Cooperation in Anonymous Dynamic Social Networks

Nicole Immorlica* Brendan Lucier† Brian Rogers*

Abstract

We study the emergence of cooperation in dynamic, anonymous social networks, such as in online communities. We examine prisoner's dilemma played under a social matching protocol, where individuals form random links to partners with whom they can interact. Cooperation results in mutual benefits, whereas defection results in a high short-term gain. Moreover, an agent that defects can escape reciprocity by virtue of anonymity: it is always possible for an agent to abandon his history and re-enter the network as a new user. We find that cooperation is sustainable at equilibrium in such a model. Indeed, cooperation allows an individual to interact with an increasing number of other cooperators, resulting in the formation of a type of social capital. This process arises endogenously, without the need for potentially harmful social enforcement rules. Additionally, for a rich class of parameter settings, our model predicts a stable coexistence of cooperating and defecting agents at equilibrium.

1 Introduction

There are many important social, political, and economic systems in which people face a choice between seeking immediate personal gain at the expense of others or cooperating to a lesser mutual benefit. In such contexts, it might be expected that an agent acting in his own best interest ought to choose an uncooperative strategy. This intuition is amplified when people can act under pseudonyms, such as over the Internet, since an agent that develops a reputation for being uncooperative can, if he chooses, simply re-enter the system with a new identity. However, the continuing success of online interaction networks such as eBay (www.ebay.com) indicates that a group of anonymous agents need not devolve into a steady state of completely uncooperative behavior. There is a large body of experimental evidence indicating that cooperation is prevalent [2, 10, 13]. In fact, many systems with a high degree of anonymity and in which agents change partners over time are characterized by a high, but less than complete, level of cooperation.

We find such behavior can be explained as constituting a steady-state equilibrium under a simple random-matching network formation model in which agents enter into persistent bilateral relationships. The key element of our model is the ability for agents to sever these relationships at will. First, this enables us to model anonymity: assuming agents have free (or cheap) access to pseudonyms, and they can always sever all their relationships and begin afresh under a new identity. Thus no agent can be forced to bear the consequences of a negative reputation. Second, due to anonymity and the resulting lack of persistent reputation, severing a relationship is the *only* mechanism via which agents can punish others for uncooperative behavior. The threat of this punishment induces cooperative behavior at equilibria as only through cooperation can agents

*Northwestern University, Evanston, IL

†University of Toronto, Toronto, ON

attract a large network of profitable relationships. The model further supports the coexistence of cooperation and defection when the gains from accumulated relationships that a cooperator expects are equal to the gains achievable by perpetually defecting on a smaller set of partners.

Our model contributes to the long line of work seeking to explain the robust phenomenon of cooperation in repeated social interactions. When the partners in the game are fixed over time and the game is repeated indefinitely, a classic application of the Folk Theorem provides a vehicle for cooperation (or any mutually beneficial payoff) [1, 6], via, e.g., trigger strategies, and can be extended to accommodate imperfect monitoring [5]. However when agents change partners over time or are anonymous, such a threat is no longer effective because a pair of agents may very well never meet again. Instead, community enforcement procedures can be used to sustain cooperation [4, 8, 11]. Such enforcement procedures typically do not provide an explanation for the stable coexistence of cooperative and defective behavior and are highly sensitive to deviations.

In the above models, partnerships are formed exogenously, be they fixed or random. When agents have choice in their partners or in the length of the relationship, new insights arise [3, 7, 9, 12, 14, 15]. In these models, cooperation is sustained because the threat of losing a relationship is costly. The cost can come in many forms, such as being cast into a matching market with friction, having to start a new relationship that requires specific investment, or having to start small in a new relationship. We also employ such an approach: agents' behavior influences network dynamics and so the reason to cooperate comes from the fact that, through cooperation, one can gradually build up a social network of other cooperators. However, unlike other related work, our model sustains, in *stable* equilibrium, a *heterogeneous* society with non-trivial fractions of both cooperating and defecting agents. Moreover, cooperation in our model arises endogenously without the use of potentially costly social enforcement protocols or inefficient trust-building behavior.

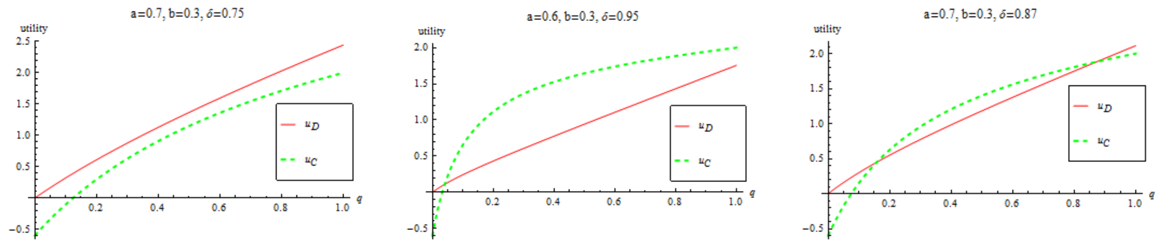
2 Model

There is a countable set of agents with finite lifetimes distributed according to a geometric random variable with mean $1/\delta$. These agents interact with each other on a directed network with fixed out-degree K (i.e., at any given time each agent sponsors K connections to other agents). An agent is generally involved both in relationships that it sponsors (outlinks) and also in relationships sponsored by others (inlinks). In each relationship, agents play a prisoner's dilemma with the following payoff matrix.

	C	D
C	1,1	-b,1+a
D	1+a,-b	0,0

We take $a, b > 0$ and $a - b < 1$ so that, while mutual cooperation is the uniquely efficient outcome, defection is strictly dominant. Relationships subsist throughout time until one of the partners dies or severs the connection. Time is discrete. Each time period proceeds as follows:

1. **Actions.** Each agent i chooses an action $\alpha_i \in \{C, D\}$. We assume that agents observe the aggregate proportion q of C behavior in the population when taking their action.
2. **Outlinks.** Each agent i with out-degree $d_i < K$ proposes $K - d_i$ new relationships with partners chosen uniformly at random from the population.
3. **Inlinks.** Proposed inlinks are accepted or rejected by receiving partner.



(a) (b) (c)

Figure 1: Utility curves corresponding to different patterns of equilibrium occurance. (a) Only the all-defection state is an equilibrium. (b) The all-cooperate and all-defect states are at equilibrium, and there is an unstable equilibrium for some $q \in (0, 1)$. (c) The all-defect state is an equilibrium, as well as two interior equilibria: the rightmost stable, the leftmost unstable.

4. **Payoffs.** Each agent receives a payoff equal to the sum of the outcomes of the PD game played with each of his connections, according to the chosen actions of the two agents and the payoff matrix given above.
5. **Severance.** Agents sever any links that they choose to.
6. **Birth and death.** Each agent dies with a given probability $1 - \delta$ in which case it is replaced with a new agent.

Agents seek to maximize the present value of expected lifetime payoffs.

3 Equilibria in a Restricted Strategy Space

How will rational agents choose their strategies in this model? We begin by analyzing a restricted strategy space in which agents are assumed to be *unforgiving*, *consistent*, and *trusting*. Agent are *unforgiving* if they always sever relationships with defectors in favor of new random partners (and always maintain relationships with cooperators). Agents are *consistent* if they commit to a strategy at birth and never change. Agents are *trusting* of strangers if they always accept proposed links.

Our first result is to derive an exact characterization of equilibrium actions under these assumptions. In particular, under these assumptions, our model supports cooperation in the society. Since relationships to defectors are broken, an agent can only build up a network of relationships by committing to cooperation. The rate at which this happens is a (nonlinear) function of the fraction of cooperators and defectors in the society. These incoming links from cooperators become a valuable asset, which can be thought of as the social capital of the agent. If the expected lifetime of an agent is sufficiently high, the promise of this asset may induce selfish agents to commit to cooperation even though the per-period-per-interaction payoff for cooperation is lower than that of defection. When the expected lifetime utility of defecting equals that of cooperating, the model supports in equilibrium a heterogeneous society in which cooperators and defectors co-exist. We also show that some of these equilibria are stable in that, should the fraction of cooperators grow (or shrink) beyond the equilibrium mixture, then newborn agents will prefer to defect (or cooperate) and thus self-correct the composition. When this happens is a function of the parameters of our model; Figure 1 shows the three possibilities and the following proposition describes the general conditions under which each occurs.

Proposition 1 *The all-defection state ($q = 0$) is always an equilibrium. The remaining equilibria are as follows:*
If $a > 1$, then:

(i) if $b < 2$ and δ is sufficiently large, there exist two interior equilibria: one stable and one unstable, with the stable equilibrium occurring with more cooperators.

(ii) otherwise ($b \geq 2$ or δ not large enough) there is only the $q = 0$ equilibrium.

If $a < 1$, then:

(iii) if δ is sufficiently large then the $q = 1$ state is an equilibrium, and there will exist an unstable internal equilibrium.

(iv) if δ is sufficiently small then only the $q = 0$ state is an equilibrium.

(v) if $b < a(1 + a)$, then there exists an intermediate range of δ for which there are two interior equilibria: one stable, and one unstable.

The intuition behind the results of Proposition 1 comes from observing that defectors and cooperators gain utility in very different ways. A defector gains utility directly by interacting with cooperators and exploiting them. Their links do not persist from one turn to the next. Thus the per-period utility of a defector is (practically) proportional to the fraction of cooperators in the system. Parameter a dictates the rate at which utility increases with cooperators. By contrast, a cooperator gains utility by building a network of relationships from which he can extract utility over his lifetime. Given sufficient time, the neighborhood of a cooperator limits to a critical size, at which point the rate of decay of existing friends matches the rate of finding other cooperators. A major factor in the payoff of a cooperator is the amount of time necessary to approach this critical neighborhood size, relative to the expected lifespan. This quantity is influenced by the number of cooperators in the system, but this influence suffers diminishing returns: when there are few cooperators present, a small increase will have large effects on the number of cooperators expected to meet each other; when there are many cooperators, they will all likely reach their critical neighborhood sizes, and thus the addition of more cooperators has little effect. The utility of a cooperator will also be affected by the losses he incurs from interacting with defectors. This is a linear effect, proportional to the number of defectors in the system. Parameter b determines the rate at which utility decreases with the number of defectors.

In summary, stable interior equilibria occur whenever (a) defecting is preferable to cooperating in a world where all agents cooperate, and (b) when defectors enter a mostly-cooperator system the rate of decay of defector utility is greater than the rate of decay of cooperator utility. Condition (b) generally requires that parameter b not be too large, and that δ not be too small. Condition (a) requires either that parameter a be large, or that δ not be too large.

4 Equilibria in General Strategies

We return to the general space of strategies, without imposing our three assumptions. We demonstrate that, subject to mild conditions on our parameters, the equilibria in which agents are unforgiving, consistent, and trusting are also equilibria in the general strategy space. Note first that if agents are consistent, then it is equilibrium behavior for them to be unforgiving (since defection in a partner indicates that no further cooperative benefit can be gained from the interaction). We next prove that if agents are trusting, it is an equilibrium for them to be consistent for a wide range of parameters.

Proposition 2 *Assume that* $\frac{1+b}{1-(1-q)\delta^2} \geq 1 + a$*. Then consistent strategies constitute equilibria under fully general strategies, given the norm of severing relationships with defectors.*

Finally, we prove that the social norm of trusting strangers is in fact equilibrium behavior for a wide range of parameters (e.g. when the marginal gain from defection is sufficiently small), assuming that agents are consistent. Note that this is equivalent to demonstrating that each agent obtains positive expected utility from accepting a proposed link.

Proposition 3 *Suppose that all agents are consistent and either* $q > 2/3$*, or* $a < 1$ *and* $q \in (0, 1]$ *corresponds to a stable equilibrium under the norm of being trusting. Then a cooperator obtains positive expected utility from accepting a proposed link.*

To summarize, we prove that, for a wide range of parameters, the equilibria that result when we assume that agents are trusting, unforgiving, and consistent are also equilibria in the general strategy space. In the full version we show how equilibrium behavior changes when the conditions of Proposition 2 and Proposition 3 are not met. In particular, consistency is replaced with behavior that allows for burning social capital, and trusting is replaced with exclusivity behavior in which cooperating agents reject proposed inlinks with some probability, thereby inducing a form of cliquishness among cooperators.

References

[1] D. Abreu. On the theory of infinitely repeated games with discounting. *Econometrica*, 56(2):383–396, 1988.

[2] R. Axelrod. *The Evolution of Cooperation.* Basic Books, New York, 1984.

[3] S. Datta. Building trust. Mimeo, Indian Statistical Institute, 1993.

[4] G. Ellison. Cooperation in the prisoner's dilemma with anonymous random matching. *The Review of Economic Studies*, 61(3):567–588, 1994.

[5] D. Fudenberg, D. Levin, and E. Maskin. The folk theorem with imperfect public information. *Econometrica*, 62(5):997–1039, 1994.

[6] D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, 1986.

[7] P. Ghosh and D. Ray. Cooperation in community interaction without information flows. *The Review of Economic Studies*, 63(3):491–519, 1996.

[8] M. Kandori. Social norms and community enforcement. *Review of Economic Studies*, 59:63–80, 1992.

[9] R. Kranton. The formation of cooperative relationships. *Journal of Law, Economics, and Organization*, 12(1):214–233, 1996.

[10] L. Lave. An empirical approach to the prisoners' dilemma game. *The Quarterly Journal of Economics*, 76(3):424–436, 1962.

[11] M. Okuno-Fujiwara and A. Postlewaite. Social norms and random matching games. *Games and Economic Behavior*, 9:79–109, 1995.

[12] G. Ramey and J. Watson. Bilateral trade and opportunism in a matching market. *Contributions to Theoretical Economics*, 1(1), 2001. Article 3.

[13] A. Rapoport and A. Chammah. *Prisoner's Dilemma.* University of Michigan Press, 1965.

[14] C. Shapiro and J. Stiglitz. Equilibrium unemployment as a worker discipline device. *American Economic Review*, 74(3):433–444, 1984.

[15] J. Sobel. A theory of credibility. *Review of Economic Studies*, 52:557–573, 1985.