"Show Me Your Friends and I'll Tell You Who You Are": A Computational Model of Reputation by Association in Noniterated Social Dilemmas

Reidar Hagtvedt
Georgia Institute of Technology School of Industrial and Systems Engineering
Interuniversity Consortium on Negotiation and Conflict Resolution
hagtvedt@ terry.uga.edu

Gregory Todd Jones
Georgia State University College of Law
Interuniversity Consortium on Negotiation and Conflict Resolution
Max Planck Institute for Research on Collective Goods
gtjones@gsu.edu

Abstract

In the context of the evolution of social behavior, an action is altruistic when it enhances the relative reproduction of another at some cost to the relative reproduction of the actor. The evolution of altruism has been explored in countless studies employing the well-known prisoner's dilemma, but plausible theory is made more difficult when exchanges are between unrelated actors in circumstances where enforcement institutions are lacking and where the unlikelihood of a repeat encounter takes the teeth out of reciprocity. This paper proposes a novel social embeddedness-based strategy (SEnS) that is successful in bringing about the evolution of cooperation in a noniterated social dilemma, without the adaptive improbability of projection strategies or the demanding cognitive overhead of detection strategies.

Contact:
Reidar Hagtvedt
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Tel: 404-877-1154
Fax: 404-463-9789
Email: hagtvedt@ terry.uga.edu

Key Words: Game Theory; Altruism; Noniterated Prisoner's Dilemma; Reputation

Acknowledgement: Much of this work was initiated while G. T. Jones was Visiting Research Scholar at the Max Planck Institute for Research on Collective Goods, Bonn, Germany. Their kind hospitality is gratefully acknowledged. The authors would like to thank Doug Yarn, Christoph Engel, and the members of the Max Planck Institute for Research on Collective Goods Rationality Working Group for stimulating discussions and many helpful comments. We particularly appreciate the insightful remarks of Martin Hellwig that planted the seed for this direction.

"Show Me Your Friends and I'll Tell You Who You Are": A Computational Model of Reputation by Association in Noniterated Social Dilemmas

Reidar Hagtvedt and Gregory Todd Jones

The General Problem of Altrusim

In the context of the evolution of social behavior, an action is altruistic when it enhances the relative reproduction of another at some cost to the relative reproduction of the actor [Sober & Wilson, 1999]. Social exchanges often involve circumstances in which an actor is faced with a choice between cooperation (engaging in altruistic behavior) or defection (cheating, or withholding altruistic behavior) without advance knowledge of how the other actor may behave. The dilemma, of course, is that in many situations, individually rational behavior may lead to collective irrationality [Kollock, 1998]. The evolution of cooperation has been explored in countless studies employing the well-known prisoner's dilemma [Axelrod, 1984] in which two actors make a dichotomous choice between cooperation and defection where: 1) defection is dominant, that is, each actor is privately better off defecting regardless of what the other actor does, and 2) there is a deficient equilibrium at mutual defection, that is, aggregate social welfare is maximized with mutual cooperation and is at its lowest with mutual defection, but there is no individual incentive for cooperation [Orbell & Dawes, 1993]. Standard solutions to this problem have involved cooperation directed towards kin [Hamilton, 1964a, 1964b], reciprocity [Trivers, 1971], and institutional controls [Heckathorn, 1993; Hechter, 1987].

Meeting Strangers in the Noniterated Social Dilemma

What about exchanges between unrelated actors in circumstances where enforcement institutions are lacking and where the unlikelihood of a repeat encounter takes the teeth out of reciprocity? As Macy and Skvoretz so elegantly put it:

> Not all exchanges . . . involve familiar faces or third-party regulation. In the dark alleys of social life the future does not cast a shadow. There strangers meet outside the watchful eye of a Leviathan capable of enforcing compliance with a negotiated agreement. In such unregulated exchanges, rapscallions can renege with impunity, without fear of future retaliation, loss of reputation, or enforcement agencies. These conditions pose the prisoner's dilemma in its purest form, as players are stripped of all institutional or structural protection against exploitation [Macy & Skvoretz, 1998].

In *Evolution of the Social Contract*, Brian Skyrms proposed what at first glance seemed to be an idea so simple as to render it trivial [Skyrms, 1996]. Skyrms suggested that if interactions could somehow be *correlated*, that is, in a population with equal representation of cooperators and defectors, if cooperators somehow (and he explicitly stated that he did not care how) were to interact with other cooperators with a probability even slightly higher than cooperators with defectors, then the selective advantage for cooperators would drive the defectors to extinction and cooperators to fixation.

Exit Strategies

One way a cooperator may insure more frequent interactions with other cooperators is to simply walk away from potential interactions with actors anticipated to be defectors. Such "exit strategies" are not new [Hirschman, 1970] and have, in fact, been explored in the context of prisoner's dilemma games [Orbell & Dawes, 1993; Yamagishi & Hayashi, 1996]. While studies have demonstrated that cooperation in prisoner's dilemma games increases where players have the option to exit [Boone & Macy, 1999], in general, the demands in actor cognitive capacity have been enormous [Sheratt & Roberts, 1998; Macy & Skvoretz, 1998; for a recent exception, see Joyce et al., 2006]. In previous research, two methods for estimating another actor's trustworthiness in noniterated interactions have been proposed: 1) *projection* of an actor's intended behavior to other parties to a potential exchange [Orbell & Dawes, 1991], and 2) *detection* of other parties intentions, possibly by the reading of signals of some type [Orbell & Dawes, 1991; Frank, 1988, 1993]. However, Orbell and Dawes admit that the evolution of projection strategies is difficult to explain [1991, p. 525], and detection strategies pose a huge adaptive overhead in cognitive capacity.

Reputation by Association

In this research, we propose an extremely straightforward procedure that relies only upon the well established mechanism of operant conditioning [Thorndike, 1911; Skinner, 1938] to bring about the social embeddedness anticipated by Macy & Skvoretz [1998] and Yamagishi & Hayashi [1996] in a spatialized prisoner's dilemma [Grim, 1995] without any significant requirement for cognitive capacity. Put simply, a primary actor, chosen from the population at random, encounters a group of secondary actors ($0 \leq n \leq 4$, a von Neumann neighborbood on a toroidal ecology). The primary actor chooses one of these secondary actors at random (if $n \geq 1$) and plays a single round of a standard prisoner's dilemma game. If in this first encounter, the primary actor encounters a cooperator, the primary actor chooses one of the other secondary actors who is a neighbor of the secondary actor first engaged and likewise plays a single round. This process continues with the primary actor playing a maximum of *n* single-round games (if all secondary actors encountered cooperate). If the primary actor encounters a defector, further play is not engaged in and the actor moves. The theory is simple. Given the ability to move away from defectors, the existence of a defector provides information to the primary actor about the defector's neighbors. Show me your friends, and I tell you who you are.

The Simulations

The simulations we report here were developed in Matlab, whose innate matrix capabilities offer an efficient means to manage parameter-rich torii [Thorngate, 2000]. A 64 x 64 torus is sparsely populated with 1000 actors with random location, initial adaptive scores allocated from a uniform distribution from 0 to 100, and equal starting proportions of five strategies: two strategies that always cooperate and either stay put (CS) or move to another location when encountering defection (CM), two strategies that always defect and either stay put (DS) or move when encountering defection (DM), and a novel strategy that we have dubbed SEnS, a social embeddedness-based strategy that always cooperates, and continues interaction within a von Neumann neighborhood when cooperation is encountered, but ceases play and moves when defection is encountered. None of these strategies require significant cognitive capacity (there is no memory requirement).

In each generation, there are 20,000 (such that the expected number of games per actor is 20) single-shot prisoner's dilemma games in which an actor is chosen at random and play proceeds with actors within the primary actor's neighborhood. Adaptive scores are summated based upon a standard prisoner's dilemma payoff matrix, with the exception that we impose a more severe penalty for cooperating in the face of defection in order to compensate for artificial inflation of the SEnS strategy's adaptive score due to the potential for multiple interactions per game.

	Cooperate	Defect
Cooperate	3,3	-3,5
Defect	5,-3	0,0

After 20,000 games, actors with adaptive scores less than or equal to zero are culled and replaced without mutation by another strategy with likelihoods equal to the relative proportion of each strategy in the current population. This is repeated for 200 generations. Costs are assessed at 100 units per actor per generation, which can be interpreted as a "cost-of-living" or aging, and 10 units for moving, which was accomplished initially by a random search away from the defector, with a finite sequential search proceeding in the same direction should the initial location be occupied. Here, if a mover is unable to locate an empty location within five attempts, the actor stays put.

Discussion & Future Research

Our preliminary results indicate that our social embeddedness-based strategy (SEnS) is successful in bringing about the evolution of cooperation in a noniterated social dilemma. And SEnS does so without the adaptive improbability of Orbell & Dawes projection strategy [1991] or the demanding cognitive overhead of Frank's detection strategy [1988, 1993]. However, SEnS does not offer the immediate gratification promised by Skyrm's correlated association [1996], largely due to the necessity of a sparse population to facilitate movement. Very quickly, the population reaches a stasis (See Figure 1) in which clusters of cooperation exist[1] among a sea of prowling defectors, forever looking for opportunities to exploit. See Figure 2. The conditions under which it may be possible to drive defectors to extinction and cooperators to fixation, if possible at all, remain a topic for future research.

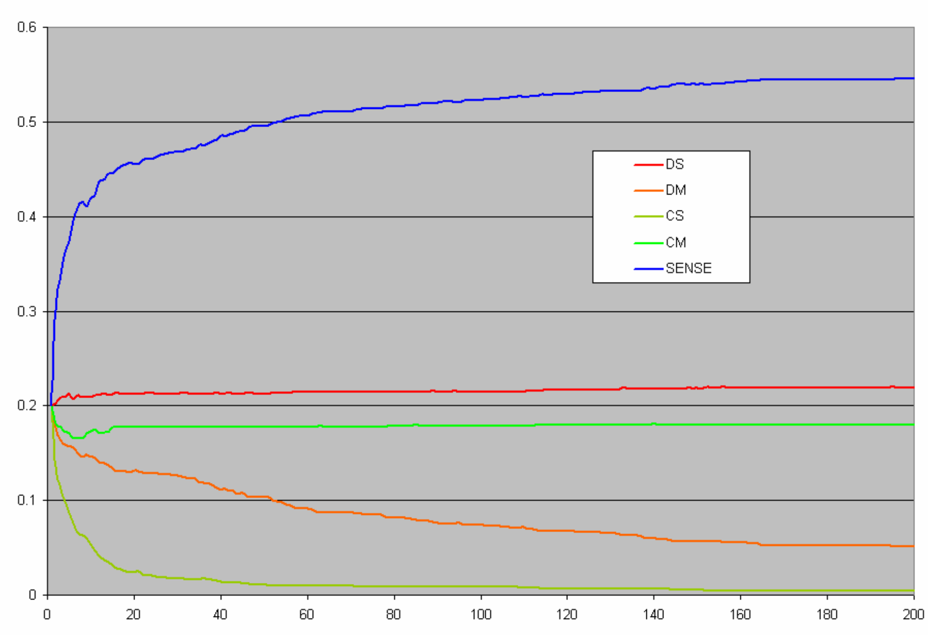

Figure 1: X-axis: Generation, Y-axis: Proportion of population. Red = Defect/Stay, Orange = Defect/Move, Lime = Cooperate/Stay, Green = Cooperate/Move, Blue = SEnS (social embeddedness strategy).

We would expect even stronger clustering if we add a probability of moving that is inversely proportional to the number of secondary actors in a primary actor's neighborhood. Further, we would expect that results will be highly sensitive to the density of the population, where more actors offer more opportunities for correlated association, but where too many actors produce congestion that inhibits movement. One possible exploration of these dynamics involves the elimination of the artificiality of the toroidal architecture and the implementation of social networks that would allow us to use existing metrics to compare clustering and neighborhood or clique formation in differing scenarios.

[1] A chi-square test for proportions was employed to compare the degrees of clustering among actors after the 200 generation simulation with a uniform distribution. Results confirm what can be ascertained visually in Figure 2. The hypothesis that the SEnS actors remained uniformly distributed was rejected with a p-value of 3.93 E-10.

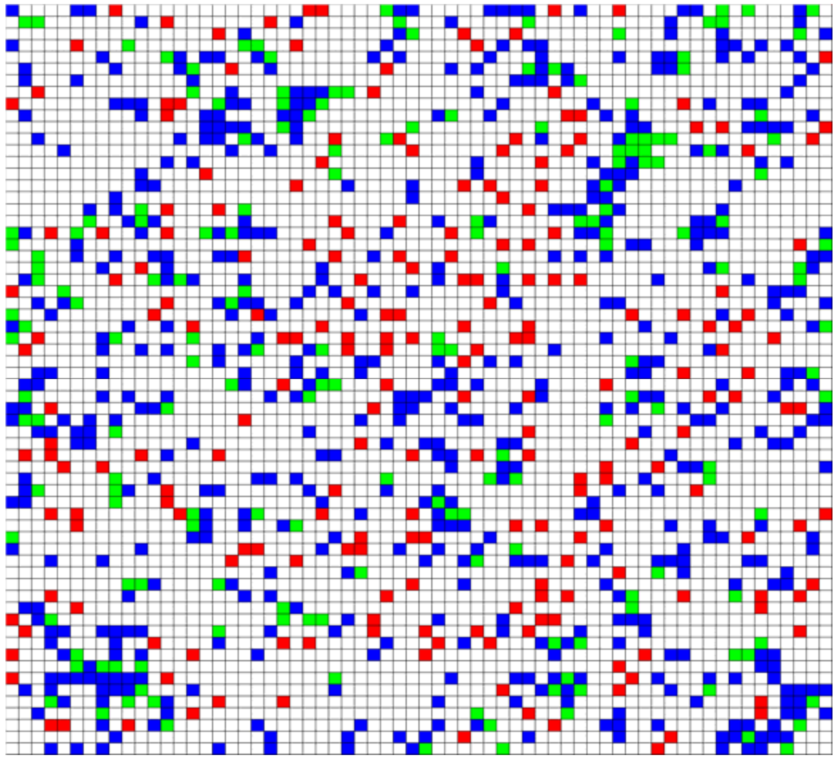

Figure 2: 64 x 64 toroidal ecology, sparsely populated randomly with 1000 agents, at generation 200, 20,000 games per generation. White = empty, Green = Cooperate/Move, Red = Defect/Stay, and Blue = SEnS (social embeddedness strategy). (Small numbers of Cooperate/Stay and Defect/Move strategies omitted for clarity).

References

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

[Boone & Macy, 1999] Boone, R.T. & Macy, M.W., 1999, "Unlocking the Door's of the Prisoner's Dilemma: Dependence, Selectivity, and Cooperation," *Social Psychology Quarterly*, 62:32-52.

[Frank, 1988] Frank, R., 1988, *Passions within Reason: The Strategic Role of the Emotions*, New York: Norton.

[Frank, 1993] Frank, R., 1993, "The Strategic Role of Emotions: Reconciling Over- and Undersocialized Accounts of Behavior," *Rationality and Society*, 5:160-84.

[Grim, 1995] Grim, P., 1995, "The Greater Generosity of the Spatialized Prisoner's Dilemma," *Journal of Theoretical Biology*, 173:353-359.

[Hamilton, 1964a] Hamilton, W.D., 1964, "The Genetical Evolution of Social Behaviour I," *Journal of Theoretical Biology*, 7:1-16.

[Hamilton, 1964b] Hamilton, W.D., 1964, "The Genetical Evolution of Social Behaviour II," *Journal of Theoretical Biology*, 7:17-52.

[Hechter, 1987] Hechter, M., 1987, *Principles of Group Solidarity*, Berkeley: University of California Press.

[Heckathorn, 1993] Heckathorn, D., 1993, "Collective Action and Group Heterogeneity: Voluntary Provision Versus Selective Incentives," *American Sociological Review*, 58:329-50.

[Hirschman, 1970] Hirschman, A., 1970, *Exit, Voice, and Loyalty*, Cambridge: Harvard University Press.

[Joyce et al., 2006] Joyce, D., et al., 2006, “My Way or the Highway: A More Naturalistic Model of Altruism Tested in an Iterative Prisoner’s Dilemma,” *Journal of Artificial Societies and Social Simulation*, 9(2).

[Kollock, 1998] Kollock, P., 1998, “Social Dilemmas: The Anatomy of Cooperation,” *Annual Review of Sociology*, 24:183-214.

[Macy & Skvoretz, 1998] Macy, M.W. & Skvoretz, J., 1998, “The Evolution of Trust and Cooperation between Strangers: A Computational Model,” *American Sociological Review*, 63:638-660.

[Orbell & Dawes, 1991] Orbell, J.M. & Dawes, R.M., 1991, “A ‘Cognitive Miser’ Theory of Cooperators’ Advantage,” *American Political Science Review*, 85:515-28.

[Orbell & Dawes, 1993] Orbell, J. M. & Dawes, R. M., 1993, “Social Welfare, Cooperator’s Advantage, and the Option of Not Playing the Game,” *American Sociological Review*, 58:787-800.

[Sheratt & Roberts, 1998] Sherratt, T.N. & Roberts, G., 1998, “The Evolution of Generosity and Choosiness in Cooperative Exchanges,” *Journal of Theoretical Biology*, 193.

[Skinner, 1938] Skinner, B.F., 1938, *The Behavior of Organisms: An Experimental Analysis*, Acton, MA: Copley.

[Skyrms, 1996] Skyrms, B., 1996, *Evolution of the Social Contract*, New York: Cambridge University Press.

[Sober & Wilson, 1999] Sober, E. & Wilson, D.S., 1999, *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge: Harvard University Press.

[Thorndike, 1911] Thorndike, E.L., 1911, *Animal Intelligence, Experimental Studies*, New York: Macmillian.

[Thorngate, 2000] Thorngate, W., 2000, “Teaching Social Simulation with Matlab,” *Journal of Artificial Societies and Social Simulation*, 3(1).

[Trivers, 1971] Trivers, R.L., 1971, “The Evolution of Reciprocal Altruism,” *Quarterly Review of Biology*, 100(C): 1073-81.

[Yamagishi & Hayashi, 1996] Yamagishi, T. & Hayashi, N., 1996, “Selective Play: Social Embeddedness of Social Dilemmas,” in Liebrand, W. & Messic, D. (Eds.), *Frontiers in Social Dilemma Research*, Berlin: Springer.