MANAGEMENT SCIENCE
Vol. 53, No. 7, July 2007, pp. 1036–1050
ISSN 0025-1909 | EISSN 1526-5501 | 07 | 5307 | 1036

DOI 10.1287/mnsc.1060.0625

Cooperation in Evolving Social Networks

Nobuyuki Hanaki
Doctoral Program in International Political Economy, Graduate School of Humanity and Social Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan, hanaki@dpipe.tsukuba.ac.jp

Alexander Peterhansl
Department of Economics, Columbia University, 1022 International Affairs Building, 420 West 118th Street, New York, New York 10027, ap11@columbia.edu

Peter S. Dodds
Department of Mathematics and Statistics, 203 Lord House, University of Vermont, 16 Colchester Avenue, Burlington, Vermont 05401, pdodds@uvm.edu

Duncan J. Watts
Institute for Social and Economic Research and Policy, Columbia University, 8th Floor, International Affairs Building, 420 West 118th Street, New York, New York 10027 and Department of Sociology, Columbia University, 413 Fayerweather Hall, 1180 Amsterdam Avenue, New York, New York 10027, djw24@columbia.edu

We study the problem of cooperative behavior emerging in an environment where individual behaviors and interaction structures coevolve. Players not only learn which strategy to adopt by imitating the strategy of the best-performing player they observe, but also choose with whom they should interact by selectively creating and/or severing ties with other players based on a myopic cost-benefit comparison. We find that scalable cooperation—that is, high levels of cooperation in large populations—can be achieved in sparse networks, assuming that individuals are able to sever ties unilaterally and that new ties can only be created with the mutual consent of both parties. Detailed examination shows that there is an important trade-off between local reinforcement and global expansion in achieving cooperation in dynamic networks. As a result, networks in which ties are costly and local structure is largely absent tend to generate higher levels of cooperation than those in which ties are made easily and friends of friends interact with high probability, where the latter result contrasts strongly with the usual intuition.

Key words: networks-graphs; theory; games-group decisions; simulation
History: Accepted by Brian Uzzi and Luis Amaral, special issue editors; received May 11, 2005. This paper was with the authors 1 month for 1 revision.

1. Introduction

Cooperation is a widely observed feature of human and animal societies (Ostrom et al. 1999, Fehr and Fischbacher 2003), arising even in populations of nonsentient organisms (Boorman and Levitt 1980). Although it manifests itself in many versions, the crux of all cooperation problems is the notion of a *social dilemma*: Individuals in a pair, group, community, organization, or society are faced with a choice between two alternative courses of action, one of which is *prosocial* (e.g., "cooperation") and the other selfish (e.g., "defection"), where the former imposes a greater direct cost or confers less benefit on the individual than the latter. The dilemma arises because each individual is by definition always better off behaving selfishly, but when all individuals do so, the collective outcome is worse for everyone than if prosocial behavior had prevailed. Several decades of mathematical modeling and analysis, laboratory experiments, field studies, and philosophical debates have yielded a variety of mechanisms by which prosocial behavior can arise, of which the following is but a partial list: reciprocity over repeated interactions (Axelrod 1984); group selection (Boyd et al. 2003) and so-called "strong reciprocity" (Bowles et al. 2003, Bowles and Gintis 2004); altruism as an observable signal of (unobservable) fitness (Gintis et al. 2001); reinforcement via stochastic learning (Kim and Bearman 1997, Macy and Flache 2002), social networks (Coleman 1988), or formal organizations (Bendor and Mookherjee 1987); and in public goods games, the specific shape of the production function (Oliver and Marwell (1985). All these proposals, however, focus exclusively on an individual's choice of actions with respect to their interaction partners, treating the choice of partners—the individual's social network—as exogenous.

The main contributions of this paper are (1) to extend the standard modeling framework to include partner choice (what we call *interaction dynamics*) as well as the usual action choice (*behavioral dynamics*) in an individual's repertoire of decisions; and, in particular, (2) to examine the effect of a *triadic closure bias* (Rapoport 1963)—the tendency of an individual to connect to a "friend of a friend"—on both interaction dynamics and behavioral dynamics. Specifically, we introduce and study a model of a multiperson prisoners' dilemma game in which agents interact locally with a small subset of partners defined by a sparse network. Agents not only learn from the behavior of others by imitating the behavior of the best-performing player they observe, but also create and sever relationships over time based on myopic cost-benefit comparisons. We show that the combination of network dynamics and learning can, under some circumstances, resolve what we call the *scalability problem*. The problem is that although decentralized cooperation may be possible in small groups, it becomes increasingly difficult to sustain in the face of free-riding (Boyd and Richerson 1988) as the group size increases. Thus, mechanisms that bring about cooperation will not necessarily scale with group size[1] (Boyd et al. 2003). By contrast, we show that when players' interaction partners evolve over time in a manner we define below, the fraction of cooperating players in the population tends to be higher in large networks.

We also present several other findings that suggest new, and in some cases counterintuitive, results affecting the fraction of cooperating players in large populations; in particular, that randomness in network dynamics and the lack of information regarding potential partners can have a positive impact on the level of cooperation. Detailed examination shows that there is an important trade-off between local reinforcement and global expansion in achieving cooperation in dynamic networks. Specifically, while unilateral tie severance and consensual tie creation strengthen the local reinforcement of cooperation, triadic closure bias hinders global expansion. These results may help us to understand phenomena such as the rapid growth of online markets like eBay. On the one hand, the feedback mechanism provided by eBay, as well as members' freedom to choose with whom to trade, function as local reinforcement mechanisms that promote cooperative members to interact among themselves. On the other hand, the existing members' apparent willingness to trade with new entrants whose previous, and therefore likely future behavior are unknown to them, acts as a global expansion mechanism. The remainder of this paper is organized as follows: §2 motivates the model, distinguishing it from related prior work; §3 describes the technical details of the model; §4 presents the results of our analysis; §5 discusses our findings; and §6 concludes.

[1] As Oliver and Marwell (1985) have pointed out, the scalability problem is not universal. It arises for some classes of public goods, but not for others. Specifically, if one assumes that the benefit b returned to each individual by a public good decreases with the size of the group N at a rate that is faster than any corresponding decrease in the cost c of contributing (typically models based on payoff matrices hold c fixed and assume that b decreases inversely with N), then invariably there will be a critical group size beyond which cooperation cannot be sustained. However, if either the benefit of the good is, for example, invariant with respect to group size (exhibiting what economists call "jointness of supply") or exhibits increasing returns to scale, then it is possible for cooperation to persist in groups of any size, and may in fact become more likely due to the increased aggregate resources of larger groups. In this paper, however, we concentrate on the problematic case outlined above, in which b decreases with N.

2. A Network Model of Cooperation

In this paper, we investigate the implications of the following two straightforward observations on individuals' behavior: (1) While cooperation necessarily entails interaction between individuals, all individuals do not interact equally with all others; rather, they can be thought of as interacting via a sparse network of social relationships. (2) Just as individuals alter their behavior subject to the logic of self-interest, they may also alter their choice of interaction partners over time.

Most previous work that considers cooperation on sparse interaction networks treats the network as exogenous with respect to the game being played, representing the network either as some kind of spatial lattice (Nowak and May 1992, Bergstrom and Stark 1993, Eshel et al. 1998) or, more recently, as a partly ordered, partly random network (Watts 1999a, b). These studies assume that the network in question remains fixed for the duration of the game, and is unaffected by it. In many social situations, however, individuals choose not only how to interact with others, but also with whom they interact. Furthermore, these two processes—what we call *behavioral dynamics* and *interaction dynamics*, respectively—coevolve in the sense that an individual's behavior is conditioned on the behavior of those with whom he is interacting (i.e., interactions affect behavior), and in turn his choice of partners will be conditioned on some assessment of their past or anticipated behavior (i.e., behavior also affects interactions).

While some limited work has explored endogenously generated relationships in the context of multiplayer games (Skyrms and Permantle 2000, Jackson and Watts 2002), it differs from our own, principally in that it focuses on coordination games, not social dilemmas. Furthermore, to the extent that networks are relevant to the results, quite different aspects of the networks in question are emphasized. In the case of Skyrms and Permantle (2000), the results actually concern the dynamics of pair formation, not networks in the sense usually intended by sociologists and, increasingly, other disciplines as well (Watts 2004). Jackson and Watts (2002) propose a model of interaction dynamics that is similar to ours in the sense that decisions both to create and sever ties are based on myopic cost-benefit comparisons. Our model differs, however, from theirs in that whereas Jackson and Watts focus on tie formation and termination exclusively between randomly chosen pairs, we explicitly incorporate features that are thought to be important to the evolution of social networks. In particular, we introduce *triadic closure bias* (Rapoport 1963, Granovetter 1973, Watts 1999a among others)—that is, the tendency for individuals to meet a "friend of a friend"—into the interaction dynamics. This bias has two important implications: (1) it makes the evolution of a network nonrandom, and (2) information regarding potential partners is available when an individual meets a "friend of a friend," whereas such information may not be available when an individual meets a stranger.

Because both these works are concerned with notions of collective behavior and also of networks that are substantively different from the kind we consider here, the relevant analyses, although quite general, cannot easily be extended to our case. Nevertheless, our findings are in agreement with the above studies in the general sense that they highlight not only the importance of the coevolution of networks and behaviors in determining possible outcomes, but also the relative speeds of the two modes of evolution (Skyrms and Permantle 2000). Furthermore, we concur with Jackson and Watts (2002) when they note (echoing an earlier warning of Oliver and Marwell 2001) that generalizations regarding prosocial behavior must be carefully qualified because the outcome tends to depend on at least some of the details of both how network and behavior are updated and, of course, the game itself.

We explore the coevolution of networks and collective behavior using a stochastic learning approach (Kim and Bearman 1997, Macy and Flache 2002) in which individuals attempt to optimize their behavior based on some limited memory of their past experience, but are otherwise myopic. Unlike forward-looking models of rationality, which also work on the principle of utility optimization, stochastic learning is a backward-looking approach, and thus assumes much lighter cognitive capabilities on the part of individuals than does traditional rationality. Furthermore, stochastic learning also lends itself naturally to a decision framework in which individuals must choose both actions and interactions in a mutually interdependent manner. That is, each individual evaluates his performance not only relative to his past performance, but also relative to the performance of his neighbors, where performance is now a function both of the payoff derived from participating in dyadic interactions, and also the cost of maintaining the interactions themselves. In keeping with the spirit of stochastic learning, individuals can sever relationships, but they do so selfishly and myopically, based solely on the relative costs and payoffs of each relationship.

3. Model Description

We consider a population of N players, each of whom repeatedly engages in a multiperson prisoners' dilemma game with an evolving subset of other players. We denote by $\Gamma_{i,t}$ the set of partners with whom player i interacts in period t $(=1, 2, 3, \ldots)$. We assume that players observe their own payoffs as well as both their partners' payoffs and strategies. Based on this information, players learn over time which strategy to use by occasionally mimicking the strategy of the best-performing player they observe (§3.1). Players are also allowed to change with whom they interact by selectively breaking ties or attempting to create them (§3.2). Players stochastically update their strategies and networks at the end of each time step with probabilities $\rho \in (0, 1)$ and $\mu \in (0, 1)$, respectively.[2]

3.1. Behavioral Dynamics

In each time step, each player i can choose either to cooperate C or to defect D with respect to its neighbors, where we assume that *players are restricted to using one strategy with respect to their entire neighborhood* (i.e., one cannot cooperate with one neighbor and defect with respect to another). We also assume that each member $j \in \Gamma_{i,t}$ is also in either one of the same two states; thus, the strategic environment faced by i can be described in terms of the potential payoffs experienced by i and its neighbors. These payoffs can be summarized in the standard manner by the payoff matrix

i, j	C	D
C	R, R	S, T
D	T, S	P, P

,

where i and j are the row and column players, respectively (i.e., if i cooperates and j defects, i receives payoff S, and j receives payoff T). Writing $\pi(a_i, a_j)$ as the

[2] An update probability of less than one implies that all updates are asynchronous, i.e., not everyone updates at the same time. See Huberman and Glance (1993) for a critique of models with synchronous updating.

payoff for player i using strategy a_i when partner j's strategy is a_j, then

$$\pi(C, C) = R, \quad \pi(C, D) = S,$$
$$\pi(D, C) = T, \quad \text{and} \quad \pi(D, D) = P.$$

For the game to embody a social dilemma, we require $T > R > P > S$ as well as $2R > (T + S)$. The first condition ensures that defection gives players a higher payoff regardless of what the opponent's strategy is. Therefore, if the game were to be played only once by rational players, who care only about their own material payoff, all players would defect and the resulting outcome would be socially suboptimal. The second condition implies that mutual cooperation is always better in the sense that it generates a higher aggregate payoff than all other cases.

Each period, players sum their payoffs over all interactions. To account for the burden of the interactions themselves, we reduce total payoffs by a quantity $\gamma(k)$, the total cost of interacting with k partners. The total net payoff (henceforth payoff) for player i obtained in period t is then

$$\Pi_{i,t} = \sum_{j \in \Gamma_{i,t}} \pi(a_{i,t}, a_{j,t}) - \gamma(k_{i,t}), \tag{1}$$

where $\gamma(k)$ is an increasing function of k. Here, we assume the specific form $\gamma(k) = ck^{\alpha}$, where $\alpha \geq 1$[3] and $0 \leq c \leq P$. When players update their strategies (with probability ρ), they copy the strategy of the player in their neighborhood (including themselves) who obtained the highest payoff in the last period.[4] For players who do not have any partners (such "isolates" inevitably appear as a result of the partner update process), their strategies are randomly set to cooperation or defection.[5] Thus, the requirement $c \leq P$ above, by allowing defectors to interact among themselves, prevents a spuriously high fraction of cooperators arising on account of defectors becoming isolated and then randomly switching to cooperation. Furthermore, we assume that all updates are subject to some error (or noise), that is, with a small probability ε, players adopt the opposite of what the rule specifies.

An important property of this construction is that in the case of all players interacting with each other equally (i.e., in an all-to-all network), and with the strategy update rule just described, defection will always generate a higher individual payoff, and so players will inevitably learn to defect. Thus, for cooperative behavior to survive at all in our model, interactions must be *local*; that is, players must interact with only a small subset of the population. It is also true that for cooperative behavior to survive, interactions should in some way promote assortative matching between cooperators (Bergstrom 2002), thus enabling cooperators to benefit from each other, while limiting defectors' opportunities to exploit them. Precisely how these properties can be generated by an interaction structure, and therefore which interaction structures represent optimal mechanisms for fostering cooperation, is an outstanding problem that we do not solve in full generality. Instead, what we will consider is a model of interaction dynamics that may, but will not necessarily, generate such interaction structures based on myopic processes of tie formation, severance, and learning.

3.2. Interaction Dynamics

At the end of each period, players also attempt to alter their local network with probability μ. Players randomly choose, with equal probability, either to break a tie with an existing partner or to create a tie with a new player. If their attempt fails, they then try the opposite action (e.g., if a player fails to break a tie, they then attempt to create one). Players can thus initiate at most one change to their local network,[6] but more than one change to a player's neighborhood can occur in a single period due to other players making their own changes.[7]

3.2.1. Termination of a Relationship. Both decisions—to break or create a tie—are based on myopic comparisons between (expected) marginal benefit and marginal cost of making the change. For the case in which player i attempts to break a tie, we assume that i first randomly chooses an existing partner j.[8]

[3] Although it is possible to set $\alpha < 1$, this choice would correspond to a diminishing marginal cost of friendship, which seems implausible. Furthermore, as we show in §4, even networks with $\alpha = 1$ generate little cooperation; thus, decreasing α further would not yield any additional insight.

[4] This assumption regarding updating players' behavior has been considered in a static network by Nowak and May (1992).

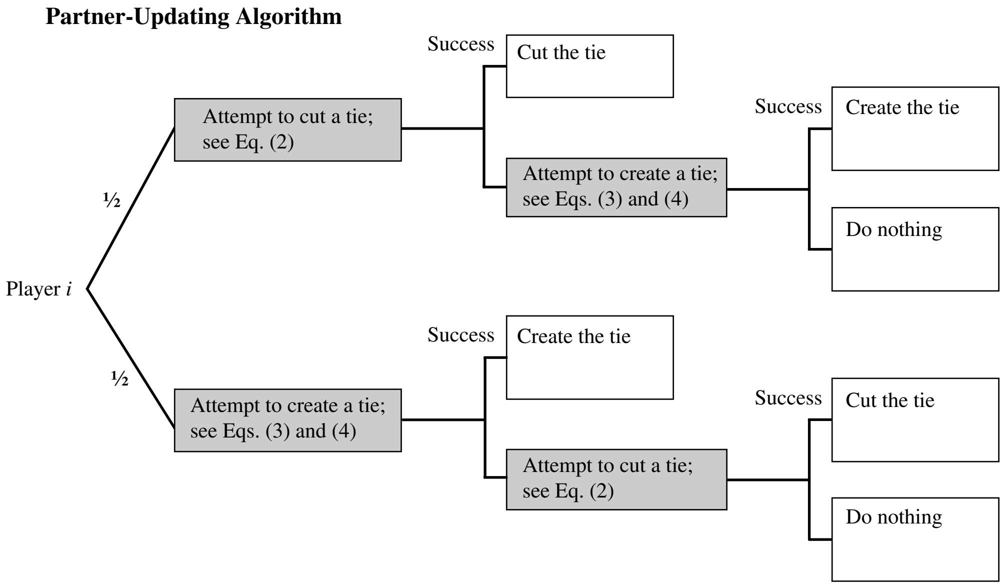
[5] We have also experimented with the alternative assumption that isolates do not change their strategies. The random alteration of isolates' strategies does not affect the main results presented here.

[6] Instead of players choosing whether to terminate a relationship or to create a new one with equal probability, we have also considered a model in which players always attempt one of the two first. The order in the ways in which players attempt to update their partners has a minor effect on the results: Models in which players always try to terminate a relationship before attempting to create a new one are slightly more likely to support cooperation than models in which the order is reversed.

[7] We require, however, that a relationship cannot be modified twice in the same period. For example, players cannot break a tie that has just been created by another player.

[8] We have also considered a stronger assumption that players always select a defecting partner to terminate. Such an assumption is conducive for sustaining a higher level of cooperation, although it weakens the expansion effect that we discuss in §5.

Figure 1 Interaction Dynamics



Notes. Each time period, a player decides with probability μ whether to cut or create a tie, choosing each with probability 1/2. If the player's chosen action fails, he tries the other, as shown. To cut a tie, a player proceeds by randomly choosing a partner and then comparing the cost and benefit of cutting that tie (Equation (2)). To create a tie, a player first decides whether to seek out a friend of a friend or a stranger with probabilities P_T and $(1-P_T)$, respectively. For the creation of any new tie, mutual consent is required so that the initiating player and the potential partner both perform the same cost-benefit calculations. In the case that a friend of a friend is sought out, a player chooses a candidate among his friends of friends according to a probability proportional to the number of shared mutual partners, and cost-benefit calculations (Equations (3) and (4)) are carried out under the full-information condition (Equation (5)). In the alternate case, a stranger is chosen randomly from the population. Cost-benefit calculations are carried out under the no-information condition based on players' trust (Equations (5) and (6)).

Player i then decides to terminate a tie with j if the gain from doing so is greater than the loss, i.e.,

$$(\gamma(k_i)-\gamma(k_i-1))-\pi(a_{i,t},a_{j,t})>0, \qquad (2)$$

where $\gamma(k_i)-\gamma(k_i-1)$ is the gain due to the reduction in interaction cost by having one less partner, and $\pi(a_{i,t},a_{j,t})$ is the expected loss from stopping the interaction. Here we assume that players naively believe that the partner will continue to act in the same manner, an assumption we maintain for the creation of a new tie as well. Termination is a unilateral decision, and no consent by the partner is required.[9]

3.2.2. Creation of a Relationship. Following Rapoport (1963) and others (Granovetter 1973, Watts 1999b), we view the creation of new network ties as a trade-off between two opposing forces: *order* and *randomness*. Order is defined in terms of a *triadic closure bias* (Rapoport 1963), i.e., the tendency of an individual to connect to a friend of a friend. Randomness means that new acquaintances are to be selected by drawing uniformly from the population at large. Specifically, we introduce a tunable bias parameter P_T, such that when player i seeks a new partner, with probability P_T a candidate is chosen from the set of i's partners' partners ($\bigcup_{j\in\Gamma_{i,t}}\Gamma_{j,t}\backslash\Gamma_{i,t}$); and with probability $1-P_T$, a candidate is instead chosen randomly from the population. When a candidate is chosen from the set of i's partners' partners, the probability of any particular player being chosen is proportional to the number of mutual partners shared by i and the player.[10] Regardless of their method of choosing each other, player i and the candidate j will commence a relationship if both their expected payoffs from interacting exceed their respective costs incurred by adding one more neighbor:

$$E(\pi(a_{i,t},a_{j,t}))-(\gamma(k_i+1)-\gamma(k_i))>0, \quad \text{and} \qquad (3)$$

$$E(\pi(a_{j,t},a_{i,t}))-(\gamma(k_j+1)-\gamma(k_j))>0, \qquad (4)$$

where in calculating the expected payoff, $E(\pi(a_{i,t}, a_{j,t}))$, players must make a prediction about their potential partner's future action. The decision rules implicit in the interaction dynamics are summarized in Figure 1.

3.2.3. Trust. When player i meets a friend of a friend j, we assume that i and j are informed by their mutual acquaintance of each other's most recent action; hence, i's expected payoff from the interaction

[9] With a small probability ε, the "incorrect" decision is made.

[10] In each period, we allow players to initiate at most one new tie, and the decision to create a tie is subjected to a small amount of noise.

is simply the payoff he would have obtained if he had interacted with j in the current period.[11] By contrast, when players meet randomly chosen strangers, they possess no information about the past behavior of their potential partner. Thus, when players interact randomly, their expected payoff must in part be a function of their *trust* in others ($\tau_{i,t} \in [0, 1]$), that is, their belief about the likelihood of others to cooperate. In these two circumstances—which we call the *full-information* and *no-information* cases, respectively—the expected gain for player i from interacting with a candidate player j becomes

$$E(\pi(a_{i,t}, a_{j,t})) = \begin{cases} \pi(a_{i,t}, a_{j,t}), & \text{full information,} \\ \tau_{i,t}\pi(a_{i,t}, C) + (1-\tau_{i,t})\pi(a_{i,t}, D), & \text{no information.} \end{cases} \tag{5}$$

In a state of universal defection, players should clearly hold little trust in each other, while when everyone cooperates, players can be fully trusting without fear of being exploited by defectors. The optimal level of trust therefore depends on the environment, where players have to learn how much they should trust strangers. To capture this behavior, we allow a player's trust to be updated every period as a weighted average of his previous level of trust, on the one hand, and his immediately preceding experiences with other players, on the other. The trust update rule for some player i is thus

$$\tau_{i,t} = \omega\tau_{i,t-1} + (1-\omega)R_{i,t-1}, \tag{6}$$

where $\omega \in (0, 1)$ is a weighting factor and $R_{i,t-1}$ is a measure of player i's experience with others.[12]

A final question remains, however: who the relevant other players are from whom an individual learns how trusting to be.[13] In this paper, we consider two alternate conditions: (1) *Open*: A player's trust is adjusted according to his experience with ongoing partners (friends) as well as new contacts (strangers), that is, $R_{i,t-1}$ is the fraction of i's *total* interactions that have been with cooperators in period $t-1$. (2) *Suspicious*: Trust depends only on a player's experiences in new encounters, that is, we define $R_{i,t-1}$ as the fraction of i's contacts *exclusively with strangers* that have been cooperators in period $t-1$.[14] All parameters in our model, and their associated ranges, are summarized in Table 1.

Table 1 **Summary of Model Parameters Along with Values Assigned for Experiments**

Parameter	Description (Range of values used in reported experiments)
N	Population size ($100 \le N \le 1{,}000$)
R	Rewarding payoff ($R = 1.0$)
S	Sucker's payoff ($S = 0.0$)
T	Temptation payoff ($1.0 \le T \le 2.0$)
P	Punishment payoff ($P = 0.1$)
ck^α	Cost of interacting with k partners ($0.0 \le c \le P$ (=0.1) and $1.0 \le \alpha \le 2.0$)
μ/ρ	Partner update rate/strategy update rate ($0.01 \le \mu/\rho \le 100$)
P_T	Triadic closure bias ($0.0 \le P_T \le 1.0$)
ω	Trust update rule weight ($0.0 \le \omega \le 1.0$)
ε	Error rate for strategy, partner, and trust update processes ($\varepsilon = 0.005$)

Notes. In addition, we have two conditions for trust evolution: (1) *Open condition*: A player's trust is adjusted according to his experience with ongoing partners (friends) as well as new contacts (strangers). (2) *Suspicious condition*: Trust depends only on a player's experiences in new encounters.

4. Results

On account of the relative complexity of our model, and thus its limited analytical tractability, we have adopted an experimental approach to study its behavior, exploring different parameter configurations via numerical simulation. In all simulations, each member of the population is assumed to start with no partners, and is assigned a random initial state (cooperate versus defect, each with probability 0.5) and trust level (uniform on $[0, 1]$). We let each simulation run for 10,000 units of time, thus allowing the initial transient behavior to run its course, and then record average statistics (e.g., level of cooperation, network density, etc.) over a subsequent 5,000 periods, which we take to be representative of the population's long-run state.

4.1. Scalable Cooperation

Our first concern is to explore the level of cooperation in our model as a function of the population size N, recalling that in traditional models of social dilemmas (i.e., in the absence of network evolution), cooperation levels diminish as N increases. Figure 2 displays, for certain specific choices of parameter values,

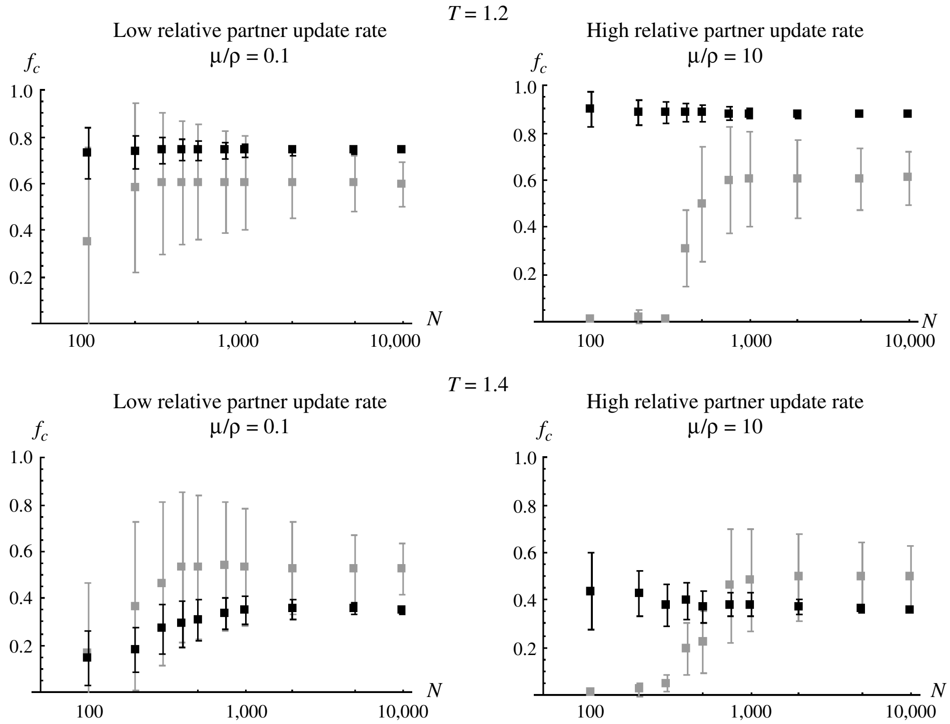
[11] We assume, as we do in the termination of relationships, that each player naively believes others will continue with their current strategy. We also have experimented with a stronger assumption, i.e., when a player meets a friend of friend, instead of meeting a randomly chosen one, he selectively chooses a cooperating player among them. This stronger assumption, however, does not qualitatively change the main results presented in this paper.

[12] With a small probability ε, $\tau_{i,t}$, is set to a random value between zero and one.

[13] If a player is isolated, we leave their trust unchanged, i.e., we set $R_{i,t-1} = \tau_{i,t-1}$.

[14] Because new relationships can be instigated, there can be more than one relationship created in period $t-1$ that involves player i even though player i himself can initiate the change at most once. If there have been no new relationships in the period, then $R_{i,t-1}$ is not defined. In this case, we assume that $R_{i,t-1} = \tau_{i,t-1}$.

Figure 2 **Steady-State Fraction of Cooperators, f_c, as a Function of Population Size N for Various Values of Temptation T and Relative Update Rate μ/ρ**



Notes. Higher network cost results ($\gamma(k) = 0.05k^2$) are in black, and lower network cost results ($\gamma(k) = 0.025k^{1.75}$) are in gray. For all plots, the triadic closure bias $P_T = 0.5$, the weighting for the past trust and recent experience is equal ($\omega = 0.5$), and trust evolves under the suspicious condition.

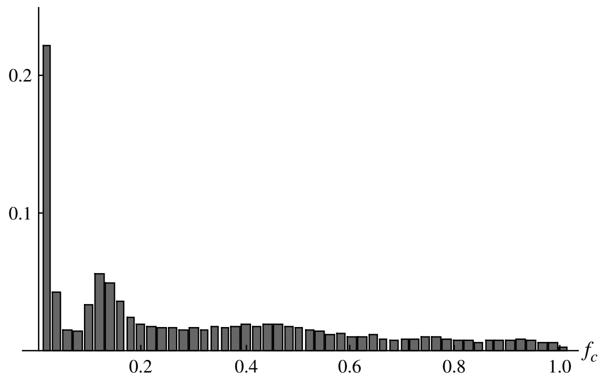
the fraction of cooperators as a function of increasing population size. For small populations ($N = 100$), our model generates low levels of cooperation, but as N increases, the level of cooperation rises dramatically before appearing to reach an asymptote around $N = 1{,}000$. Thus, the choice of $N = 1{,}000$ seems sufficiently high to capture the behavior of large populations, but not so high that we cannot run the required number of simulations. We can also conclude that the fraction of cooperators in very large evolving networks appears to be invariant with respect to the population size; that is, the most favorable scaling relationship possible. Therefore, at least for some parameter values of our model, high levels of cooperation are not only possible in very large populations, but actually appear to benefit from increasing population size. In the remainder of this paper, we attempt to specify more precisely which parameter ranges are more conducive to cooperation, and why.

4.2. Cooperation Requires Sparse Networks

Figure 3 displays a discrete probability distribution of steady-state cooperation levels for a fixed population size $N = 1{,}000$ for 10,000 different combinations of the parameters $\{T, \mu, \rho, \alpha, c, P_T\}$ and our two trust evolution rules, where the combinations are sampled uniformly[15] from the full ranges of all the parameters (as summarized in Table 1). There are two striking features of this distribution: (a) there is a large spike at zero cooperation, and (b) there is a long tail that extends over the entire range of possible steady-state values. These observations suggest that without any restrictions on the parameter space, the introduction of interaction dynamics cannot on its own consistently generate high levels of cooperation. Here we note that varying the tie cost parameters c and α through their respective ranges is analogous to moving from a part of the parameter space in which new ties can be added to the network virtually free of cost (i.e., when $c \simeq 0$) to one in which the cost of adding a new tie rapidly exceeds even the maximum benefit that an additional relationship can deliver (effectively imposing an upper limit on the degree of nodes). Thus, by varying c and α, we can examine the effect of increasing tie cost and, indirectly, network density. This effect can be seen in Figure 4, which splits the frequency distribution of Figure 3 into two distributions, corresponding to high-cost ($1.5 < \alpha \leq 2.0$ and $0.01 < c \leq 0.1$) and low-cost ($1.0 < \alpha \leq 1.5$ or $0 \leq c \leq 0.01$)

[15] For μ/ρ, we first choose either μ or ρ with equal probability from the range [0.1, 0.5]; the other is then chosen from the range $[0.01\xi, \xi]$ (where $\xi \in \{\mu, \rho\}$). For each set of randomly drawn parameter values, we run two separate simulations, corresponding to the open and suspicious conditions, respectively.

Figure 3 **Distribution of the Steady-State Fraction of Cooperators f_c over 10,000 Randomly Sampled Parameter Sets**

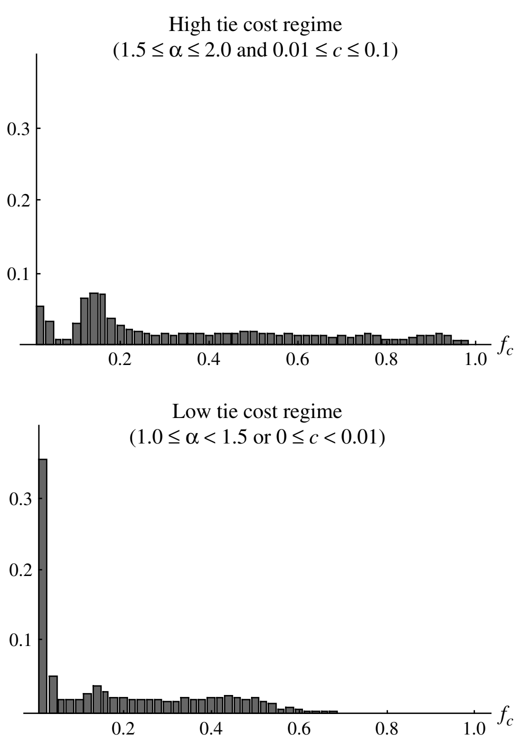


Note. Average cooperation level $\langle f_c \rangle = 0.274$.

regimes, respectively. The average level of cooperation in the high network cost regime is much higher than in the low-cost regime, with the latter possessing the majority of the zero cooperation outcomes observed in Figure 3. Thus, the sustainability of cooperation in our model depends not only on the network being dynamic, but also being sparse.

Because the effect of very low network cost is so pronounced, the other network-related effects in our model are substantially minimized in low-cost regimes. These effects, however, remain of interest once we restrict our attention to high-cost (i.e., sparse) dynamic networks. We explore more systematically the variation in steady-state levels of cooperation generated by our model over the parameter space sampled within the high-cost regime by using regression analysis. Although regression analysis is relatively uncommon in simulation studies, it is not unprecedented (Marwell et al. 1988, Kim and Bearman 1997), and is useful in this case for the same reason that it is useful in analyzing empirical data: because it can be applied to high-dimensional data for which the underlying model is unknown or (as in this case) intractable, and can identify general trends that are not easily discernible to the eye in straightforward plots or cross-tabulations. We emphasize that because the behavior of our model is, in general, nonlinear, our conclusions must be interpreted with some caution, and should only be considered valid in a qualitative sense. As we show below, however, our qualitative conclusions do hold up under a simple robustness check. The output of our regression analysis is summarized in Table 2, in which we consider three main dependent variables of interest: (1) the average proportion of cooperating players in the population; (2) the volatility of cooperation, i.e., the standard deviation of the fraction of the cooperating players; and (3) the average trust in the population.

Figure 4 **Distribution of Steady-State Fraction of Cooperators f_c for High Tie Cost Regime ($1.5 \leq \alpha \leq 2.0$ and $0.01 \leq c \leq 0.1$) and Low Tie Cost Regime ($1.0 \leq \alpha < 1.5$ or $0 \leq c < 0.01$), Respectively**



Note. Average levels of cooperation and number of observations are, for the high-cost regime, $\langle f_c \rangle = 0.374$ and 4,490, and, for the low-cost regime, $\langle f_c \rangle = 0.192$ and 5,510.

Table 2 **Result of Regression Analysis: High Network Cost Regime ($1.5 \leq \alpha \leq 2.0$ and $0.01 \leq c \leq 0.1$) Only**

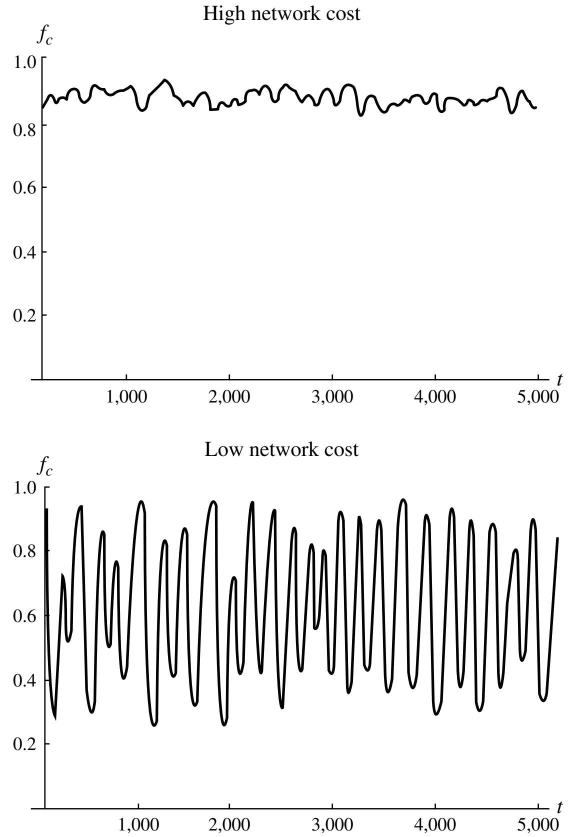
Independent variables	Cooperation	Volatility	Trust
Temptation (T)	−0.829	−0.065	−0.637
	(0.005)	(0.003)	(0.005)
Network cost 1 (c)	0.316	−1.963	−0.035*
	(0.056)	(0.029)	(0.054)
Network cost 2 (α)	0.048	−0.277	0.021*
	(0.010)	(0.005)	(0.010)
Relative network update rates (μ/ρ)	0.0022	−0.0011	0.0016
	(0.0002)	(0.0001)	(0.0002)
Triadic closure bias (P_T)	−0.110	−0.006*	−0.027
	(0.005)	(0.003)	(0.005)
Weight on current trust (ω)	0.044	0.034	0.131
	(0.005)	(0.003)	(0.005)
Open trust evolution[(1)]	−0.051	−0.006	0.086
	(0.003)	(0.001)	(0.003)
Constant	1.563*	0.771*	1.238
	(0.020)	(0.010)	(0.019)
R-squared	0.74	0.46	0.66

Notes. Number of observations is 10,000 (5,000 realizations for each of the two trust evolution conditions). Parameters are randomly sampled from the range shown in Table 1, except for the restriction placed on network costs: $1.5 \leq \alpha \leq 2.0$ and $0.01 \leq c \leq 0.1$.

[(1)]The dummy variable takes value one (zero) if trust evolves under open (suspicious) condition.

*Not statistically significant at the 1% level.

Figure 5 **Examples of Dynamic Fluctuations in Cooperation for High Network Cost** ($\gamma(k) = 0.05k^2$) **and Low Network Cost** ($\gamma(k) = 0.025k^{1.75}$)



Note. For these simulations, $P_T = 0.5$ (partial triadic closure), the trust condition is suspicious, low temptation payoff is low ($T = 1.2$), the relative partner update rate μ/ρ is 10, and the weight in the trust update rule is one-half ($\omega = 0.5$).

These results are all taken from the last 5,000 periods of our simulations.

4.3. Positive Effect of Tie Cost

The positive effect of tie cost on the level of cooperation (demonstrated above) is also confirmed in the regression analysis, with positive coefficients for α and c. In addition, cooperation is less volatile when ties are costly, as shown by the negative coefficients for α and c in the volatility column of Table 2. This result is also shown in Figure 5, which shows the time evolution of the fraction of cooperators for two network costs. In both time series,[16] there are periods where almost all players are cooperating; however, when interaction costs are lower, the population experiences repeated "crashes" in the level of cooperation. We explain these cooperation crashes at lower network costs by recalling that players occasionally make a mistake when updating their strategy; that is, there is a small chance ε that players choose the opposite of what the strategy update rule specifies. This "random shock" is present in the higher network cost case as well. The more partners some defector i has, however, the greater the benefit of defection is to i, and the less likely it is that all of i's cooperating partners will successfully break their ties with i before they themselves "learn" (from i) that defection delivers a higher payoff, and begin defecting. Thus, a single defection spreads more rapidly and easily in highly connected networks (in particular from highly connected individuals), and is harder to contain than in poorly connected networks. By two measures, therefore—average cooperation and volatility around the mean—high-cost, hence sparse, networks support cooperation.[17] This result may also explain why higher cooperation can be achieved in a larger population: For a given interaction cost, the larger the population size, the sparser the network becomes.

4.4. Positive Effect of Updating Partners

Table 2 also shows that more frequent partner updating tends to generate higher levels of cooperation (i.e., the coefficient for μ/ρ is positive), indicating that unilateral termination of (or players walking away from) unproductive relationships is effective in sustaining global cooperation. This conclusion agrees with earlier findings by, for example, Orbell and Dawes (1993), and is also analogous to strategies such as ostracism that are often employed in models of group selection. In general, the ability of players to terminate an undesirable relationship can be seen as a punishment strategy that can support cooperation (Fudenberg and Tirole 1991) by lowering the excluded partner's payoff. However, in this case the punishment aspect of cutting ties is at best indirect because—unlike in the case of ostracism—any player excluded from one relationship can avoid being punished simply by choosing new partners. Furthermore, our implementation of stochastic learning prevents individuals from responding deterministically to punishment; hence, even if being excluded from some particular interaction can be construed as punishment, it does not mean that the excluded individual will necessarily learn to cooperate. In fact, as we discuss in §5, some additional mechanism for recruiting new cooperators is also required. Thus, partner updating, although in some ways analogous to punishment, is more subtle

[16] These time series are taken from the last 5,000 periods of simulation.

[17] The higher average proportion of cooperating players does not necessarily mean that the population average payoff is higher. In particular, when the network cost is high, the payoff net of interaction cost can be lower even though there are more cooperating players in the population. Nevertheless, the average proportion of cooperating players and the population average payoff are positively correlated with the coefficient of correlation being 0.48.

and indirect, but is still apparently an effective mechanism for encouraging cooperation.

"Walking away" is also critical to the scalability of cooperation (i.e., its persistence as network size increases), not only because it acts as a way of punishing defectors, albeit indirectly, but because as an enforcement mechanism it too is scalable. Because individuals need only make decisions about the partnerships in which they themselves are participating, and from which they personally stand to gain or lose, then the burden associated with monitoring and enforcement (i.e., tie severance) grows only with the *degree* (number of neighbors) of each node; thus, it is invariant with respect to the size of the population. Unilateral tie severance is therefore not only an effective way of maintaining scalable cooperation, but also one that is simpler than other sanctioning systems, such as ostracism, which typically require either group-level coordination or else the resolution of a second-order dilemma problem (Oliver 1980).

4.5. Negative Effect of Triadic Closure Bias

In light of Coleman's (1988) observation that triadic closure in social networks should facilitate cooperative behavior, and subsequent simulation results (Watts 1999b) that support the same conclusion in a static network, perhaps our most surprising result is that when new contacts are made via referrals from existing partners—that is, through triadic closure—the average global fraction of cooperators is less than when individuals choose new partners randomly from the population at large (i.e., the coefficient of P_T in Table 2 is negative).

There are two possible effects of triadic closure bias that could account for this result: a *structural effect*, and an *informational effect*. The former comes from the observation that triadic closure generates networks with high clustering coefficients (Watts and Strogatz 1998), and the latter comes from the assumption that when meeting friends of friends, individuals have full information about one another, and no information when meeting random strangers. We note that neither effect seems a likely candidate for impeding global cooperation, as local clustering is generally thought to improve the potential for cooperation in a static network (i.e., through local reinforcement), and full information of the previous actions of others enables individuals to better avoid "bad risks." While these intuitive arguments are plausible, neither turns out to be true in our model, with the opposite effect pertaining in each case. After presenting these additional results, we will discuss the reason behind these counterintuitive results in §5.

We first separate the structural and informational effect of triadic closure bias by running the model *with* and *without* information transmission about potential partners who are currently "a friend of a friend" (i.e., without information transmission, the past behavior of all potential partners is unknown). In the regression analysis reported in Table 3, we introduce a dummy variable I, which takes the value one for experiments with information transmission and zero, otherwise. The coefficients of P_T and $P_T \times I$ together, $-0.092 - 0.066 = -0.158$, capture the effects of triadic closure bias—both structural and informational—on the dependent variable. The result shows that approximately two-fifths of the negative effect of triadic closure derives from better information being provided about potential partners, while the structural effect (i.e., local clustering) accounts for the other three-fifths. Thus, both the information and network structure aspects of triadic closure impede cooperation.

Table 3 **Triadic Closure With and Without Information Transmission**

Independent variables	Cooperation	Volatility	Trust
Temptation (T)	−0.843	−0.056	−0.644
	(0.004)	(0.002)	(0.004)
Network cost 1 (c)	0.124	−1.705	−0.183
	(0.043)	(0.018)	(0.040)
Network cost 2 (α)	0.007*	−0.249	−0.004*
	(0.008)	(0.003)	(0.007)
Relative network update rates (μ/ρ)	0.0027	−0.001	0.0012
	(0.0002)	(0.0001)	(0.0002)
Triadic closure bias			
Structural effect (P_T)	−0.092	−0.027	−0.042
	(0.004)	(0.002)	(0.004)
Information effect[(1)] ($P_T \times I$)	−0.066	0.032	−0.003*
	(0.004)	(0.002)	(0.004)
Weight on current trust (ω)	0.016	0.020	0.125
	(0.004)	(0.002)	(0.004)
Open trust evolution[(2)]	−0.126	−0.016	0.038
	(0.002)	(0.001)	(0.002)
Constant	1.747	0.700	1.233
	(0.015)	(0.007)	(0.014)
R-squared	0.72	0.45	0.64

Notes. The number of observations is 20,000 (5,000 realizations for each of the two trust evolution conditions and information conditions). Parameters are randomly sampled from the range shown in Table 1, except for the restriction placed on network costs: $1.5 \leq \alpha \leq 2.0$ and $0.01 \leq c \leq 0.1$.

[(1)]The dummy variable I takes value one (zero) if expected payoff is calculated under full (no)-information condition when players meet friends of friends.

[(2)]The dummy variable takes value one (zero) if trust evolves under the open (suspicious) condition.

*Not statistically significant at the 1% level.

The informational effect apparent in triadic closure can also be investigated in isolation from the structural effect by studying a variant of our model without triadic closure bias (i.e., with $P_T = 0$), and comparing cases in which individuals updating their (randomly chosen) partners are always informed of a potential partner's previous action (full information) with those in which they never receive such information (no information). As Table 4 shows, the

Table 4 Information Transmission Without Triadic Closure ($P_T = 0$)

Independent variables	Cooperation	Volatility	Trust
Temptation (T)	−0.771	−0.045	−0.621
	(0.008)	(0.004)	(0.008)
Network cost 1 (c)	0.586	−1.495	0.656
	(0.093)	(0.039)	(0.088)
Network cost 2 (α)	0.007*	−0.237*	−0.007*
	(0.017)	(0.007)	(0.016)
Relative network update rates (μ/ρ)	0.0010	−0.0013*	0.00003*
	(0.0004)	(0.0002)	(0.00004)
Information effect(1)	−0.091	0.030*	0.177*
	(0.005)	(0.002)	(0.005)
Weight on current trust (ω)	−0.010*	0.008*	0.048
	(0.008)	(0.003)	(0.008)
Open trust evolution(2)	−0.108	0.001*	−0.085
	(0.005)	(0.002)	(0.005)
Constant	1.645	0.631*	1.278
	(0.034)	(0.014)	(0.032)
R-squared	0.71	0.43	0.68

Notes. The number of observations is 4,000 (1,000 realizations for each of the four combinations of trust evolution and information conditions). Parameters are randomly sampled from the range shown in Table 1, except for the restriction placed on network costs: $1.5 \leq \alpha \leq 2.0$ and $0.01 \leq c \leq 0.1$, as well as triadic closure bias being zero.

(1)The dummy variable I takes value one (zero) if expected payoff is calculated under full (no)-information condition.

(2)The dummy variable takes value one (zero) if trust evolves under the open (suspicious) condition.

*Not statistically significant at the 1% level.

effect of having information about previous behavior is the same strength and sign as in the comparisons involving triadic closure bias, that is, in our model, *lack of information* about potential contact in deciding whether to start interacting has a *positive* effect on the level of cooperation, regardless of the network structure.

4.6. Positive Effect of Suspicion

Because populations in which individuals have no information about strangers appear to sustain higher levels of cooperation than populations in which they have full information, it is interesting to investigate, for the no-information case, the effect of varying trust between strangers on cooperation. Not surprisingly, we find that open communities, in which the distinction between friends and strangers is overlooked, develop higher levels of trust than suspicious communities, in which individuals compute trust levels solely in terms of past interactions with strangers ("Trust" column in Tables 2 and 3). Surprisingly, however, cooperation is much higher in suspicious communities than in open ones (Tables 2–4)—to such an extent, in fact, that the average level of trust in a suspicious community is lower than it would be if it were based on the actual average behavior of the population.

How can trust be lower in suspicious communities that at the same time achieve a higher level of cooperation? The answer depends on the players' tendency to maintain ties with cooperating partners and sever ties with defectors. Consequently, defectors are more likely than cooperators to be looking for new partners, and so when players are seeking partners, contacts with strangers are more likely to be with defectors than with other cooperators. Suspicious players who base their trust solely on past encounters with strangers will therefore estimate the likelihood of encountering a cooperator more accurately than open players who, by including their experiences with ongoing partners, will tend to overestimate it. A lower trust in suspicious communities, in turn, results in a higher level of cooperation because it limits the number of partners players have (cf. Equation (5)). As we have seen above in the discussion on the positive effect of interaction costs, a higher level of cooperation can be achieved when players have a smaller number of partners. Our results therefore suggest an intriguing contrast between information about the past behavior of strangers and trust in their future behavior: Limited information appears to enhance cooperation, but so does limited trust.

4.7. A Simple Robustness Check

Considering the strong nonlinearity generally prevalent in networks, one might reasonably suspect the robustness of these results, based as they are on linear regression analysis. It is possible, for example, that triadic closure bias has varying effects on the level of cooperation, depending on the proportion of cooperating players in the population, taking a negative value when most of the players are cooperating and a positive value when the majority of players are defecting.[18] Given the potential for nonlinear behavior, we have performed a simple robustness check on our conclusions by separating our data into four subsets (based on the average proportion of cooperating players in the population, $\langle f_c \rangle$), and running the same basic regression that we reported in Table 2 separately for each subset. Table 5 reports the results of this exercise. The table shows that the coefficient of triadic closure bias is negative and significant for all the subsets except for the case reported in the second column, $0.25 < \langle f_c \rangle \leq 0.5$, where the coefficient is not significantly different from zero. Notably, the fits (as measured by R^2) are lower than those of the original regressions, which use the entire sample; nevertheless, they suggest that in spite of the possible presence of nonlinearity in the behavior of the model, our conclusions regarding the negative effect of triadic closure are robust—at least qualitatively. Similar

[18] We thank an anonymous referee for pointing out this possibility.

Table 5 **Result of Regression Analysis for Four Different Ranges of Average Proportion of Cooperating Players in the Population, $\langle f_c \rangle$**

Independent variables	Average fraction of cooperating players in the population			
Subsample	$0.0 < \langle f_c \rangle \leq 0.25$	$0.25 < \langle f_c \rangle \leq 0.5$	$0.5 < \langle f_c \rangle \leq 0.75$	$0.75 < \langle f_c \rangle \leq 1.0$
Temptation (T)	−0.88	−0.25	−0.28	−0.33
	(0.004)	(0.01)	(0.01)	(0.015)
Network cost 1 (c)	0.95	−0.30	0.86	0.33
	(0.03)	(0.06)	(0.06)	(0.06)
Network cost 2 (α)	0.01*	−0.03	0.12	0.09
	(0.006)	(0.01)	(0.01)	(0.01)
Relative network update rates (μ/ρ)	0.0008	−0.001	0.0008	0.002
	(0.0002)	(0.0002)	(0.0003)	(0.0002)
Triadic closure bias (P_T)	−0.01	0.002*	−0.066	−0.013
	(0.003)	(0.005)	(0.006)	(0.005)
Weight on current trust (ω)	0.03	−0.0002*	−0.032	−0.007*
	(0.003)	(0.005)	(0.005)	(0.005)
Open trust evolution[(1)]	−0.002*	−0.025	−0.048	−0.009
	(0.002)	(0.003)	(0.003)	(0.003)
Constant	0.19	0.94	0.78	1.05
	(0.01)	(0.02)	(0.02)	(0.02)
R-squared	0.25	0.28	0.28	0.31
Number of observations	4,771	2,013	1,797	1,419

Note. Parameters are randomly sampled from the range shown in Table 1, except for the restriction placed on network costs: $1.5 \leq \alpha \leq 2.0$ and $0.01 \leq c \leq 0.1$.
[(1)]The dummy variable takes value one (zero) if trust evolves under open (suspicious) condition.
*Not statistically significant at the 1% level.

conclusions also apply to the other parameters of our model.

5. Discussion

While some of our results regarding the emergence of cooperation in dynamic networks are surprising, we are nevertheless able to account for them, at least qualitatively, in terms of a trade-off between two conflicting forces: *local reinforcement* and *global expansion*. Both forces have clear implications: Mechanisms that enhance local reinforcement tend to result in assortative matching of cooperators, which is well known to support cooperation (Axelrod 1984, Bergstrom 2002); and mechanisms that promote global expansion of cooperation clearly raise the level of cooperation, at least to the extent that they do not undermine local reinforcement. What is less clear is how myopic processes of network evolution can generate both effects simultaneously, especially as they *do* tend to undermine one another. That is, expansion requires cooperators to interact with defectors (thus diminishing assortativity), and local reinforcement discourages exploration (thus impeding expansion). Investigating this trade-off uncovers two mechanisms at work, both of which depend on the evolution of network ties: (1) exclusion as a means of reinforcing cooperative behavior, and (2) recruiting of defectors as a means of expanding cooperation.

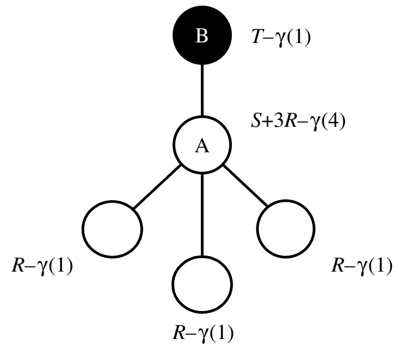
5.1. Exclusion as Local Reinforcement

For static networks, one way to bring about assortative matching is to build in local network structure (Coleman 1988, Watts 1999b), such that if a cooperator is interacting with other cooperators, those cooperators will also interact preferentially with others, thus creating a self-reinforcing cluster of cooperation capable of resisting outside attack. However, in dynamic networks—where individuals can make and break ties—assortativity can be generated even in networks with no local structure through pure pairwise exclusion. Furthermore, our results suggest that pairwise exclusion alone, based on unilateral severance and consensual tie creation, generates enough assortativity that structure-generating processes like triadic closure, even if they further enhance assortativity, do not do so sufficiently to overcome their debilitating impact on expansion.

5.2. Expansion via Recruitment of Defectors

The second key insight required to understand the expansion of cooperative behavior is that defectors can be recruited by direct exposure to cooperating partners, in spite of the apparent temptation to defect. The mechanism is best understood graphically, as displayed in Figure 6. At time t, a new tie is formed between a cooperator, who we assume is connected to other cooperators, and a defector, who we assume is otherwise isolated. Because Player A—the white

Figure 6 A Schematic of the Recruitment of a Defector by a Cooperator



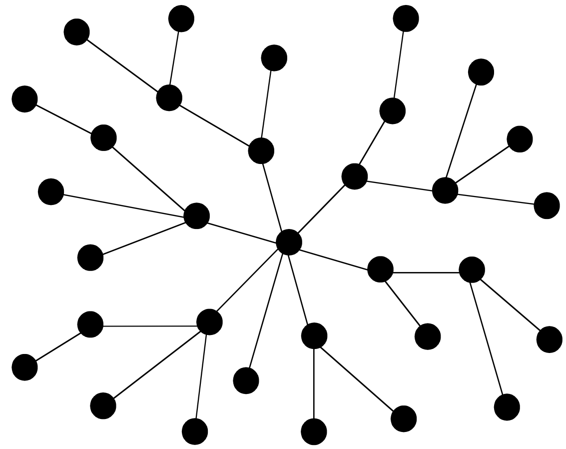
Notes. White nodes are cooperators and the black node is currently a defector. According to the payoff matrix described in §3.1, the node in the center (A) will receive a combined payoff of $S + 3R - \gamma(4)$, while the other cooperators will each receive $R - \gamma(1)$ and the lone defector (B) will receive $T - \gamma(1)$. While the defector thus outscores three of the four cooperators, he is exposed only to one, who by virtue of his other interactions receives a greater payoff. Thus, the defector is persuaded that cooperation pays better than defection, and is thereby converted.

player in the center of the figure—receives payoffs from four interactions, while Player B—the black player at the top of the figure—receives a payoff from only one interaction, then B observes that A's payoff exceeds his own, even though A's average payoff per interaction is lower than B's. Furthermore, B has no other partner with whom to compare A's payoff. Thus, when B has an opportunity to revise his behavior, he will switch to cooperation—a rational decision in the sense that B is merely adopting a strategy observed to deliver greater payoffs than his own. Note, however, that if B were also connected to A's other neighbors—that is, if local clustering were present—B would outscore all others, and so would continue to defect.

5.3. The Importance of Sparseness

For cooperation to survive, it is therefore critical (a) that individuals do not interact equally with all others (i.e., that the network be sparse globally); and (b) that the partners with which an individual interacts do *not* all interact with each other (i.e., that the network be sparse locally as well). The reason is twofold. First, as discussed in §3.1, if a defector and a cooperator have the same set of interaction partners, then the defector will *always* outscore the cooperator, and so everyone will eventually learn to defect. It matters, in other words, that individuals learn from others who do not, in turn, learn from each other. Thus, cooperation in evolving networks fares better in networks with low clustering (Watts 1999a), i.e., networks rich in structural holes, and scales well with the network size because large networks are more likely to be sparsely connected than small ones. Second, high clustering may also compromise cooperation by making networks harder to disconnect. Random networks, for example, exhibit a local structure that looks like a branching tree, as shown in Figure 7. The absence of cycles (i.e., clustering) renders the network vulnerable to disconnection by severing only a few ties—a feature that is generally viewed negatively because such networks are not robust (Albert et al. 2000, Callaway et al. 2000, Dodds et al. 2003), but which can be desirable if what is spreading is (collectively) undesirable.

Figure 7 Locally, a Large, Sparse, Random Network Will Resemble a Pure Branching Network, as Shown Here; Thus, Defectors Can Be Isolated Easily Through Dyadic Exclusion (Tie Severance)



5.4. The Importance of Heterogeneity

Coupled with global sparseness and low network clustering, the success of cooperation as a strategy depends on cooperators having, on average, more partners than defectors, and thus being able to profit from more interactions. The combination of varying connectivity and learning is permitted by the dynamic nature of the network ties: Because cooperators will tend to sever ties with defectors, then (a) cooperators will be more likely than defectors to interact with other cooperators, thus supporting assortative matching, and in turn local reinforcement of cooperators; and (b) defectors will be more likely than cooperators to be isolated (or have low degree), thus inclining them to imitate, and hence be recruited by, successful cooperators who "accidentally" link to them. As demonstrated in §4, however, while differences in connectivity are central to sustaining and expanding cooperation, too much heterogeneity can be counterproductive. When, for example, the cost of creating additional ties is low, individuals who are initially cooperators will tend to accrue large numbers of partners, at which point their benefit from defecting will be correspondingly large. Therefore, if a highly connected individual defects, even rarely (say by a random fluctuation), their high connectivity will have the greatest possible impact on other individuals (because they expose many others to their negative influence, and also because they are hard to isolate), thus destabilizing global cooperation as well.

6. Conclusion

In recent years, both the collective action and the networks literatures have considered, on the one hand, the evolution of different kinds of behavior on static networks (May et al. 1995; Watts 1999a, b; Strogatz 2001; Newman 2002); and on the other hand, the evolution of networks both with (Skyrms and Permantle 2000, Jackson and Watts 2002) and without (Albert and Barabasi 2002) strategic behavior. As yet, however, the coevolution of networks and behavior has not received as much attention as it deserves. In this paper, we have investigated a specific, but important, case of this very general problem area by introducing, simulating, and analyzing a dynamic network model of a multiplayer social dilemma in which players can modify both their actions and their interactions through stochastic learning.

Our model admits a number of insights that appear to surmount the details of the model itself. Our main result is that cooperation can persist in sparse, dynamic networks of effectively unlimited size, and in fact tends to fare better in large networks than in small ones. Furthermore, if players imitate others' behavior, networks in which an individual's interaction partners do not themselves interact—that is, networks that are rich in structural holes (Burt 1992)—appear to support higher levels of cooperation than networks in which local density (or clustering, Watts 1999a) is high. This last finding is precisely the opposite of the standard intuition regarding cooperation and network structure.

Although counterintuitive, most of our results, including the main one of scalable cooperation, can be understood in terms of a trade-off between two effects: *expansion versus reinforcement*. Roughly speaking, this principle states that the maintenance of a high global level of cooperation requires that two conditions be satisfied: (a) cooperation is reinforced by assortative matching of cooperators; and (b) the cooperative community must expand by "recruiting" defectors. However, these conditions tend to work against each other, as the former works, in effect, by shielding cooperators from defectors, while the latter works precisely by exposing cooperators to defectors, anticipating that the latter will learn to cooperate as well. Thus, overall cooperation can only be maximized by seeking some trade-off between the two. A relatively high rate of tie updating, for example, along with a high cost of adding ties and a suspicion of strangers, are all reinforcement mechanisms, whereas ignorance of strangers and a preference for random interactions over referrals by mutual friends both favor expansion.

In this paper, we have considered only a very simple behavioral rule—players can either cooperate or defect against all of their partners—and our description of the social network formation process is equally simplistic; thus, many obvious extensions are easily conceivable. As we have seen, however, even such a simple model can generate considerable complexity at the level of collective dynamics—a result that suggests such extensions and complications should be approached with caution. Nevertheless, the coevolution of interaction and behavioral dynamics does appear to have new and counterintuitive consequences for collective outcomes, and thus is worthy of further development.

Acknowledgments

The authors gratefully acknowledge the financial support of the National Science Foundation (SES 0094162 and SES 0339023), the Office of Naval Research, the McDonnell Foundation, and Legg Mason Funds.

References

Albert, R., A.-L. Barabasi. 2002. Statistical mechanics of complex networks. *Rev. Modern Phys.* **74**(1) 47–97.

Albert, R., H. Jeong, A.-L. Barabasi. 2000. Attack and error tolerance of complex networks. *Nature* **406** 378–382.

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

Bendor, J., D. Mookherjee. 1987. Institutional structure and the logic of ongoing collective action. *Amer. Political Sci. Rev.* **81**(1) 129–154.

Bergstrom, T. C. 2002. Evolution of social behavior: Individual and group selection. *J. Econom. Perspect.* **16** 67–88.

Bergstrom, T. C., O. Stark. 1993. How altruism can prevail in an evolutionary environment. *Amer. Econom. Rev.* (*Papers Proc.*) **83** 149–155.

Boorman, S. A., P. R. Levitt. 1980. *The Genetics of Altruism*. Academic Press, New York.

Bowles, S., H. Gintis. 2004. The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoret. Population Biol.* **65**(1) 17–28.

Bowles, S., E. Fehr, H. Gintis. 2003. Strong reciprocity may evolve with or without group selection. *Theoret. Primatology Project Newsletter* **1**.

Boyd, R., P. J. Richerson. 1988. The evolution of reciprocity in sizable groups. *J. Theoret. Biol.* **132** 337–356.

Boyd, R., H. Gintis, S. Bowles, P. J. Richerson. 2003. The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. USA* **100**(6) 3531–3535.

Burt, R. S. 1992. *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.

Callaway, D. S., M. E. J. Newman, S. H. Strogatz, D. J. Watts. 2000. Network robustness and fragility: Percolation on random graphs. *PRL* **85** 5468–5471.

Coleman, J. S. 1988. Free riders and zealots: The role of social networks. *Sociol. Theory* **6** 52–57.

Dodds, P. S., D. J. Watts, C. F. Sabel. 2003. Information exchange and the robustness of organizational networks. *Proc. Natl. Acad. Sci. USA* **100**(21) 12516–12521.

Eshel, I., L. Samuelson. 1998. Altruists, egoists, and hooligans in a local interaction model. *Amer. Econom. Rev.* **88** 157–179.

Fehr, E., U. Fischbacher. 2003. The nature of human altruism. *Nature* **425** 785–791.

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

Gintis, H., E. A. Smith. 2001. Costly signaling and cooperation. *J. Theoret. Biol.* **213** 103–119.

Granovetter, M. S. 1973. The strength of weak ties. *Amer. J. Sociol.* **78** 1360–1380.

Huberman, B. A., N. S. Glance. 1993. Evolutionary games and computer simulations. *Proc. Natl. Acad. Sci. USA* **90**(16) 7716–7718.

Jackson, M. O., A. Watts. 2002. On the formation of interaction networks in social coordination games. *Games Econom. Behav.* **41** 265–291.

Kim, H., P. S. Bearman. 1997. The structure and dynamics of movement participation. *Amer. Sociol. Rev.* **62** 70–93.

Macy, M. W., A. Flache. 2002. Learning dynamics in social dilemmas. *Proc. Natl. Acad. Sci. USA* **99**(Suppl. 3) 7229–7236.

Marwell, G., P. E. Oliver. 1988. Social networks and collective action: A theory of the critical mass. III. *Amer. J. Sociol.* **94** 502–534.

May, R. M., S. Bohoeffer, M. A. Nowak. 1995. Spatial games and evolution of cooperation. *Proc. Third European Conf. Adv. Artificial Life* **929** 749–759.

Newman, M. E. J. 2002. Spread of epidemic disease on networks. *Physical Rev. E* **66**, article 016128.

Nowak, M. A., R. M. May. 1992. Evolutionary games and spatial chaos. *Nature* **359** 826–829.

Oliver, P. E. 1980. Rewards and punishments as selective incentives for collective action: Theoretical investigations. *Amer. J. Sociol.* **85** 1356–1375.

Oliver, P. E., G. Marwell. 1985. A theory of the critical mass. I. Interdependence, group heterogeneity, and the production of collective action. *Amer. J. Sociol.* **91**(3) 522–556.

Oliver, P. E., G. Marwell. 2001. Whatever happened to critical mass theory? A retrospective and assessment. *Sociol. Theory* **19**(3) 292–311.

Orbell, J. M., R. M. Dawes. 1993. Social welfare, cooperators' advantage, and the option of not playing the game. *Amer. Sociol. Rev.* **58**(6) 787–800.

Ostrom, E., J. Burger, C. B. Field, R. B. Norgaard, D. Policansky. 1999. Revisiting the commons: Local lessons, global challenges. *Science* **284** 278–282.

Rapoport, A. 1963. Mathematical models of social interaction. R. D. Luce, R. R. Bush, E. Galanter, eds. *Handbook of Mathematical Psychology*, Vol. 2. Wiley, New York, 493–579.

Skyrms, B., R. Permantle. 2000. A dynamic model of social network formation. *Proc. Natl. Acad. Sci. USA* **97**(16) 9340–9346.

Strogatz, S. H. 2001. Exploring complex networks. *Nature* **410** 268–276.

Watts, D. J. 1999a. Networks, dynamics, and the small-world phenomenon. *Amer. J. Sociol.* **105** 493–527.

Watts, D. J. 1999b. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press, Princeton, NJ.

Watts, D. J. 2004. The "new" science of networks. *Annual Rev. Sociol.* **30** 243–270.

Watts, D. J., S. H. Strogatz. 1998. Collective dynamics of "small-world" networks. *Nature* **393** 440–442.