Learning dynamics in social dilemmas

Michael W. Macy*† and Andreas Flache‡

*Department of Sociology, Cornell University, Ithaca, NY 14853; and ‡Interuniversity Center for Social Science Theory and Methodology, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands

The Nash equilibrium, the main solution concept in analytical game theory, cannot make precise predictions about the outcome of repeated mixed-motive games. Nor can it tell us much about the dynamics by which a population of players moves from one equilibrium to another. These limitations, along with concerns about the cognitive demands of forward-looking rationality, have motivated efforts to explore backward-looking alternatives to analytical game theory. Most of the effort has been invested in evolutionary models of population dynamics. We shift attention to a learning-theoretic alternative. Computational experiments with adaptive agents identify a fundamental solution concept for social dilemmas––stochastic collusion––based on a random walk from a self-limiting noncooperative equilibrium into a self-reinforcing cooperative equilibrium. However, we show that this solution is viable only within a narrow range of aspiration levels. Below the lower threshold, agents are pulled into a deficient equilibrium that is a stronger attractor than mutual cooperation. Above the upper threshold, agents are dissatisfied with mutual cooperation. Aspirations that adapt with experience (producing habituation to stimuli) do not gravitate into the window of viability; rather, they are the worst of both worlds. Habituation destabilizes cooperation and stabilizes defection. Results from the two-person problem suggest that applications to multiplex and embedded relationships will yield unexpected insights into the global dynamics of cooperation in social dilemmas.

Learning Theory and Social Dilemmas

Why are some communities, organizations, and relationships characterized by cooperation, trust, and solidarity whereas others are fraught with corruption, discord, and fear? Viewed from the top down, the answer centers on hierarchical institutions that coordinate and regulate individual behavior to conform to the functional requirements of the system at a higher level of aggregation. These overarching institutions are backed by central authorities and supported by cultural norms to which individuals are socialized to conform.

Agent-based models assume a very different world, where decision making is distributed and global order self-organizes out of multiplex local interactions among autonomous interdependent actors. From this "bottom-up" perspective, socially efficient outcomes are possible but problematic. In striking contrast to the image of a functionally integrated society, a bottom-up world is one that is likely to be riddled with "social dilemmas" (1) in which decisions that make sense to each individual can aggregate into outcomes in which everyone suffers. Although some relationships proceed cooperatively, others descend into spiraling conflict, corruption, distrust, or recrimination that all parties would prefer to avoid. Can agents find their way out, and if so, how do they do it?

At the most elementary level, social dilemmas can be formalized as a mixed-motive two-person game with two choices––cooperate (be honest, truthful, helpful, etc.) or defect (lie, cheat, steal, and the like). These two choices intersect at four possible outcomes, each with a designated payoff. R (reward) and P (punishment) are the payoffs for mutual cooperation and defection, respectively, whereas S (sucker) and T (temptation) are the payoffs for cooperation by one player and defection by the other. Using these payoffs, we can define a social dilemma as any ordering of these four payoffs such that the following four conditions are satisfied:

1. $R > P$: players prefer mutual cooperation (CC) over mutual defection (DD).
2. $R > S$: players prefer mutual cooperation over unilateral cooperation (CD).
3. $2R > T + S$: players prefer mutual cooperation over an equal probability of unilateral cooperation and defection.
4. $T > R$: players prefer unilateral defection (DC) to mutual cooperation (greed)

or $P > S$: players prefer mutual defection to unilateral cooperation (fear).

These four conditions create the tension in social dilemmas between individual and collective interests. This tension is apparent when the preferred choices of each player intersect at the outcome that both would prefer to avoid: mutual defection. This outcome is Pareto deficient in all social dilemmas, that is, there is always another outcome–mutual cooperation–that is preferred by everyone (given $R > P$). Yet mutual cooperation may be undermined by the temptation to cheat (if $T > R$) or by the fear of being cheated (if $P > S$) or by both. In the game of Stag Hunt ($R > T > P > S$) the problem is fear but not greed, and in the game of Chicken ($T > R > S > P$) the problem is greed but not fear. The problem is most challenging when both fear and greed are present, that is, when $T > R$ and $P > S$. Given the assumption that $R > P$, there is only one way this can happen, if $T > R > P > S$, the celebrated game of Prisoner's Dilemma (PD).

If the game is played only once, analytical game theory can identify precise solutions to social dilemmas based on the Nash equilibrium, a set of pure or mixed strategies from which no player has an incentive to unilaterally deviate. The Nash solution is discouraging: with one exception (mutual cooperation in Stag Hunt), all pure- and mixed-strategy Nash equilibria, across all possible one-shot social dilemma games, are Rawls deficient (2), that is, there is at least one other outcome that both players prefer, assuming they must choose under a "veil of ignorance," before knowing who will receive the higher payoff.§

If the game is repeated, the result is discouraging for game theory but not for the players. The good news for the players is

---

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Adaptive Agents, Intelligence, and Emergent Human Organization: Capturing Complexity through Agent-Based Modeling," held October 4–6, 2001, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: PD, Prisoner's Dilemma; SRE, self-reinforcing equilibrium; SCE, self-correcting equilibrium; BM, Bush–Mosteller.

†To whom reprint requests should be addressed. E-mail: mwm14@cornell.edu.

§Unilateral defection is not Pareto deficient, that is, there is no other outcome that both players prefer. However, the constraint $2R > T + S$ makes unilateral defection Rawls deficient. With equal probabilities for earning T and S, the expected CD payoff is $0.5(T + S)$, which leaves both players preferring CC under Rawls' constraint.

that there are now many more possibilities for mutual cooperation to obtain as a Nash equilibrium in the super game (a game of games). The bad news for game theory is that there are so many possibilities that the solution concept is robbed of predictive power.¶

An added concern is that analytical game theory imposes heroic assumptions about the knowledge and calculating abilities of the players. Simply put, game theorists tend to look for solutions for games played by people like themselves. For everyone else, the theory prescribes how choices ought to be made−−choices that bear little resemblance to actual decision making, even by business firms where forward-looking calculated rationality seems most plausible (4).

These problems have prompted the search for alternative solution concepts with greater predictive accuracy as well as precision. We would like an alternative to the Nash equilibrium that predicts a relatively small number of possible solutions to repeated social dilemma games. Equally important, where more than one equilibrium is possible, we want to understand the dynamics by which populations move from one equilibrium to another, a task for which analytical game theory is ill-equipped.

Agent-based models are an effective tool on both counts. They can be readily applied to backward-looking adaptive behavior, and they are useful for studying autopoetic and path-dependent dynamics. These models relax the conventional assumption of forward-looking calculation and explore backward-looking alternatives based on evolutionary adaptation (5) and learning (6). Evolution modifies the frequency distribution of strategies in a population, whereas learning modifies the probability distribution of strategies in the repertoire of an individual. In both evolution and learning, the probability that any randomly chosen individual uses a given strategy increases if the associated payoff is above some benchmark and decreases if below. In evolution, the benchmark is typically assumed to be the mean payoff for the population (7). In learning, the benchmark depends on the individual's aspirations.

The evolutionary approach has yielded a new solution concept, evolutionary stability. A strategy is evolutionarily stable if it cannot be invaded by any possible mutation of that strategy. All evolutionarily stable strategies (ESS) are Nash equilibria but the reverse is not true, which means that we now have a more restrictive solution concept for repeated games (7). However, the problem of indefiniteness remains. There is no strict ESS for the repeated PD (8).

Until recently, far less attention has been paid to learning theory as an alternative to evolutionary approaches. Fudenberg and Levine (6) give the first systematic account of the emerging new theory of learning in games. Our article builds on their work, and on earlier work by Roth and Erev (9) and Macy (10), which applies learning theory to the problem of cooperation in mixed-motive games. We identify a dynamic solution concept, stochastic collusion, based on a random walk from a self-correcting equilibrium (SCE) to a self-reinforcing equilibrium (SRE). These concepts make much more precise predictions about the possible outcomes for repeated games.

The learning-theoretic approach is based on connectionist principles of reinforcement learning, with origins in the cognitive psychology of William James. If a behavioral response has a favorable outcome, the neural pathways that triggered the behavior are strengthened, which "loads the dice in favor of those of its performances that make for the most permanent interests of the brain's owner" (11). Thorndike (12) later refined the theory as the Law of Effect, based on the principle that "pleasure stamps in, pain stamps out." This connectionist model of changes in the strength of neural pathways has changed very little during the 100 years since it was first presented and closely resembles the error back-propagation used in contemporary neural networks (13).

Applications of learning theory to the problem of cooperation do not solve the social dilemma, they merely reframe it: Where the penalty for cooperation is larger than the reward, and the reward for selfish behavior is larger than the penalty, how can penalty-aversive, reward-seeking agents elude the trap of mutual punishment?

The earliest answer was given by Rapoport and Chammah (14), who used learning theory to propose a Markov model of PD with state transition probabilities given by the payoffs for each state. Macy (10) elaborated Rapapport and Chammah's analysis by using computer simulations of their Bush–Mosteller (BM) stochastic learning model. Macy showed how a random walk may lead adaptive agents out of the social trap of a PD and into lock-in characterized by stable mutual cooperation, a process he characterized as stochastic collusion, the backward-looking equivalent to the tacit collusion engineered by forward-looking players.

More recently, Roth and Erev (9, 15) have proposed an alternative to the earlier BM formulation. Their payoff-matching model draws on the Matching Law, which holds that adaptive agents will choose between alternatives in a ratio that matches the ratio of reward. Applied to social dilemmas, both the BM and Roth–Erev models identify a key difference with analytical game-theoretic solutions: the existence of a cooperative equilibrium that is not Nash equivalent, even in Stag Hunt games where mutual cooperation is also a Nash equilibrium.

However, the generality of this solution may be questioned for two reasons. First, most theoretical attention has centered on PD games. We will show that the dynamics observed in PD cannot be generalized to games with only fear (Stag Hunt) or only greed (Chicken). Second, both the BM and Roth–Erev models have hidden assumptions about aspiration levels that invite skepticism about the simulation results. There are strong reasons to suspect that these earlier results are artifacts of hidden assumptions about payoff aspirations. By exploring a wider range of aspiration levels, we discover a previously unnoticed obstacle to cooperation among adaptive agents: a noncooperative equilibrium with much stronger attraction than the equilibrium for mutual cooperation.

We elaborated the BM model to parameterize variable aspiration levels and then applied the model to all three classes of social dilemma: PD, Stag Hunt, and Chicken. We show that the cooperative equilibrium based on stochastic collusion obtains in all social dilemmas in which each side is satisfied if the partner cooperates. We then show how attainment of this equilibrium depends on the level and adaptability of aspirations.

Principles of Reinforcement Learning

Analytical game theory assumes that players have sufficient knowledge and cognitive skill to make accurate predictions about the consequences of alternative choices. Learning theory lightens the cognitive load by allowing players to base these predictions on experiential induction rather than logical deduction. Learning theory also differs from game theory in positing two entirely separate cognitive mechanisms that guide decisions toward better outcomes, approach and avoidance. Rewards induce approach behavior, a tendency to repeat the associated choices even if other choices have higher utility. In contrast, punishments induce avoidance, leading to a search for alternative outcomes, including a tendency to revisit alternative choices whose outcomes are even worse.

Approach and avoidance imply two dynamic alternatives to the traditional Nash equilibrium. Rewards produce a SRE in which the equilibrium strategy is reinforced by the payoff, even

¶This result is known as the folk theorem of the theory of repeated games. The theorem asserts that if the players are sufficiently patient then any feasible, individually rational payoffs can be enforced by an equilibrium in an indefinitely repeated social dilemma game (3). The restriction of individual rationality of payoffs is a weak limitation for possible equilibria; it only assures that players do not accept less than their maximin payoff, which is *P* in the PD and Stag Hunt and *S* in Chicken.

if an alternative strategy has higher utility. The expected change in the probability of cooperation is zero when the probability of repeating the equilibrium strategy attains unity. The number of SRE in a social dilemma depends on the number of outcomes in which both players are rewarded.

A mix of rewards and punishments can produce a SCE in which outcomes that punish cooperation or reward defection (causing the probability of cooperation to decrease) balance outcomes that reward cooperation or punish defection (causing the probability of cooperation to increase). The expected change in the probability of cooperation is zero when the dynamics pushing the probability higher are balanced by the dynamics pushing in the other direction, like a tug-of-war between two equally strong teams. There can be at most one SCE in a social dilemma and there may be none.

The difference between approach and avoidance means that the effect of an outcome depends entirely on whether or not it is regarded as satisfactory. If the payoff exceeds aspirations, the probability increases that the behavior will be repeated rather than searching for a superior alternative, a behavioral tendency March and Simon (16) call "satisficing." While satisficing is suboptimal when judged by forward-looking game-theoretic criteria, it may be more effective in leading agents out of social traps than if they were to use more sophisticated decision rules, such as testing the waters to see whether they could occasionally get away with cheating. Satisficing produces SRE by precluding search for superior alternatives.

If the payoff falls below aspirations, the probability decreases that the behavior will be repeated. Avoidance precludes the opportunity to minimize losses by remaining with the lesser of two evils, a pattern we call "dissatisficing." Dissatisficing produces SCE by inducing search for alternatives to punished behavior.

Changing behavior is not the only way that adaptive agents can respond to an aversive stimulus. They can also respond by adapting their aspiration level, a process known as habituation (17). Habituation can lead to desensitization to a recurrent stimulus, whether reward or punishment, and to increased sensitivity to change in the stimulus. Thus, habituation to reward increases sensitivity to punishment. Conversely, habituation to punishment has a numbing effect that increases sensitivity to reward.

BM: An Agent-Based Model of Reinforcement Learning

In general form, the BM model consists of a stochastic decision rule and a learning algorithm in which the consequences of decision create positive and negative stimuli (rewards and punishments) that update the probability p that the decision will be repeated. Applied to two-person social dilemmas, the model assumes binary choices (C or D) that intersect at one of four possible outcomes (CC, CD, DC, or DD), each with an associated payoff (R, S, P, and T, respectively) that is evaluated as satisfactory or unsatisfactory relative to an aspiration level. Although the BM model implies the existence of some aspiration level relative to which cardinal payoffs can be positively or negatively evaluated, the model imposes no constraints on its determinants. Whether aspirations are high or low or change with experience (habituation) depends on assumptions that are exogenous to the model. Given some aspiration level, satisfactory payoffs present a positive stimulus (or reward) and unsatisfactory payoffs present a negative stimulus (or punishment). Rewards and punishments then modify the probability of repeating the associated action. Diagram 1 shows how the probability of action a ($a \in \{C, D\}$) is updated by the associated stimulus s.

The classification of payoffs as satisfactory or unsatisfactory requires that agents hold an aspiration level relative to which payoffs are positively or negatively evaluated on a standard scale. Formally, the stimulus s_a is calculated as

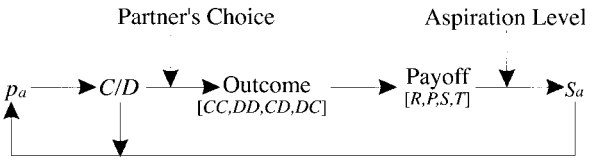

Diagram 1. Schematic representation of the BM stochastic learning model.

$$s_a = \frac{\pi_a - A}{\sup[|T - A|, |R - A|, |P - A|, |S - A|]}, a \in \{C, D\}, \quad [1]$$

where π_a is the payoff associated with action a (R or S if $a = C$, and T or P if $a = D$). A is the aspiration level and s_a is a positive or negative stimulus derived from π_a. The denominator in Eq. **1** represents the upper value of the set of possible differences between payoff and aspiration. With this scaling factor, stimulus s is always equal to or less than unity in absolute value, regardless of the magnitude of the corresponding payoff.

If the aspiration level A is fixed, s will be unaffected by habituation to repeated stimuli. If A changes with experience, habituation to π will cause s to decline over time in absolute value. The tendency for s to approach zero with repeated reinforcement corresponds to satiation, whereas the tendency for s to approach zero with repeated punishment corresponds to desensitization. To implement habituation, we assume that the aspiration level adapts to recent payoffs π. More precisely, the updated aspiration level, A_{t+1}, is a weighed mean of the prior aspiration level at time t and the payoff π that was experienced at t. Formally,

$$A_{t+1} = (1 - h)A_t + h\pi_t, \quad [2]$$

where h indicates habituation, i.e., the degree to which the aspiration level floats toward the payoffs. With $h = 0$, the aspiration level is constant, that is, recent payoffs are ignored and the initial aspiration A_0 is preserved throughout the game. With $h = 1$, aspirations float immediately to the payoff that was received in the previous iteration. The aspiration level then provides the benchmark for positive or negative evaluation of the payoffs, as given by Eq. **1**.

This evaluation, in turn, determines whether the probability of taking the associated action increases or decreases. Applied to behavior in social dilemmas, the model updates probabilities after an action a (cooperation or defection) as follows:

$$p_{a,t+1} = \begin{cases} p_{a,t} + (1 - p_{a,t})ls_{a,t}, & \text{if } s_{a,t} \geq 0 \\ p_{a,t} + p_{a,t}ls_{a,t}, & \text{if } s_{a,t} < 0 \end{cases}, a \in \{C, D\}. \quad [3]$$

In Eq. **3**, $p_{a,t}$ is the probability of action a at time t, l is the learning rate ($0 < l < 1$), and s_{at} is the positive or negative stimulus experienced after action a in t. The change in the probability for the action not taken obtains from the constraint that probabilities always sum to unity.

Eq. **3** implies a tendency to repeat rewarded behaviors and avoid punished behaviors, consistent with the Law of Effect. The model implies a cooperative SRE if (and only if) both agents' aspiration levels are lower than the payoff for mutual cooperation (R). Aspirations above this point necessarily preclude a mutually cooperative solution to social dilemmas. If both players' aspiration levels fall below R but exceed maximin (the largest possible payoff they can guarantee themselves), then there is a unique SRE in which both players receive a reward, namely, mutual cooperation.

Very low aspirations do not preclude mutual cooperation as an equilibrium but may prevent adaptive agents from finding it. In social dilemmas, if aspiration levels are below maximin, then mutual or unilateral defection may also be mutually reinforcing, even though these outcomes are socially deficient. Multiple SRE

Table 1. Treatment groups for exploration of parameter space for the BM model, with π = [4,3,1,0]

Aspiration level	Parameters	Satisf.	Dissatisf.	Habituation
Fixed low	$h = 0$, $A = 0.5$	Yes	No	No
Fixed high	$h = 0$, $A = 3.5$	No	Yes	No
Floating	$h = 0.2$	After punish	After reward	Yes

mean that learning may get trapped in an alternative basin of attraction before locking in mutual cooperation.

The SRE is a black hole from which escape is impossible. In contrast, players are never permanently trapped in SCE; a chance sequence of fortuitous moves can lead both players into a self-reinforcing stochastic collusion. The fewer the number of coordinated moves needed to lock-in SRE, the better the chances. Thus, the odds of attaining lock-in are highest if $l \approx 1$ and behavior approximates a "win-stay, lose-change" heuristic, in which choices always repeat when rewarded and always change when punished (18).

In sum, the BM model identifies stochastic collusion as a backward-looking solution for social dilemmas. However, this solution is available only if both players have aspirations levels below R, such that each is satisfied when both partners cooperate. Stochastic collusion is guaranteed if an additional condition can be met: Both players have aspirations fixed above their maximin and below R, such that mutual cooperation is the unique SRE.

By manipulating the aspiration level, we can test both the existence and the attraction to SRE under different payoff inequalities. A fixed aspiration level can be manipulated to study the effects of satisficing (when aspirations are low) and dissatificing (when aspirations are high). A floating aspiration level can be used to study the effects of habituation. Habituation is modeled as an aspiration level that floats in the direction of a repeated stimulus, regardless of the associated behavior. If the payoff exceeds aspirations, the aspiration level increases, leading to satiation on reward and sensitization to punishment. If the payoff falls below aspirations, the aspiration level decreases, leading to sensitization on reward and desensitization to punishment.

The model in Eqs. **1** and **2** allows us to independently manipulate satisficing (precluded by fixed high aspirations), dissatisficing (precluded by fixed low aspirations), and habituation (precluded by fixed aspirations of any level), using just two parameters (h and A). Table 1 illustrates the corresponding manipulations for the vector of payoffs π = (4,3,1,0).

We apply this model to the three classic types of social dilemma (PD, Chicken, and Stag Hunt) and perform a series of experiments that systematically explore the solutions that emerge with different parameter combinations over each of the three games.

Learning-Theoretic Solution Concepts for Social Dilemmas

We begin by testing the generality of stochastic collusion as a solution concept, with aspirations fixed midway between maximin and minimax. This aspiration level can also be interpreted as the expected payoff when behavioral propensities are uninformed by prior experience ($p_a = 0.5$), such that all four payoffs are equiprobable ($p_\pi = 0.25$). We define this as a neutral aspiration, designated A^0, while A^+ and A^- designate aspirations above minimax and below maximin, respectively.

Macy (10) identified mutual cooperation as the unique SRE (or lock-in) for repeated PD. He also identified a mixed-strategy SCE in which the ratio of bilateral (*CC* or *DD*) to unilateral (*CD*) outcomes balances the ratio of unilateral to bilateral payoffs (relative to aspirations).

Assuming neutral aspirations (A^0, we can show analytically that these results generalize to Chicken and Stag Hunt by solving for the level of cooperation at which the expected change in the probability of cooperation is zero. The expected change is zero when, for both players, the probability of outcomes that reward cooperation or punish defection, weighted by the absolute value of the associated change in propensities, equals the probability of outcomes that punish cooperation or reward defection, weighted likewise. For all possible payoffs that conform to the four conditions we impose for social dilemmas, SRE obtains at $p_c = 1$. In PD, the cooperation rate in the SCE is below $p_c = 0.5$, which is the asymptotic upper bound that the equilibrium approaches as R approaches T and P approaches S simultaneously. The lower bound is $p_c = 0$ as P approaches A^0. In Chicken, the corresponding upper bound is $p_c = 1$ as R approaches T and S approaches A^0. The lower bound is $p_c = 0$ as R, S, and P all converge on A^0 (retaining $R > A^0 > S > P$). Only in Stag Hunt is it possible that there is no SCE, if $R - T > A^0 - S$. The lower bound is $p_c = 0$ as T approaches R and P approaches A^0.

These analytical results do not tell us much about the dynamics or the probability of moving from one equilibrium to another. To that end, we ran computational experiments with the payoffs ordered from the set π = (4,3,1,0) for each of the three social dilemma payoff inequalities. With this set, minimax is always 3, maximin is always 1, giving $A^0 = 2$, such that mutual cooperation is a unique SRE in each of the three games, a key scope condition for stochastic collusion. With $h = 0$, there are two analytical solutions in each of the three games. The first solution is the SCE at $p_c = 0.37$ in PD and at $p_c = 0.5$ in Chicken and Stag Hunt. The second solution is a unique SRE at $p_c = 1$ in each of the three games.

We observe stochastic collusion by setting the learning rate l high enough to facilitate coordination by random walk into the SRE at *CC*. The parameter $l = 0.5$ allows lock-in on mutual cooperation within a small number of coordinated *CC* moves. Fig. 1 confirms the possibility of stochastic collusion in PD and extends this result for the two other social dilemma games. Fig. 1 charts the change in the probability of cooperation, p_c (which in this case is statistically identical across the two players).

Fig. 1 shows that the characteristic pattern of stochastic collusion occurs in all three games, but not with equal probability. Mutual cooperation locks in most readily in Stag Hunt and least readily in PD. To test the robustness of this difference, we simulated 1,000 replications of this experiment and measured the

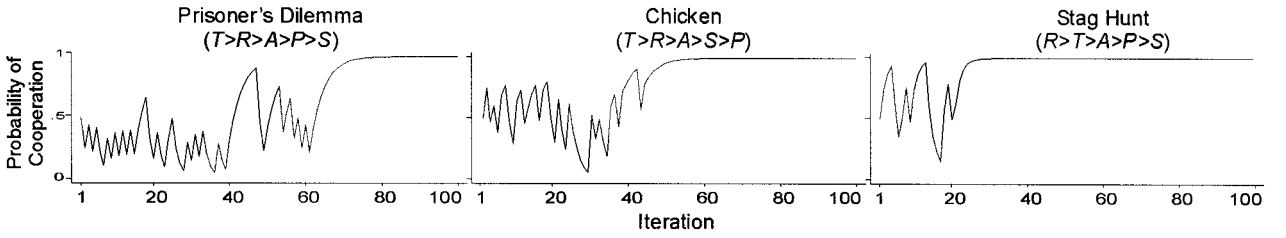

Fig. 1. Change in p_C over 100 iterations [π = (4,3,1,0), $A^0 = 2$, $l = 0$, 5, $h = 0$, $p_{C,1} = 0.5$].

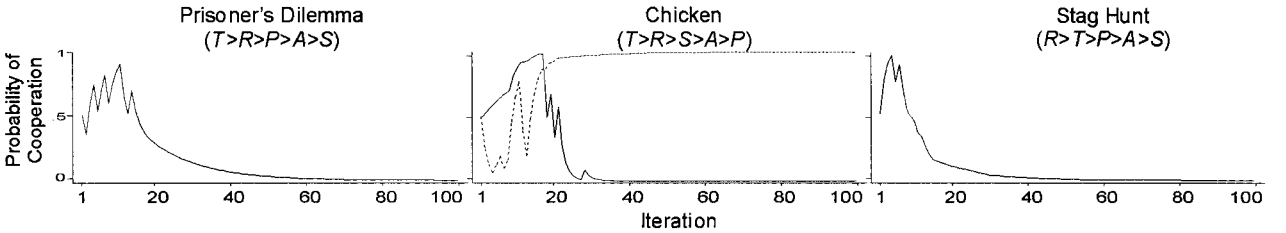

Fig. 2. Change in p_C over 100 iterations with low aspirations [$\pi = (4,3,1,0)$, $A^- = 0.5$, $l = 0.5$, $h = 0$, $p_{C,1} = 0.5$]. Broken line indicates p_C for player 2 in Chicken.

proportion of runs that locked in mutual cooperation within 100 iterations. We would expect the lock-in rate to be lowest in PD, and that is confirmed, with a relative frequency of 0.13. We might also expect the lock-in rate to be similar in Chicken and Stag Hunt, because the model parameters yielded identical equilibrium solutions ($p_c = 1$ and $p_c = 0.5$ in both games). Surprisingly, this was not the case. The probability of lock-in was nearly twice as high in Stag Hunt (0.96) compared with Chicken (0.45).

This finding suggests that, all things being equal, the problem of cooperation in PD is caused more by greed than fear, even when the decomposed K indices[‖] for fear and greed are identical—in this case, $K_f = (P - S)/(T - S) = 0.25$ and $K_g = (T - R)/(T - S) = 0.25$. Holding all else constant, removing fear (by setting $S > P$) only improved the probability of cooperation 32 points (from 0.13 to 0.45), whereas removing greed (by setting $R > T$) improved the probability of cooperation an impressive 83 points (from 0.13 to 0.96).

A moment's reflection shows why. Escape from the SCE by random walk depends on a chance string of bilateral moves, either *CC* (which rewards cooperation) or *DD* (which punishes defection). At $p_c = 0.5$, the equilibrium probabilities for *CC* and *DD* are equal in Chicken and Stag Hunt. However, with each step in the random walk, the probability of *CC* increases relative to *DD*, such that R plays an increasingly greater role in the escape. Greed constrains the value of R (relative to T) whereas fear does not.

Effect of Aspiration Level: Low, High, or Floating. The simulations with neutral aspirations ($A^0 = 2$) make it possible for players to satisfice when mutual cooperation is achieved and dissatisfice when it fails. Intuitively, dissatisficing is crucial for stochastic collusion. Otherwise, players could lock in on the lesser of two evils, with zero probability of mutual cooperation. Dissatisfaction with the social costs of a socially deficient outcome motivates players to continue searching for a way out of the social trap. The players then wander about in SCE with a nonzero probability of cooperation that makes possible the chance sequence of bilateral cooperative moves needed to escape the social trap by random walk, as illustrated in Fig. 1.

Mutual cooperation is always the optimal outcome in Stag Hunt, a game in which there is no temptation to cheat ($R > T$). However, satisficing is crucial for stochastic collusion in PD and Chicken. Appreciation that the payoff for mutual cooperation is good enough motivates players to stay the course despite the temptation to cheat (given $T > R$). Otherwise, mutual cooperation will not be self-reinforcing.

The dilemma is that dissatisficing and satisficing are complementary with respect to the aspiration level. Higher aspirations increase the level of dissatisfaction with P, thereby increasing the rate of cooperation in SCE. Lower aspirations increase the level of satisfaction with R, thereby strengthening reinforcement for mutual cooperation. This dilemma poses an interesting puzzle: Where is the optimal balance point between satisficing and dissatisficing, and how does the optimum vary across the three games?

In Stag Hunt, the absence of a temptation to cheat makes satisficing less important than in the other two games (where $T > R$). Conversely, in Chicken, the high cost of mutual defection (given $P < S$) suggests that dissatisficing is relatively less important. And in all three games, an aspiration level below the maximin payoff generates a second SRE besides mutual cooperation. This alternative basin of attraction may trap the players in a self-reinforcing dynamic of mutual or unilateral defection.

Fixed Aspirations: Low vs. High. To test this idea, we assumed a fixed aspiration level and manipulated it, ranging from below the maximin payoff (which limits dissatisficing) to above the payoff for mutual cooperation (which limits satisficing). More precisely, for low aspirations, we fixed the aspiration level midway between the minimum ($\pi = 0$) and maximin ($\pi = 1$) payoffs ($A^- = 0.5$). For high aspirations, we fixed the aspiration level midway between the maximum ($\pi = 4$) and minimax ($\pi = 3$) payoffs ($A^+ = 3.5$). Fig. 2 shows typical simulation runs under low aspirations. The graph for the Chicken game also reports the probability of cooperation for the second player (broken line) whose probabilities differ because of the possibility for reinforcement of unilateral cooperation (given $S > A^-$).

Fig. 2 presents typical simulation runs for the low-aspiration condition in each of the three social dilemmas. The results show that low aspirations—and thus a high level of satisficing relative to dissatisficing—can undermine stochastic collusion. With aspirations below the maximin payoff, a socially deficient outcome (either mutual or unilateral defection) becomes an alternative SRE that competes with mutual cooperation as an attractor. In PD and Stag Hunt, both players are attracted to mutual defection as good enough. Cooperation in Stag Hunt almost attains lock-in around $t = 5$ (because of $R > T$) but this result is not sufficient to overcome the far greater pull of the competing attractor. In Chicken, unilateral defection becomes a competing attractor because $S > A^-$.

We tested the robustness of these results by measuring the proportion of trials that locked into one of the SRE within 100 iterations of the game. In all three games, the deficient attractor dominated mutual cooperation. The proportion of trials that converged on the deficient SRE was 0.89 in PD, 0.72 in Chicken, and 0.59 in Stag Hunt. All runs that did not converge on a deficient equilibrium ended up in *CC* as an SRE.

The stronger attraction to mutual cooperation in Chicken (0.28) compared with PD (0.13) reflects the aversion to mutual defection that is present in Chicken (given $P < A^-$) but not in PD. This aversion increases both players' cooperative propensities, and with a relatively high learning rate, this can jump-start random walk into mutual cooperation. In PD, players are aversive only to the S payoff, which teaches the opposite lesson: it does not pay to cooperate. The only payoff that increases both

[‖]The K index was invented by Rapoport and Chammah (14) to predict the level of cooperation in mixed-motive games, where $K = (R - P)/(T - S)$. We decomposed the index into fear (K_f) and greed (K_g), such that $K = 1 - (K_f + K_g)$.

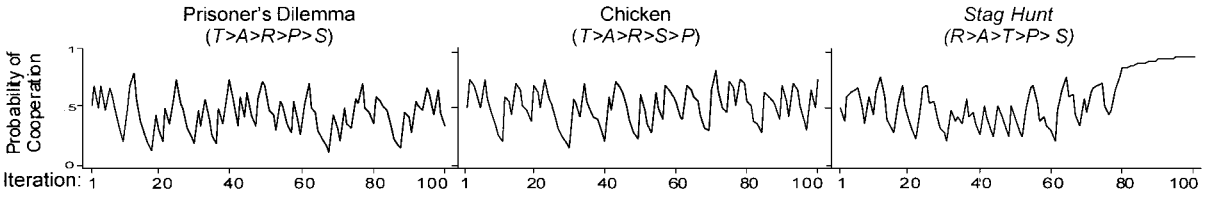

Fig. 3. Change in p_C over 100 iterations with high aspirations [π = (4,3,1,0), A^+ = 3.5, l = 0.5, h = 0, $p_{C,1}$ = 0.5].

players' cooperative propensities is R, which quickly becomes out of reach as players are pulled into the noncooperative SRE.

The adverse effect of low aspirations might seem to suggest that high aspirations are a possible solution to social dilemmas. However, Fig. 3 shows that this is not the case. With the exception of Stag Hunt, high aspirations preclude convergence on mutual cooperation. The reason is that with $A^+ > R$, there is no SRE in PD and Chicken, a prerequisite for stochastic collusion. Only in Stag Hunt is mutual cooperation an SRE, and even here, robustness tests show that this equilibrium is attained in only 120 of 1,000 replications, a rate of 0.12.

To test the effects of aspiration more systematically, we varied the aspiration level A across the entire range of payoffs (from 0 to 4) in steps of 0.1. The graphs in Fig. 4 show the proportion of runs that lock in mutual cooperation within 250 iterations, based on 1,000 replications at each level of aspiration.

Fig. 4 confirms the optimal aspiration level between maximin and R. Below maximin, attraction to a deficient SRE reduces the probability of stochastic collusion to about 0.1 in PD, 0.25 in Chicken, and 0.4 in Stag Hunt. The probability then sharply increases in all three games as soon as aspirations exceed maximin ($A > 1$) and continues to do so in PD and Chicken until a turning point is reached at approximately $A = 2$. For $A > 2$, mutual cooperation loses its appeal as an outcome that is good enough, leading to zero chance of stochastic collusion as aspirations approach R (at $A = 2.5$ in PD and in Chicken and $A = 3.5$ in Stag Hunt).

To test the generality of this qualitative result across different payoff vectors, we replicated the simulation with fear and greed varied relative to the baseline payoff vector π = (4,3,1,0). We assumed higher greed ($T = 5$) in PD and Chicken and greater fear ($S = -1$) in PD and Stag Hunt. Fig. 5 shows the results of increased fear and greed relative to the superimposed baseline results of Fig. 4 (broken line).

Fig. 5 supports the generality of the nonlinear effect of aspiration A. At the same time, the results reveal that both higher greed and higher fear inhibit stochastic collusion, all else being equal. Moreover, the manipulation of fear and greed in PD confirms what we also find in the comparison between games. Greed is a bigger problem for stochastic collusion than fear. In PD, the 2-fold increase in greed from $T - R = 1$ to $T - R = 2$ causes a decline in the maximum rate of stochastic collusion, from 0.8 in the baseline condition to about 0.2. However, the 2-fold increase in fear reduces the maximum rate of stochastic collusion only to a level of 0.5. Correspondingly, although the effect of greed on the peak rate in Chicken is substantial (a decline from about 0.95 to about 0.6), fear does not reduce the lock-in rate in Stag Hunt. The only effect of fear in Stag Hunt is a slight reduction of the range of aspiration levels under which stochastic collusion can be obtained.

Floating Aspirations: Habituation. Given the adverse effects of both low and high aspirations, one might conclude that the best way to find the optimal balance point is to let the agents find it themselves, by allowing the aspiration level to float. However, on closer inspection it seems that the opposite is more likely the case. Floating aspirations are the worst of both worlds, causing players to become easily dissatisfied with mutual cooperation and numb to the social costs of socially deficient outcomes. Moreover, the tendency to adapt to a repeated stimulus also increases sensitivity to changes in the stimulus. Thus, agents who became habituated to reward in SRE will respond more aversively to unexpected punishment (and vice versa). Accordingly, habituation not only attenuates the self-reinforcing effect of R, it also amplifies the destabilizing effects of a chance defection.

Fig. 6 confirms the expected destabilizing effects, based on conditions identical to those in Fig. 1 except for a modest increase in habituation. With h = 0.2, agents' reference points represent a moving average of past payoffs based on a long memory. Relatively slow adaptation of aspirations gives the players sufficient time for stochastic collusion, but once SRE is obtained, they eventually will habituate to rewards.

Fig. 6 shows that, in all three games, players can temporarily achieve stochastic collusion but cannot sustain it. Habituation to reward destabilizes mutual cooperation as players lose their appetite for R. High aspirations also amplify disappointment when cooperation is eventually "suckered," leading quickly to the SCE. This in turn restores the appetite for mutual cooperation, especially in Chicken where the payoff for mutual defection is highly aversive.

Reliability tests confirm the pattern in Fig. 6. Over 1,000 replications, the rate of mutual cooperation in iteration 500 was higher in Chicken (0.49) than in Stag Hunt (0.39) and lowest in PD (0.16). Cooperation in PD suffers from less aversion to mutual defection in SCE (compared with Chicken) and less attraction to mutual cooperation in SRE (compared with Stag Hunt).

Comparison with the corresponding results with fixed neutral aspirations (Fig. 1) clearly demonstrates the destabilizing effect of habituation in all three games. Over 1,000 replications with aspirations fixed at $A^0 = 2$, the rate of mutual cooperation was

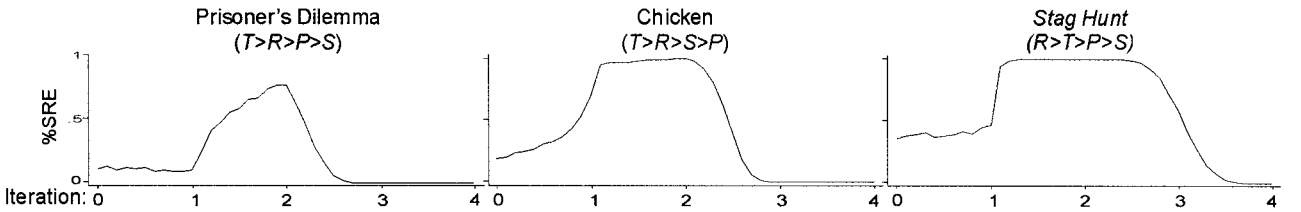

Fig. 4. Effects of aspiration levels on SRE within 250 iterations [π = (4,3,1,0), l = 0.5, h = 0, pC,1 = 0.5, n = 1,000].

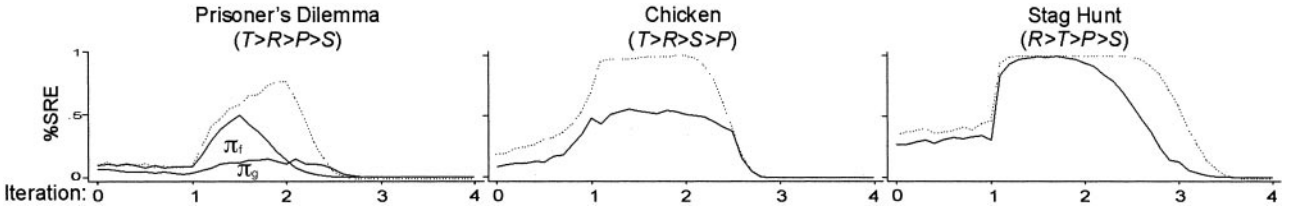

Fig. 5. Effect of aspirations ($0 < A < 4$) on SRE within 250 iterations, by fear [$\pi_f = (4,3,1,-1)$] and greed [$\pi_g = (5,3,1,0)$]. $l = 0.5$, $h = 0$, $p_{C,1} = 0.5$, $n = 1{,}000$. Broken line shows baseline results with $\pi = (4,3,1,0)$.

nearly universal by iteration 500 in PD (0.95), Chicken (0.98), and Stag Hunt (1.0).

How Habituation Can Also Promote Cooperation. The destabilizing effect of habituation on mutual cooperation suggests a way that adaptation to stimuli may also promote cooperation through the destabilization of noncooperative SRE. Previously we observed how low fixed aspirations generate a competing attractor that diverts random walk away from mutual cooperation and into mutual or unilateral defection. However, if agents adapt to repeated exposure to the associated payoffs, they may be able to eventually escape. Moreover, the increased sensitivity to a change in stimuli should then amplify the attraction to mutual cooperation.

To test this hypothesis, we combined a low initial aspiration level ($A_0^- = 0.5$) with modest habituation ($h = 0.2$). Fig. 7 shows representative examples of the dynamics that result.

Comparison with corresponding results without habituation (Fig. 2) confirms the hypothesized effect of adaptation to payoffs. With $h = 0$, satisficing with low initial aspirations generally led to quick convergence on the SRE with lowest coordination complexity. Mutual cooperation was preempted by mutual defection in PD and unilateral defection in Chicken. By contrast, Fig. 7 shows that moderate habituation weakens these deficient attractors relative to mutual cooperation, which remains viable as a punctuated equilibrium.

Curiously, in Stag Hunt, habituation has the opposite effect, destabilizing mutual cooperation in favor of temporary convergence on mutual defection. Reliability tests confirm this pattern. Table 2 shows the interaction between satisficing and habituation on the rate of stochastic collusion within 500 iterations, over 1,000 replications of the three social dilemma games. Mutual cooperation decreases with habituation when aspirations are initially neutral ($A_0^0 = 2.0$) in all three games. However, when aspirations are initially low ($A_0^- = 0.5$), mutual cooperation decreases with habituation only in Stag Hunt; in the other two games, mutual cooperation increases. Put differently, in the absence of habituation ($h = 0$), mutual cooperation suffers when aspirations are low, in all three games. However, with moderate habituation ($h = 0.2$), low initial aspirations have almost no effect on mutual cooperation.

Discussion and Conclusion

The Nash equilibrium, the main solution concept in analytical game theory, cannot make precise predictions about the outcome of repeated mixed-motive games. Nor can it tell us much about the dynamics by which a population of players can move from one equilibrium to another. These limitations, along with concerns about the cognitive demands of forward-looking rationality, have led game theorists to explore backward-looking alternatives based on evolution and learning. Considerable progress has been made in agent-based modeling of evolutionary dynamics (5), but much less work has been invested in learning-theoretic approaches. This lacunae is curious. Evolution operates on the global distribution of strategies within a given population, whereas learning operates on the local distribution of strategies within the repertoire of each individual member of that population. From the perspective of multilevel theorizing, learning can be viewed as the cognitive microfoundation of extragenetic evolutionary change. Just as evolutionary models theorize autopoetic population dynamics from the bottom up, learning models theorize social and cultural evolution from the bottom up, beginning with a population of strategies competing for the attention of a single individual.

Previous applications of reinforcement learning to the evolution of cooperation have used the BM stochastic learning model and the Roth–Erev payoff-matching model. Elsewhere (19) we explore some interesting differences between these models, centered on Blackburn's Power Law of Learning. Here, we focus on the convergent prediction––the possibility of stochastic collusion in which adaptive agents escape a socially deficient equilibrium by random walk.

The generality of this solution has two important limitations. First, previous applications were restricted to PD games, which precluded disaggregation of the dynamic properties of fear and greed. Second, aspiration levels were arbitrarily fixed, which precluded analysis of the dynamic properties of satisficing, dissatisficing, and habituation.

This study maps the landscape for the game dynamics at the cognitive level, beginning with the simplest possible iterated social dilemma, a two-person repeated game. Using a BM stochastic learning model, we identify a fundamental solution concept for the long-term dynamics of backward-looking behavior in all social dilemmas–stochastic collusion–based on random walk into SRE. However, the viability of this solution is sensitive to the aspiration level and the relative magnitude of fear and greed.

With fixed aspirations between R and maximin, stochastic collusion is much more robust against an increase in fear compared with an increase in greed. Greed undermines attrac-

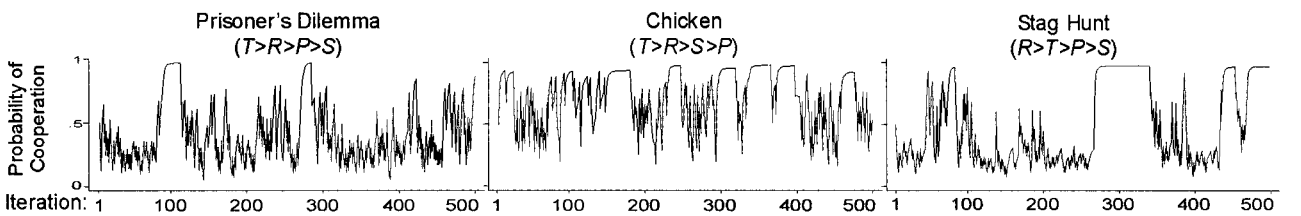

Fig. 6. Change in p_C over 500 iterations with floating aspirations [$\pi = (4,3,1,0)$, initial $A_0^0 = 2.0$, $l = 0.5$, $h = 0.2$, $p_{C,1} = 0.5$].

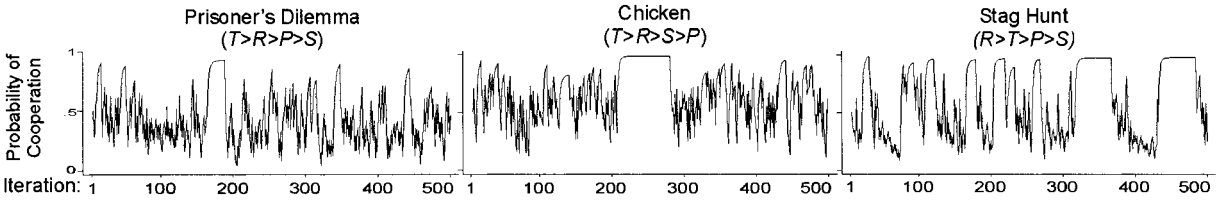

Fig. 7. Change in p_C over 500 iterations with initially low floating aspirations [$\pi = (4,3,1,0)$, $A_0^0 = 0.5$, $l = 0.5$, $h = 0.2$, $p_{C,1} = 0.5$].

tion to mutual cooperation as the SRE, making it more difficult to escape from a socially deficient SCE through a fortuitous sequence of bilateral moves. Thus, when fear and greed are equal in magnitude, the probability of stochastic collusion is greater in Stag Hunt than in Chicken.

Stochastic collusion is also highly sensitive to aspirations. If aspirations are too low, mutual cooperation can suffer from insufficient dissatisfaction with socially deficient outcomes. Mutual cooperation can then be preempted by attraction to an alternative SRE that is socially deficient––mutual defection (in PD and Stag Hunt) or unilateral defection (in Chicken). If aspirations are too high, agents may not feel sufficiently rewarded by mutual cooperation to avoid the temptation to defect. High aspirations are the greater problem in Chicken, whereas low aspirations are the greater problem in Stag Hunt.

We also explored the effects of habituation–aspirations that adjust to experience. Habituation destabilizes stochastic collusion in all three games. Paradoxically, the problem is particularly acute in Stag Hunt, where the low reward for exploitation (T) facilitates accommodation to the social costs of mutual defection. On the other hand, habituation can also destabilize the socially deficient SRE created by low baseline aspirations.

Although we are struck by the amount that could be learned about the emergent dynamics of cooperation from a model as simple as BM, these results need to be interpreted cautiously. Our exploration of learning-theoretic solutions to social dilemmas is necessarily incomplete. We have limited the analysis to symmetrical two-person simultaneous social dilemma games within a narrow range of possible payoffs. Previous work (10, 20) suggests that the coordination complexity of stochastic collusion increases with the number of players and payoff asymmetry and decreases with social influence. Much more work needs to be done to study the evolution of cooperation at the cognitive level, especially where dyadic games are embedded in dynamic social networks, a problem addressed by Macy and Sato (23). Going forward, we anticipate much more compelling insights with multiplex models applied to socially embedded games.

We also assumed agents with minimal cognitive complexity. Previous research by Hegselmann and Flache (21) suggests that agents might be better off using more sophisticated strategies for conditional cooperation, such as Tit for Tat (which retaliates against defection by defecting on the next play) or Grim Trigger (which retaliates by defecting forever). In future research, cognitive game theory faces the challenge to show how more sophisticated strategy choices might emerge from simple learning principles. As a first step, Macy (22) used artificial neural networks to see whether adaptive agents could learn to cooperate based on conditionally cooperative super-game strategies, but found that the coordination complexity of stochastic collusion increased exponentially with the strategy space. Much more work is needed to see how adaptive agents might also learn to find nodal points or other solutions to the problem of coordination complexity.

Although theoretical elaborations are clearly needed, we should not lose sight of the elementary principles suggested by a simple learning model of the dynamics of cooperation. Social order may ultimately depend less on top-down global coordination or enforcement than on bottom-up emergence of self-enforcing norms for cooperation in everyday life. If so, then the emphasis in agent-based modeling of the evolution of cooperation may need to shift downward, from evolutionary dynamics at the population level to cognitive dynamics at the level of the individual. We see the simple BM learning model for two-person games as a cautious step in this direction.

Table 2. Effects of habituation on stochastic collusion within 500 iterations, by initial aspiration level ($l = 0.5$, $p_{a,1} = 4$, $N = 1{,}000$)

Aspiration level	PD		Chicken		Stag hunt	
	$h = 0$	$h = 0.2$	$h = 0$	$h = 0.2$	$h = 0$	$h = 0.2$
$A_0^0 = 2.0$	0.96	0.17	0.99	0.48	1.00	0.35
$A_0^- = 0.5$	0.10	0.18	0.30	0.49	0.39	0.36

We thank Yoshimichi Sato, Douglas Heckathorn, and other members of the Janus Workshop on Backward-Looking Rationality at Cornell University. The research was supported by grants to M.W.M. from the U.S. National Science Foundation (SES-9819249 and SES-0079381). A.F.'s research was made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

1. Dawes, R. M. (1980) *Annu. Rev. Psychol.* **31,** 169–193.
2. Rawls, J. (1971) *A Theory of Justice* (Harvard Univ. Press, Cambridge, MA).
3. Fudenberg, D. & Tirole, J. (1991) *Game Theory* (MIT Press, Cambridge, MA).
4. Simon, H. (1992) in *Decision Making: Alternatives to Rational Choice Models*, ed. Zey, M. (Sage, Newbury Park, CA), pp. 32–53.
5. Binmore, K. G. & Samuelson, L. (1992) *J. Econ. Theory* **57,** 278–305.
6. Fudenberg, D. & Levine, D. (1998) *The Theory of Learning in Games* (MIT Press, Cambridge, MA).
7. Weibull, J. W. (1998) *Eur. Econ. Rev.* **42,** 641–649.
8. Boyd, R. & Lorderbaum, J. (1987) *Nature (London)* **327,** 58–59.
9. Roth, A. E. & Erev, I. (1995) *Games Econ. Behav.* **8,** 164–212.
10. Macy, M. W. (1991) *Am. J. Soc.* **97,** 808–843.
11. James, W. (1981) *Principles of Psychology* (Harvard Univ. Press, Cambridge, MA).
12. Thorndike, E. L. (1898) *Animal Intelligence: An Experimental Study of the Associative Processes in Animals* (MacMillan, New York).
13. Rummelhart, D. E. & McLell, J. L. (1988) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (MIT Press, Cambridge, MA).
14. Rapoport, A. & Chammah, A. M. (1965) *Prisoner's Dilemma: A Study in Conflict and Cooperation* (Michigan Univ. Press, Ann Arbor).
15. Erev, I. & Roth, A. E. (1998) *Am. Econ. Rev.* **88,** 848–879.
16. March, J. G. & Simon, H. A. (1958) *Organizations* (Wiley, New York).
17. Sokolov, Y. N. (1963) *Perception and the Conditioned Reflex* (Pergamon, Oxford).
18. Nowak, M. A. & Sigmund, K. (1993) *Nature (London)* **364,** 56–58.
19. Flache, A. & Macy, M. W. (2002) *J. Conflict Res.*, in press.
20. Flache, A. & Macy, M. W. (1996) *J. Math. Soc.* **21,** 3–28.
21. Hegselmann, R. & Flache, A. (2000) *Analyse Kritik* **22,** 75–97.
22. Macy, M. W. (1996) *Soc. Methods Res.* **25,** 103–137.
23. Macy, M. W. & Sato, Y. (2002) *Proc. Natl. Acad. Sci. USA* **99,** Suppl. 3, 7214–7220.