry, T., Using an interactive route-choice simulator to investigate driver's compliance with route guidance information, *Transpn. Res. Rec.* **1306**, 59–68 (1991).

[9] Bottazzi, G. and Devetag, G., Coordination and self-organization in minority games: Experimental evidence, Working Paper 2002/09, Sant'Anna School of Advances Studies, May, 2002.

[10] Braess, D., Über ein Paradoxon der Verkehrsplanung [A paradox of traffic assignment problems], *Unternehmensforschung* **12**, 258–268 (1968). For about 100 related references see http://homepage.ruhr-uni-bochum.de/Dietrich.Braess/#paradox

[11] Browning, L. and Colman, A. M., Evolution of coordinated alternating reciprocity in repeated dyadic games, *J. Theor. Biol.* **229**, 549–557 (2004).

[12] Camerer, C. F., *Behavioral Game Theory: Experiments on Strategic Interaction* (Princeton University Press, Princeton, 2003).

[13] Camerer, C. F., Ho, T.-H. and Chong, J.-K., Sophisticated experience-weighted attraction learning and strategic teaching in repeated games, *J. Econ. Theory* **104**, 137–188 (2002).

[14] Cetin, N., Nagel, K., Raney, B. and Voellmy, A., Large scale multi-agent transportation simulations, *Computer Physics Communications* **147**(1–2), 559–564 (2002).

[15] Challet, D. and Marsili, M., Relevance of memory in minority games, *Phys. Rev.* **E62**, 1862–1868 (2000).

[16] Challet, D. and Zhang, Y.-C., Emergence of cooperation and organization in an evolutionary game, *Physica* **A246**, 407–418 (1997).

[17] Cohen, J. E. and Horowitz, P., Paradoxial behaviour of mechanical and electrical networks, *Nature* **352**, 699–701 (1991).

[18] Colman, A. M., *Game Theory and its Applications in the Social and Biological Sciences*, 2nd edn. (Butterworth-Heinemann, Oxford, 1995).

[19] Colman, A. M., Depth of strategic reasoning in games, *Trends Cogn. Sci.* **7**(1), 2–4 (2003).

[20] Crowley, P. H., Dangerous games and the emergence of social structure: evolving memory-based strategies for the generalized hawk-dove game, *Behav. Ecol.* **12**, 753–760 (2001).

[21] Eriksson, A. and Lindgren, K., Cooperation in an unpredictable environment, in *Proc. Artificial Life VIII* (eds. Standish, R. K., Bedau, M. A. and Abbass, H. A.) (MIT Press, Sidney, 2002), pp. 394–399; and poster available at http://frt.fy.chalmers.se/cs/people/eriksson.html

[22] Garcia, C. B. and Zangwill, W. I., *Pathways to Solutions, Fixed Points, and Equilibria* (Prentice Hall, New York, 1981).

[23] Glance, N. S. and Huberman, B. A., The outbreak of cooperation, *J. Math. Soc.* **17**(4), 281–302 (1993).

[24] Greenshield, B. D., A study of traffic capacity, in *Proc. Highway Research Board*, Vol. 14 (Highway Research Board, Washington, D. C., 1935), pp. 448–477.

[25] Hardin, G., The tragedy of the commons, *Science* **162**, 1243–1248 (1968).

[26] Helbing, D., Dynamic decision behavior and optimal guidance through information services: Models and experiments, in *Human Behaviour and Traffic Networks*, eds. Schreckenberg, M. and Selten, R. (Springer, Berlin, 2004), pp. 47–95.

[27] Helbing, D., Schönhof, M. and Kern, D., Volatile decision dynamics: Experiments, stochastic description, intermittency control, and traffic optimization, *New J. Phys.* **4**, 33.1–33.16 (2002).

[28] Hofbauer, J. and Sigmund, K., *The Theory of Evolution and Dynamical Systems* (Cambridge University Press, Cambridge, 1988).

[29] Hogg, T. and Huberman, B. A., Controlling chaos in distributed systems, *IEEE Trans. Syst. Man Cy.* **21**(6), 1325–1333 (1991).

[30] Hu, T.-Y. and Mahmassani, H. S., Day-to-day evolution of network flows under real-time information and reactive signal control, *Transport. Res.* **C5**(1), 51–69 (1997).

[31] Iwanaga, S. and Namatame, A., The complexity of collective decision, *Nonlinear Dynam. Psychol. Life Sci.* **6**(2), 137–158 (2002).

[32] Kagel, J. H. and Roth, A. E. (eds.), *The Handbook of Experimental Economics* (Princeton University, Princeton, NJ, 1995).

[33] Kephart, J. O., Hogg, T. and Huberman, B. A., Dynamics of computational ecosystems, *Phys. Rev.* **A40**(1), 404–421 (1989).

[34] Klügl, F. and Bazzan, A. L. C., Route decision behaviour in a commuting scenario: Simple heuristics adaptation and effect of traffic forecast, *JASSS* **7**(1), Jan. (2004).

[35] Korilis, Y. A., Lazar, A. A. and Orda, A., Avoiding the Braess-paradox in non-cooperative networks, *J. Appl. Prob.* **36**, 211–222 (1999).

[36] Laureti, P., Ruch, P., Wakeling, J. and Zhang, Y.-C., The interactive minority game: A Web-based investigation of human market interactions, *Physica* **A331**, 651–659 (2004).

[37] Lee, K., Hui, P. M., Wang, B. H. and Johnson, N. F., Effects of announcing global information in a two-route traffic flow model, *J. Phys. Soc. Japan* **70**, 3507–3510 (2001).

[38] Lo, T. S., Chan, H, Y., Hui, P. M. and Johnson, N. F., Theory of networked minority games based on strategy pattern dynamics, *Phys. Rev.* **E70**, 056102 (2004).

[39] Lo, T. S., Hui, P. M. and Johnson, N. F., Theory of the evolutionary minority game, *Phys. Rev.* **E62**, 4393–4396 (2000).

[40] Macy, M. W. and Flache, A., Learning dynamics in social dilemmas, in *Proc. National Academy of Sciences USA*, Vol. 99, Suppl. 3 (2002), pp. 7229–7236.

[41] Mahmassani, H. S. and Jou, R. C., Transferring insights into commuter behavior dynamics from laboratory experiments to field surveys, *Transport. Res.* **A34**, 243–260 (2000).

[42] Mansilla, R., Algorithmic complexity in the minority game, *Phys. Rev.* **E62**, 4553–4557 (2000).

[43] Marsili, M., Mulet, R., Ricci-Tersenghi, F. and Zecchina, R., Learning to coordinate in a complex and nonstationary world, *Phys. Rev. Lett.* **87**, 208701 (2001).

[44] McNamara, J. M., Barta, Z. and Houston, A. I., Variation in behaviour promotes cooperation in the prisoner's dilemma game, *Nature* **428**, 745–748 (2004).

[45] Michor, F. and Nowak, M. A., The good, the bad and the lonely, *Nature* **419**, 677–679 (2002).

[46] Milinski, M., Semmann, D. and Krambeck, H.-J., Reputation helps solve the 'tragedy of the commons', *Nature* **415**, 424–426 (2002).

[47] Monderer, D. and Shapley, L. S., Fictitious play property for games with identical interests, *J. Econ. Theory* **1**, 258–265 (1996).

[48] Monderer, D. and Shapley, L. S., Potential games, *Games Econ. Behav.* **14**, 124–143 (1996).

[49] Nowak, M. A., Sasaki, A., Taylor, C. and Fudenberg, D., Emergence of cooperation and evolutionary stability in finite populations, *Nature* **428**, 646–650 (2004).

[50] Novak, M. and Sigmund, K., A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game, *Nature* **364**, 56–58 (1993).

[51] Nowak, M. A. and Sigmund, K., The alternating prisoner's dilemma, *J. Theor. Biol.* **168**, 219–226 (1994).

[52] Nowak, M. A. and Sigmund, K., Evolution of indirect reciprocity by image scoring, *Nature* **393**, 573–577 (1998).

[53] Pigou, A. C., *The Economics of Welfare* (Macmillan, London, 1920).

[54] Posch, M., Win-stay, Lose-shift strategies for repeated games — Memory length, aspiration levels and noise, *J. Theor. Biol.* **198**, 183–195 (1999).

[55] Queller, D. C., Kinship is relative, *Nature* **430**, 975–976 (2004).

[56] Rapoport, A., Exploiter, leader, hero, and martyr: The four archtypes of the 2 × 2 game, *Behav. Sci.* **12**, 81–84 (1967).

[57] Rapoport, A. and Guyer, M., A taxonomy of 2×2 games, *Gen. Systems* **11**, 203–214 (1966).

[58] Reddy, P. D. V. G. *et al.*, Design of an artificial simulator for analyzing route choice behavior in the presence of information system, *J. Math. Comp. Mod.* **22**, 119–147 (1995).

[59] Riolo, R. L., Cohen, M. D. and Axelrod, R., Evolution of cooperation without reciprocity, *Nature* **414**, 441–443 (2001).

[60] Rosenthal, R. W., A class of games possessing pure-strategy Nash equilibria, *Int. J. Game Theory* **2**, 65–67 (1973).

[61] Roughgarden, T. and Tardos, E., How bad is selfish routing?, *J. ACM* **49**(2), 236–259 (2002).

[62] Schelling, T. C., *Micromotives and Macrobehavior* (WW Norton and Co, New York, 1978), pp. 224–231, 237.

[63] Schreckenberg, M. and Selten, R. (eds.), *Human Behaviour and Traffic Networks* (Springer, Berlin, 2004).

[64] Schweitzer, F., Behera, L. and Mühlenbein, H., Evolution of cooperation in a spatial prisoner's dilemma, *Adv. Complex Syst.* **5**(2/3), 269–299 (2002).

[65] Selten, R. *et al.*, Experimental investigation of day-to-day route-choice behaviour and network simulations of autobahn traffic in North Rhine-Westphalia, in *Human Behaviour and Traffic Networks*, eds. Schreckenberg, M. and Selten, R. (Springer, Berlin, 2004), pp. 1–21.

[66] Selten, R., Schreckenberg, M., Pitz, T., Chmura, T. and Kube, S., Experiments and simulations on day-to-day route choice-behaviour, see http://papers.ssrn.com/sol3/papers.cfm?abstract_id=393841

[67] Semmann, D., Krambeck, H.-J. and Milinski, M., Volunteering leads to rock-paper-scissors dynamics in a public goods game, *Nature* **425**, 390–393 (2003).

[68] Spirakis, P., Algorithmic aspects of congestion games, Invited talk at the *11th Colloquium on Structural Information and Communication Complexity*, Smolenice Castle, Slovakia, June 21–23, 2004.

[69] Szabó, G. and Hauert, C., Phase transitions and volunteering in spatial public goods games, *Phys. Rev. Lett.* **89**, 118101 (2002).

[70] Wahle, J., Bazzan, A. L. C., Klügl, F. and Schreckenberg, M., Decision dynamics in a traffic scenario, *Physica* **A287**, 669–681 (2000).

[71] Wardrop, J. G., Some theoretical aspects of road traffic research, in *Proc. the Institution of Civil Engineers II*, Vol. 1, 1952, pp. 325–378.

[72] Wolpert, D. H. and Tumer, K., Collective intelligence, data routing and Braess' paradox, *J. Artif. Int. Res.* **16**, 359–387 (2002).

[73] Yamashita, T., Izumi, K. and Kurumatani, K., Effect of using route information sharing to reduce traffic congestion, *Lect. Notes Comput. Sci.* **3012**, 86–104 (2004).

[74] Yuan, B. and Chen, K., Evolutionary dynamics and the phase structure of the minority game, *Phys. Rev.* **E69**, 067106 (2004).