Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Theoretical Biology 223 (2003) 523–531

Journal of Theoretical Biology

www.elsevier.com/locate/jtbi

Evolution of indirect reciprocity by social information: the role of trust and reputation in evolution of altruism

Mojdeh Mohtashemi*, Lik Mui

Laboratory for Computer Science, MIT, 200 Technology Square, Room 419, Cambridge, MA 02139, USA

Received 26 August 2002; received in revised form 27 February 2003; accepted 25 March 2003

Abstract

The complexity of human's cooperative behavior cannot be fully explained by theories of kin selection and group selection. If reciprocal altruism is to provide an explanation for altruistic behavior, it would have to depart from direct reciprocity, which requires dyads of individuals to interact repeatedly. For indirect reciprocity to rationalize cooperation among genetically unrelated or even culturally dissimilar individuals, information about the reputation of individuals must be assessed and propagated in a population. Here, we propose a new framework for the evolution of indirect reciprocity by social information: information selectively retrieved from and propagated through dynamically evolving networks of friends and acquaintances. We show that for indirect reciprocity to be evolutionarily stable, the differential probability of trusting and helping a reputable individual over a disreputable individual, at a point in time, must exceed the cost-to-benefit ratio of the altruistic act. In other words, the benefit received by the trustworthy must out-weigh the cost of helping the untrustworthy.

Keywords: Reciprocity; Social information; Trust; Reputation; Collective memory; Growth; Order

1. Introduction

Throughout social history, humans have exhibited large-scale cooperative behavior toward individuals or groups, with whom they are neither acquainted nor share genetic, phenotypic, or cultural traits. Such behavior is difficult to account for by theories of kin selection (Hamilton, 1963, 1964), group selection (Williams, 1971; Wilson and Sober, 1994), or direct reciprocation (Axelrod and Hamilton, 1981; Axelrod, 1984; Axelrod and Dion, 1988).

Alexander (1987) coined the term *indirect reciprocity* to refer to the commonly practiced act of cooperation in human societies, where the donor of a good deed does not necessarily expect to be rewarded by the recipient but perhaps by another individual who may be the recipient of other good deeds by other donors[1]. Several authors have since attempted to formalize the mechanisms by which indirect reciprocity can evolve. Boyd and Richerson (1989) developed a mathematical model of 'circular reciprocity' where the donor of a good deed is to be rewarded by the last individual in a ring of n reciprocating individuals. Their results suggested that indirect reciprocity is unlikely to be important unless interacting groups are relatively small. Alexander, however, has further hypothesized that "indirect reciprocity is a consequence of direct reciprocity occurring in the presence of others" (Alexander, 1987). Those who observe direct reciprocation between individuals will then be in the position of assessing the reputation of members of a population. Hence, reputation is a key concept in Alexander's premise and he conjectures that indirect reciprocity requires that the reputation and status of members of a group be continually assessed and reassessed. Pollock and Dugatkin (1992) investigated the significance of observation in guiding behavioral choice by studying a variant of tit-for-tat (TFT), where players behave like TFT in the absence of information about a new co-player but defect if the co-player has been observed defecting in his last

*Corresponding author. Tel.: +1-617-253-3512; fax: +1-617-258-8682.

E-mail address: mojdeh@lcs.mit.edu (M. Mohtashemi).

[1] Trivers (1971), who introduced the notion of reciprocal altruism in his seminal paper, referred to this type of reciprocation as *generalized altruism* in order to differentiate it from direct reciprocity.

0022-5193/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0022-5193(03)00143-7

interaction. They showed that when TFT fails to be evolutionarily stable, its variant is indeed evolutionarily stable. To our knowledge, this is the first work to have introduced and utilized the notion of reputation, although binary and minimal, in the context of a theoretical framework for reciprocity and evolution of cooperation.

Following Alexander's conjecture and studies by Pollock and Dugatkin, Nowak and Sigmund (1998a, b) developed a model of indirect reciprocity by image scoring to study the role of observers in assessing the reputation of donating players. Under this model, every player has an image score. When a player is selected as donor, his[2] image score is modified locally by the recipient of the action and a few randomly selected observers. Thus, different individuals may have different perceptions about the same player. Cooperators are rewarded for their altruistic acts through increases of their image scores for their recipients and observers. Every player also has a numeric strategy. A potential donor cooperates only if the recipient's image is at least as high as his own strategy. The underlying premise in this framework is that if image-based information about members of a population can be obtained, then an informed donor only helps those who are likely to help others. This in turn will improve the reputation of the donor thereby introducing feedback into the system. Although an altruistic act entails a cost, if a potential donor helps today, his reputation will improve, which then increases the likelihood that others will help him in the future (Ostrom, 1998).[3] Indeed as baseline experimentation, Nowak and Sigmund (1998a) simulated a population wherein everyone's image is broadcasted globally to all members of the population. Under this model, cooperation evolves irrespective of the population size. This result well demonstrates the role of reputation-based information in the evolution of indirect reciprocity. When information is locally available to observers, however, in order for this information to be of utility, potential donors would have to interact with observed agents since there is no other mechanism for the transfer of information in the population. Nowak and Sigmund's results suggest that when information is localized, cooperation can be established. However, a greater level of interactions per generation, or a larger number of observers, is needed for cooperation to be sustainable in a larger population.

2. Social information: trust, reputation, and collective memory

Although making observations is one of the mechanisms for acquiring information, it is not the principal manner by which humans process information about other's actions. Furthermore, information is seldom randomly propagated in populations.

Embedded in every social network is a web of trust with nodes representing members of the web and edges representing the amount of trust between pairs of friends or acquaintances. When faced with social dilemmas, such as to cooperate or not, individuals make use of social information embedded in their social networks to reduce uncertainty. A key mechanism by which individuals acquire information about others is by seeking the opinions of trusted friends and reputable acquaintances. Parents are the first members of such trust-based networks who provide their children with instruction and advice.

Because it is not possible for one to observe or remember all possible events even in small populations, the information content of such a web of trust serves as one's *collective memory*. Such a trust-based body of information seems to be a fundamental element of social information and an inherent aspect of the processes by which humans make decisions. Aside from this trust-based component of social information, which assumes direct links between members of a social network, there is another vital, while not as apparent, element of social information consequent of indirect links between members of different social networks, where one attempts to seek the opinion of a reputable k-degree acquaintance who is not a direct acquaintance but is connected through a chain of k other individuals. In other words, trust can be inferred in a transitive manner using the notion of reputation: if A trusts B, who trusts C, even if A has not met C before, it is quite likely that A will cooperate with C if C's good reputation is communicated to A by B. Such a process of decision making not only requires the ability to induce friendship and make acquaintances, it also entails the cognitive ability to learn and communicate.

What seems common in most models of evolution of altruism in human populations is the player's passive treatment of information and their limited ability to remember only the near past. Even if we assume that human memory is not functional on a longer timeline, it is important to note the existence of a collective memory upon which humans routinely rely to gain more information and make decisions more effectively. Therefore, as long as we can communicate with our acquaintances, even if everyone in our social web can only remember one past action, at any point in time we have access to a body of information much richer than the information content of our own personal experiences and memory.

[2] Throughout this article, we use the male pronoun for donor and female pronoun for recipient.

[3] Elinor Ostrom in her 1998 paper demonstrates the reinforcing relationship between trust, reputation, and reciprocity, and how reciprocity affects the level of cooperation which affects the overall net benefit in a society.

The resulting network then consists of many small but densely connected clusters of close ties consisting of friends and family, connected to each other by occasional links between few members of different clusters (Granovetter, 1973, 1985).[4] Therefore, the underlying network connectivity, which in turn defines the routes of information transfer, contains *order*. This results in selective retrieval and dissemination of information in populations, a property far from random acquisition of information routinely adopted in the literature.

Another property of social networks which is often ignored in models of evolution of cooperation is dynamics and growth. Even if we assume closed populations of constant size, due to interaction between members of the population, new links are created and new acquaintances are made at all times.[5] This means that: (1) the topology of a social network changes over time making the network a dynamic entity; (2) as new links are added to the network the average path between two individuals becomes shorter over time, increasing the probability that any two randomly selected individuals know each other, thereby causing growth in the network.

In the next section we develop a model of indirect reciprocity by social information in which we make the following assumptions: (1) players can communicate and inquire information about the reputation of their co-players. (2) Information about the reputation of co-players is not obtained randomly, but rather players selectively acquire information from their acquaintance networks by taking advantage of the collective memory of the social networks to which they belong. (3) Information is not propagated randomly in a population; new information resulting from new interactions modifies the content of the collective memory of a recipient and is therefore selectively propagated through the recipient acquaintance network. (4) A social network is an evolving dynamic entity; with new interactions, new links are created, which in turn increases the likelihood of any two randomly selected players to know each other.

3. Model: simulation environment

To make clear the concepts introduced thus far, in this section we develop a simulation environment for the evolution of indirect reciprocity by social information. Consider a population of n individuals divided into non-overlapping groups of acquaintances of the same size. Here, we model the underlying graph structure of a group as a clique. Every generation consists of a fixed number of rounds, m. In every round two players are selected at random, one as donor and the other as recipient. The donor has the option of helping or defecting upon the recipient. If the donor cooperates it will cost him a value of c and the recipient receives a benefit value of b ($b > c$). At the beginning of each generation every player is born into a unique clique of acquaintances. At the end of each generation every one dies and produces offspring in proportion to the total payoff they receive throughout the generation. We assume that every agent j possesses a numeric strategy, k_j, and has an image score, s_{ij}, for agent i (see Nowak and Sigmund, 1998a). At the start of a generation all players have image score of zero. A donor performs one of the two actions, *cooperate* or *defect*. If a donor cooperates, his image score is increased by one unit; otherwise it is decreased by one unit. A potential donor cooperates if the image score of the recipient is at least as high as his own strategy.

Acquaintanceship is defined as a one-way (unidirectional) or a two-way (bidirectional) relation. An acquaintance set is then the union of one- and two-way acquaintances. An initial clique is a bidirectional network. Once the donor performs an action, he becomes a one-way acquaintance of the recipient and her acquaintances if he is not already. By definition of collective memory all the donor's acquaintances will also become one-way acquaintances of the recipient and her acquaintances. However, this is a one-way relationship because the donor gains no new information about the recipient by donating to her. But the recipient who observes the donor in action can use this information to her advantage in future encounters. This means that in future rounds if the recipient is chosen as donor, in addition to her current group of acquaintances she will also inquire the opinion of her newly acquired acquaintance, the donor. Note that the newly acquired acquaintance may be untrustworthy if he defected upon the recipient, in which case the recipient may consider performing the opposite of what the donor recommends. Here, however, we do not impose a distribution upon trust and assume that one equally trusts all one's acquaintances (see Section 4 for further elaboration). Upon performing his action, the donor's image score is updated for his acquaintances and the recipient. However, once the recipient has a new image score for the donor, this knowledge will become part of the collective memory embedded in the recipient's acquaintance set.

Therefore, a new link is created between two players in every round of the game if the donor (or any of his acquaintances) is not in the recipient's acquaintance

[4] Granovetter, in his influential paper *The strength of weak ties*, for the first time noted the existence of order in social networks, in which small clusters of close ties, organized into cliques, are connected by links of acquaintanceship.

[5] In game theoretic models with non-overlapping generations no one dies before the end of the generation. Hence, we assume that links are only added and not removed.

set already. However, this is a one-way relationship since the donor is being added to the recipient's acquaintance set but not vice versa. Such a one-way link may become bidirectional in a future round under three conditions: (1) if the same two players are selected again in opposite roles; (2) if the donor is selected to play as recipient against a donating member of the recipient's acquaintance set; and (3) if a member of the donor's acquaintance set is selected to play as recipient against a donating member of the recipient's acquaintance set (including the recipient herself). To avoid the effects of direct reciprocity on the evolution of cooperation here, we assume that the probability of the same pair being selected again, i.e. condition (1), is negligible. Conditions (2) and (3) are simply implications of the notion of the collective memory. Therefore, when a one-way link is created between a recipient and a donor, by definition all members of the recipient's acquaintance set are also linked to the donor and his acquaintance set in a one-way manner.

If a potential donor does not know the image score of the recipient, he will make use of the social information available to him by asking all his acquaintances if they have ever played in the recipient role against the current recipient, i.e. if the current recipient is a one-way acquaintance of the donor's acquaintance set. If no information is learned, the donor will assume an image score of zero; otherwise, he assesses the reputation of the recipient by adding up her image scores, provided by members of his acquaintance set, and dividing by the total number of encounters. He then compares the final score to his own strategy. He cooperates if the final score is at least as large as his own strategy and defects otherwise. Therefore, the outcome will depend on the probability of knowing the recipient's image, which is derived from the collective memory embedded in the donor's acquaintance set. As the game continues and players meet over time, the underlying topology of the social network also evolves, which causes the probability of knowing a randomly selected recipient to increase over the lifetime of a generation. Fig. 1 illustrates the steady rise of this probability in a simulation environment of a hypothetical population of 100 individuals with an initial clique size of four, i.e. three friends per agent, over a generation consisting of 10 000 encounters. In Section 5 we derive this dynamic quantity analytically (see Eq. (1)–(3) in Methods).

4. Simulation results

Fig. 2 shows the results of computer simulations for varying population sizes with the initial acquaintance clique size of four (three initial friends per agent). Every generation consists of a fixed number of rounds, m. Children inherit neither the image score of their parents nor their parental acquaintance structure. They only inherit the strategy of their parents unless they are subject to mutation. At the beginning of each generation all players are randomly assigned to unique cliques of the same size. The game is played for many generations to subject the population to selective pressure. We say that cooperation is established if the average winning strategy (k) for all individuals at the end of the game is less than or equal to zero. We find that under our framework even with as few as three initial acquaintances[6] cooperation is evolved and sustained after the game is played for many generations (see Fig. 2).

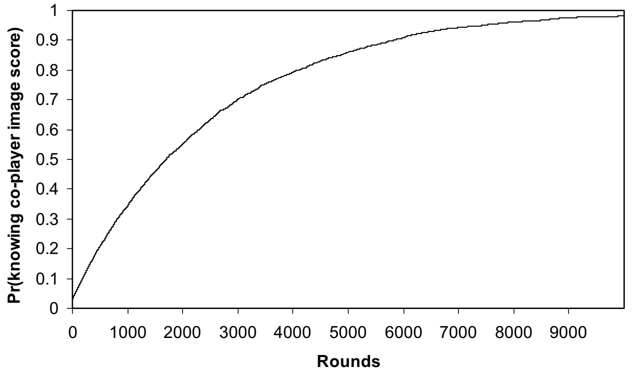

Fig. 1. Dynamics of acquaintanceship. The probability of knowing a randomly selected co-player's image score increases with number of rounds. Initially, a population of $n = 100$ individuals is divided into non-overlapping cliques of size 4, i.e. three friends per agent. As the game is continued, agents meet and make new connections. The average number of acquaintances per player increases over time thereby increasing the probability of knowing the opponent's image score. Here a generation consists of $m = 10\,000$ number of rounds. Towards the end of the generation life the probability of knowing a randomly selected member of the population approaches 1, an indication that if a generation lives long enough one will eventually have met everyone in the population. The rise in the probability of familiarity is consistent with the analytic result in Eq. (1) in the Methods section.

It is interesting to note the effect of the size of the initial social clique on the evolution of indirect reciprocity. Fig. 3 shows the results of computer simulations for a population of 100 individuals under varying initial clique sizes. With more initial acquaintances cooperation is clearly evolved and sustained in a more secure manner. Under similar conditions, in a population of 'all loners', i.e. when the initial clique size is one and no one is initially born into a social group, cooperation fails to evolve. Due to inherent growth factor in the network connectivity, however, even a slight increase in the number of rounds per generation can help a mixture of cooperating strategies to be established in a population of all loners (simulation result not shown). Therefore, either a generation must live long enough so that in the absence of initial social

[6] Sociologists estimate that everyone knows about 150 people on a regular basis and 15 people on an intimate basis.

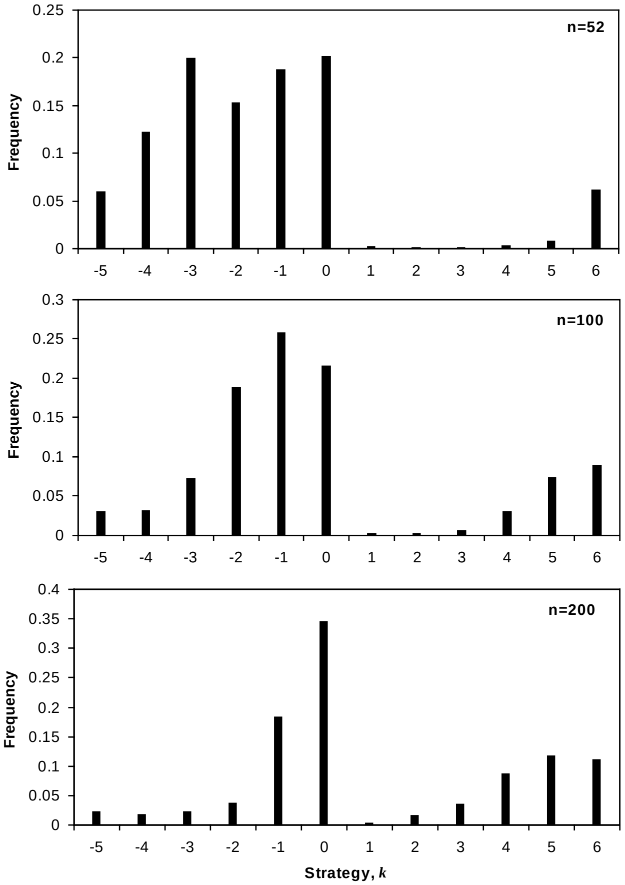

Fig. 2. Evolution of indirect reciprocity by trust and reputation under varying population sizes. A population of n individuals is organized into non-overlapping cliques of size 4, representing three initial acquaintances per player at the beginning of each generation. Modeling after Nowak and Sigmund's simulation framework (Nowak and Sigmund, 1998a), the strategy k ranges from -5 to 6 where $k = -5$ represents unconditional cooperators, $k = 6$ represents defectors, and $k = 0$ represents the most discriminating. The image scores range from -5 to 5. A potential donor cooperates only if the image score of the recipient is at least as large as his own strategy. The children inherit the strategies of their parents unless they are subject to mutation at a rate of 0.001. We sampled the frequency distribution of strategies over 10^6 generations for population sizes $n = 52$, 100, and 200. Each generation consists of $m = 10n$ number of rounds. Cooperation is evolved and sustained in populations after the game is played for many generations.

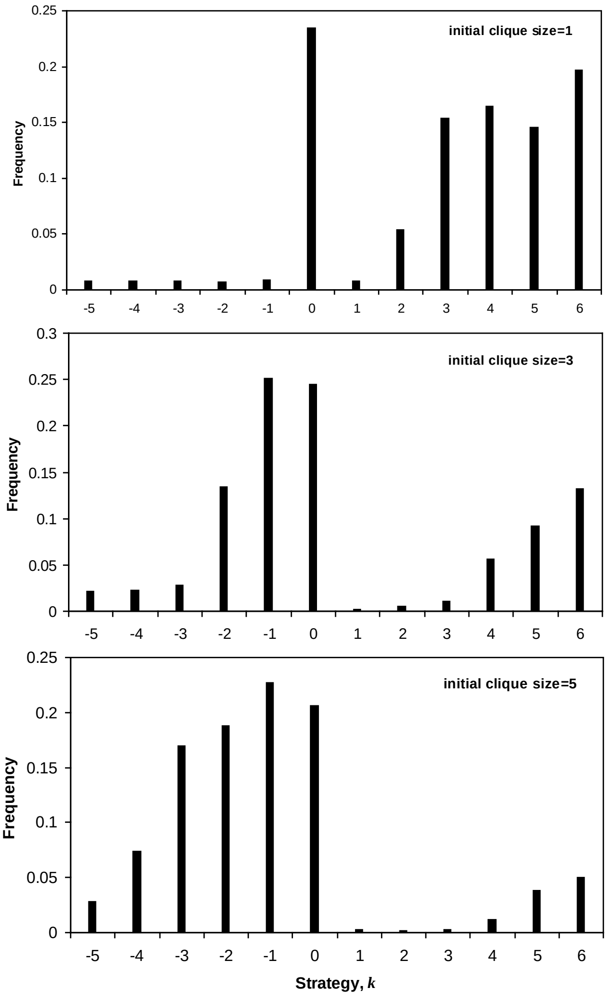

Fig. 3. Evolution of indirect reciprocity by trust and reputation under varying initial clique sizes. We conducted the same simulation experiment as in Fig. 2 except we experimented with different initial clique sizes of 1, 3, and 5, for a population of $n = 100$ individuals. As the initial clique size increases, cooperation is evolved and sustained more securely. In a population of 'all loners', i.e. when the initial clique size is one and no one is born into social groups, defectors win.

structures, over time, interactions between individuals would create interconnected social networks for information to transfer, or if the generation life is short then the existence of initial trust relations gives the information transfer a head start.

When information is scarce, in the absence of a dynamically evolving network of acquaintances, cooperators (players with $k \leqslant 0$) are less fit and can be easily defeated by defectors. As soon as we allow for channels of information to evolve by letting individuals make new connections as they interact, the likelihood of dissemination of information about player's reputation is increased, thereby increasing the likelihood of discriminators to rightfully discriminate against exploiters. Cooperators benefit from this situation: discriminators donate to exploiters less and less, causing exploiter's average fitness and progeny number to decrease; at the same time, discriminators donate to cooperators more and more, causing cooperator's average fitness and progeny number to increase. This is why cooperation can evolve under this scenario despite the fact that discriminators are the only ones who can make use of information. Evolution of cooperation is then a consequence of informed discrimination. Phrasing after Alexander: "indirect reciprocity is a consequence of direct reciprocity occurring in the presence of others", we add that indirect reciprocity is also a consequence of selective inquiry about direct reciprocity.

4.1. Variation on the theme

In the framework presented here we assume that a potential donor takes the advice of all his acquaintances at their face values and simply adds up the image scores reported by them. In particular, even when a donor defects upon his recipient, he still becomes a one-way acquaintance of the recipient. In future rounds, the recipient may inquire the opinion of such a mistrusted acquaintance. Although in social networks clusters of acquaintanceship appear to capture the underlying graph structure of a clique (Granovetter, 1973, 1985), i.e. it may be quite likely that any two acquaintances of an individual are also acquaintances, it is not realistic to assume that one equally trusts all one's friends and acquaintances. In a model of social networks capturing the possibility of misinformation or mistrust, links in the graph must have weights associated with them representing the amount of familiarity or trust between pairs of acquaintances. Furthermore, a pair of acquaintances may have different values of trust for each other. These properties, namely varying degrees of trust for one's acquaintances and asymmetry in pair-wise trust relations should be captured both in the initial cliques and new links resulting from acquiring new acquaintances. But this is a subject for future works.

5. Methods: analytic results

Let A_i be the average number of acquaintances per player at round i, $0 \leqslant i \leqslant m$, where m is the number of rounds per generation. This variable represents the average network connectivity per player at round i (including self). Let $A_0 \geqslant 2$ be the initial clique size at the beginning of each generation, and define q_i as the probability that in round i the donor knows the recipient. Thus in effect, q_i represents the average amount of information available to discriminators at round i since defectors and unconditional cooperators do not make use of information to make decisions.

If we assume that only one donor–recipient pair interacts in each round, as in Sections 3 and 4, then for $1 \leqslant i \leqslant m$ we have

$$q_i = \frac{A_{i-1} - 1}{n - 1}. \tag{1}$$

Eq. (1) asserts that for a discriminator the likelihood of knowing the opponent's reputation at round i is one less than the average number of acquaintances per player from the previous round (assuming that a player cannot play as both recipient and donor) over $n-1$ possible ways a recipient can be selected from a population of size n (excluding the donor). The average network connectivity per player at round i can then be expressed as the following recurrence:

$$A_i = A_{i-1} + (1 - q_i)\frac{A_{i-1}}{n}. \tag{2}$$

This means that a new link is created between two players in every round of the game if the donor or his acquaintances do not know the recipient, i.e. if the donor is not in the recipient's acquaintance set already, in which case the donor is added to the recipient's acquaintance set, resulting in A_{i-1} new links in the recipient's acquaintance set or A_{i-1}/n new links per player. If we rewrite Eq. (2) using Eq. (1) we get

$$A_i = \left(\frac{n}{n-1}\right) A_{i-1} - \frac{A_{i-1}^2}{n(n-1)}. \tag{3}$$

Define $X_i = A_i/n^2$. Then Eq. (3) can be rewritten in canonical form $X_i = (n/(n-1))X_{i-1}(1 - X_{i-1})$, which is the familiar logistic equation. As $n \to \infty$, the growth rate, $n/(n-1) \to 1$. Therefore, for $n \geqslant 2$ the system is stable with the non-trivial fixed point $A^* = n$. This is the maximum network connectivity per player, i.e. the maximum number of acquaintances an individual can have in a population of size n (including self).

In the remainder of this section, we will show how q_i can be used to provide a qualitatively sensible interpretation of a time-varying upper bound on the cost-to-benefit ratio of the altruistic act in order for cooperation to evolve.

5.1. A simple three-strategy game

Here, we modify the basic model of indirect reciprocity with incomplete information, proposed by Nowak and Sigmund (1998b), according to the assumptions underlying the notion of collective memory to derive analytic conditions under which indirect reciprocity can evolve. Unlike their framework, wherein all players are paired up to play as either recipients or donors in each round, here we assume that only one donor–recipient pair will be selected to play in a round. Later in this section we make explicit comparison between our results and those of Nowak and Sigmund's (1998a, b) for a two-strategy game.

Consider a population of size n consisting of three types of players: defectors who never help, cooperators who always help, and discriminators who only help players with good image. Let x denote the frequency of discriminators, y be the frequency of unconditional defectors, and z be that of unconditional cooperators. As in Nowak and Sigmund's (1998a, b) for a discriminating donor, a recipient has a good image if she is known to have cooperated last time; otherwise she has a bad image. We modify the model of indirect reciprocity by Nowak and Sigmund by imposing a social structure on acquaintance relations so that players can acquire information through their acquaintance network. If discriminating donors do not learn new information about their recipients by asking their acquaintances, they always cooperate. This means that in the absence of information defectors can be mistaken for good scorers by discriminators. Therefore at the beginning of each generation, when no information about players' behaviors is available, all players are assumed to be good scorers. Let g_i be the proportion of good scorers at round i. Then $g_i = x_g(i) + y_g(i) + z_g(i)$, where $x_g(i), y_g(i), z_g(i)$ represent the respective frequency of good scoring discriminators, defectors, and cooperators at round i. For cooperators $z_g(i) = z$ for $1 \leqslant i \leqslant m$ since they always help. For discriminators and defectors $x_g(1) = x$ and $y_g(1) = y$, since at the beginning of the game everyone has a good image.

For defectors, $y_g(i) = (1 - (1/n))y_g(i-1)$, $2 \leqslant i \leqslant m$. This is because the proportion of good scoring defectors in round i changes only if a good scoring defector is selected to play as donor with probability $y_g(i-1)$, in which case the number of good scoring defectors is reduced by one (or by $1/n$ in proportion). Hence, for $1 \leqslant i \leqslant m$ we have

$$y_g(i) = \left(1 - \frac{1}{n}\right)^{i-1} y. \tag{4}$$

The proportion of good scoring discriminators that changes in every round due to selection of a discriminating donor is x/n. There are two conditions under which the proportion of good scoring discriminators in round i can change: (1) if a bad scoring discriminator is selected to play as donor and he cooperates, which happens with probability $(x - x_g(i-1))(1 - q_{i-1} + q_{i-1}g_{i-1})$, in which case the number of good scoring discriminators will increase by one (or by $1/n$ in proportion); (2) if a good scoring discriminator is selected to play as donor and defects, which happens with probability $x_g(i-1)q_{i-1}(1-g_{i-1})$, in which case the number of good scoring discriminators will decrease by one (or by $1/n$ in proportion). Therefore, we have

$$x_g(i) = \left(1 - \frac{x}{n}\right)x_g(i-1) + \frac{x^2}{n}(1 - q_{i-1} + q_{i-1}g_{i-1}). \tag{5}$$

Let $F_x(i), F_y(i), F_z(i)$ denote the respective expected payoff for discriminators, unconditional defectors and unconditional cooperators at round i. Then we have

$$F_x(i) = \frac{1}{n}\left\{-c(q_i g_i + 1 - q_i) + b(x+z) - bq_i x\left(1 - \frac{x_g(i)}{x}\right)\right\}, \tag{6}$$

$$F_y(i) = \frac{b}{n}\left\{z + x(1-q_i) + xq_i\frac{y_g(i)}{y}\right\}, \tag{7}$$

$$F_z(i) = \frac{1}{n}\{-c + b(x+z)\}. \tag{8}$$

Because only one donor–recipient pair interacts in each round and therefore the payoff must be averaged over the population of interacting strategies, the cost c and benefit b are divided by the respective number of discriminators, defectors and cooperators, xn, yn, zn. At the same time, each equation is multiplied by the respective probability of selecting a discriminator, defector and cooperator, x, y, z. Hence, the ratio $1/n$ appears in all three equations. Next, we normalize the payoffs by subtracting Eq. (7) from all three equations to get

$$\hat{F}_x(i) = \frac{1}{n}\left\{-c(q_i g_i + 1 - q_i) + bq_i x \times \left(\frac{x_g(i)}{x} - \frac{y_g(i)}{y}\right)\right\}, \tag{9}$$

$$\hat{F}_z(i) = \frac{1}{n}\left\{-c + bq_i x\left(1 - \frac{y_g(i)}{y}\right)\right\}. \tag{10}$$

Therefore, the total expected payoffs for discriminators and cooperators over one generation consisting of m rounds are

$$P_x = \frac{1}{n}\left\{-mc + c\sum_{i=1}^{m} q_i(1-g_i) + bx\sum_{i=1}^{m} q_i\left(\frac{x_g(i)}{x} - \frac{y_g(i)}{y}\right)\right\}, \tag{11}$$

$$P_z = \frac{1}{n}\left\{-mc + bx\sum_{i=1}^{m} q_i\left(1 - \frac{y_g(i)}{y}\right)\right\}. \tag{12}$$

Note that $P_z > 0$ iff

$$\frac{c}{b} < \frac{x\sum_{i=1}^{m} q_i\left(1 - y_g(i)/y\right)}{m} \tag{13}$$

and $P_x > 0$ iff

$$\frac{c}{b} < \frac{x\sum_{i=1}^{m} q_i\left(x_g(i)/x - y_g(i)/y\right)}{m - \sum_{i=1}^{m} q_i(1-g_i)}. \tag{14}$$

Inequality (13) asserts that for cooperators to attain a positive fitness over a generation, the average per-round probability of a discriminator defecting on a defector must exceed the cost-to-benefit ratio of the altruistic act. This is a sensible condition since in the absence of discriminators, unconditional cooperators will be less fit and easily defeated by defectors. In other words, for cooperators to prosper, they need discriminators who can use information to rightfully deny defectors help. However, a reasonable precondition to discriminators defecting on defectors is for them to survive, i.e. to attain positive fitness. Therefore, if inequality (14) is satisfied, so is inequality (13) in which case a mixture of cooperating strategies will be established (unconditional cooperators and discriminators). To make intuitive sense of the upper bound cross multiply inequality (14) and divide both sides by m to get

$$c\left(1-\frac{1}{m}\sum_{i=1}^{m} q_i(1-g_i)\right) < \frac{bx}{m}\sum_{i=1}^{m} q_i\left(\frac{x_g(i)}{x}-\frac{y_g(i)}{y}\right). \tag{15}$$

The ratio $(1/m)\sum_{i=1}^{m} q_i(1-g_i)$ is the average per round probability of a discriminator not cooperating. This is because $1-g_i$ is the proportion of bad scorers at round i, which is the probability that a randomly selected recipient is a bad scorer, and q_i is the probability that a discriminator knows the recipient. The term $1-(1/m)\sum_{i=1}^{m} q_i(1-g_i)$ is then the average per round probability of a discriminator cooperating. The ratio

$$\frac{x}{m}\sum_{i=1}^{m} q_i\left(\frac{x_g(i)}{x}-\frac{y_g(i)}{y}\right)$$

is the average per round differential probability of a discriminator being helped over a defector. Inequality (15) then asserts that for cooperators to outperform the defectors, the average per round differential benefit received by a discriminator over a defector must exceed the average per round cost to a discriminator for cooperating. In other words, trust pays off only if it is placed upon the trustworthy.

5.2. *Two-strategy game*

To compare the outcome of the approach presented here with that of the two-strategy model of incomplete information proposed by Nowak and Sigmund (1998b), we first derive a condition analogous to inequality (15) for a game with two strategies consisting of discriminators and defectors, where $g_i = x_g(i) + y_g(i)$. For a game with two strategies inequality (15) reduces to

$$c\left(1-\frac{1}{m}\sum_{i=1}^{m} q_i(1-g_i)\right) < \frac{b}{m}\sum_{i=1}^{m} q_i\left(g_i-\frac{y_g(i)}{y}\right). \tag{16}$$

Inequality (16) has a similar interpretation to inequality (15): for discriminators to outperform the defectors, the average per round differential benefit received by a discriminator over a defector must exceed the average per round cost to a discriminator for cooperating. If we rewrite inequality (16) in terms of the cost-to-benefit ratio we derive that for a game with two strategies we must have

$$\frac{c}{b} < \frac{1/m\sum_{i=1}^{m} q_i\left(g_i - y_g(i)/y\right)}{1-1/m\sum_{i=1}^{m} q_i(1-g_i)}. \tag{17}$$

For a two-strategy game, under similar assumptions about the population composition, Nowak and Sigmund derived the condition that for discriminators to be evolutionarily stable we must have: $q > c/b$, where q is a constant representing the probability of knowing the co-player's image score. Nowak and Sigmund aptly pointed out the similarity between their result and Hamilton's rule for altruism through kin selection, where the parameter for genetic relatedness is replaced by q, i.e. the likelihood of knowing the opponent's reputation. However, when we impose a constantly evolving social structure under selectivity and order in information transfer, the parameter of *familiarity*, whether genetic, cultural, or simply due to observation or interaction, is replaced by a time-varying ratio that amounts to the average differential likelihood of knowing and helping a trustworthy individual (discriminator) over an untrustworthy individual (defector) out of the proportion of those that a discriminator helps at a point in time. As we have shown, this dynamic quantity can be derived from the collective knowledge embedded in the social network of acquaintances once we remove randomness in observation and introduce a trust-based bias in the selectivity and propagation of information.

6. Discussion

We proposed a new framework for the evolution of indirect reciprocity by social information and its vital constituents, trust and reputation, under two organizing principles of social networks: order and growth (Barabasi, 2002). We defined 'collective memory' as the information content of networks of friends and acquaintances, and argued that at any time the amount of information available to an individual is much larger than that of one's own memory. Such body of information can be retrieved by communicating to friends and acquaintances. This would then imply that in social networks: (1) information is selectively retrieved from and propagated through networks of acquaintances due to the existing clustering effect of social ties; (2) Dynamics and growth constantly modify the topology of the underlying network thus modifying the content of the collective memory.

These observations are a significant departure from the prevailing view of social networks as random and static entities, which can be utilized to identify new sources of information thereby increasing the average amount of information available at any point in time and the likelihood for informed cooperation. Under these assumptions, we analytically derived the condition under which indirect reciprocity can evolve. Our results suggest that for cooperators to be evolutionarily stable, on average the benefit received by a trusting and reputably cooperative individual must out-weigh the cost of trusting and helping a disreputable individual. In other words, trust pays off only if it is placed upon the trustworthy. But to make the right decision as to whom to trust, information on trustworthiness must be obtained by communicating to others. Hence, implicit in our framework is the premise that the evolution of human language has played a critical role in the evolution and sustenance of indirect reciprocity in human populations.

Are some populations more cooperative than others due to different processes of information transfer? Different political and economical structures, among many other factors, appear to result in different underlying networks of information transfer in different societies. Under more rigid economic and political conditions, where the processes of information transfer are rather stagnant and the cost of trust and cooperation is too high (sometimes one's life), people tend to distrust strangers. Lack of trust and cooperation then results in disreputability, which introduces negative feedback into the network driving the system into all-defection. By analogy, less rigidity in the political and economic infrastructure should result in people exhibiting more cooperative behavior due to the relative openness of channels of information and lower costs of trusting and cooperating with strangers. However, this question should be put to test more rigorously through further experimental and theoretical studies (see Henrich et al., 2001, 2002).[7]

Have humans become increasingly more cooperative throughout their social history? If the availability of information about others was the single driving force behind evolution of cooperation, and if, on average, information has become more available due to higher degree of mobility and technological advancement throughout human history then one may also expect to observe higher levels of cooperation over time. Our goal here, however, was to study the evolution of cooperation through indirect reciprocity by social information, a task much more modest than providing a sensible answer to the posed question. Nevertheless we find the question worthy of note and leave it open for future studies.

Acknowledgements

The authors would like to thank Dr. Ali Saidi, Professor Peter Szolovits and two anonymous reviewers for their insightful comments.

References

Alexander, R.D., 1987. The Biology of Moral Systems. Aldine de Gruyter, New York.

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

Axelrod, R., Dion, D., 1988. The further evolution of cooperation. Science 242, 4884.

Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. Science 211, 1390.

Barabasi, A.-L., 2002. Linked: the New Science of Networks. Perseus Publishing, Cambridge, MA.

Boyd, R., Richerson, P.J., 1989. Evolution of indirect reciprocity. Soc. Networks 11, 213–236.

Granovetter, M.S., 1973. The strength of weak ties. Am. J. Soc. 78, 1360–1380.

Granovetter, M.S., 1985. Economic action and social structure: the problem of embeddedness. Am. J. Soc. 91, 481–510.

Hamilton, W.D., 1963. The evolution of altruistic behavior. Am. Nat. 97, 354–356.

Hamilton, W.D., 1964. The genetical evolution of social behavior. J. Theor. Biol. 7, 1–16.

Henrich, J., Boyd, R., Bowles, S., Gintis, H., Fehr, E., 2001. In search of homo economicus: experiments in 15 small-scale societies. Am. Econ. Rev. 91, 73–78.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Hill, K., Gil-White, F., Gurven, M., Marlowe, F., Patton, J.Q., Smith, N., Tracer, D., 2002. 'Economic man' in cross-cultural perspective: behavioral experiments in 15 small-scale societies. Unpublished manuscript. http://webuser.bus.umich.edu/henrich/gameproject.htm.

Nowak, M.A., Sigmund, K., 1998a. Evolution of indirect reciprocity by image scoring. Nature 393, 573–577.

Nowak, M.A., Sigmund, K., 1998b. The dynamics of indirect reciprocity. J. Theor. Biol. 194, 561–574 (doi:10.1006/jtbi.1998.0775).

Ostrom, E., 1998. A behavioral approach to the rational-choice theory of collective action. Am. Pol. Sci. Rev. 92, 1–22.

Pollock, G., Dugatkin, L.A., 1992. Reciprocity and the emergence of reputation. J. theor. Biol. 159, 25–37.

Trivers, R., 1971. The evolution of reciprocal altruism. Q. Rev. Biol. 46, 35–57.

Williams, G.C., 1971. Group Selection. Aldine-Atherton, Chicago.

Wilson, D.S., Sober, E., 1994. Reintroducing group selection to the human behavioral sciences. Behav. Brain Sci. 17, 585–654.

[7] The preliminary results of a series of experimental field studies by the authors suggest that the more market integrated a society the more cooperative its people are. It is important to note, however, that in these studies cooperation is defined in economic terms.