# Coevolution of honest signaling and cooperative norms by cultural group selection

István Scheuring [a,b,*]

[a] Department of Plant Taxonomy and Ecology, Research Group of Theoretical Biology and Ecology, Eötvös Loránd University and the Hungarian Academy of Sciences, Budapest, H-1117, Pázmány Péter sétány, 1/c, Hungary
[b] Konrad Lorenz Institute, Altenberg, A-3422, Adolf Lorenz Gasse 2, Austria

## ARTICLE INFO

## ABSTRACT

Evolution of cooperative norms is studied in a population where individual and group level selection are both in operation. Individuals play indirect reciprocity game within their group and follow second order norms. Individuals are norm-followers, and imitate their successful group mates. Aside from direct observation individuals can be informed about the previous actions and reputations by information transferred by others. A potential donor estimates the reputation of a potential receiver either by her own observation or by the opinion of the majority of others (indirect observation). Following a previous study (Scheuring, 2009) we assume that norms determine only the probabilities of actions, and mutants can differ in these probabilities. Similarly, we assume that individuals follow a stochastic information transfer strategy. The central question is whether cooperative norm and honest social information transfer can emerge in a population where initially only non-cooperative norms were present, and the transferred information was not sufficiently honest.

It is shown that evolution can lead to a cooperative state where information transferred in a reliable manner, where generous cooperative strategies are dominant. This cooperative state emerges along a sharp transition of norms. We studied the characteristics of actions and strategies in this transition by classifying the stochastic norms, and found that a series of more and more judging strategies invade each other before the stabilization of the so-called generous judging strategy. Numerical experiments on the coevolution of social parameters (e.g. probability of direct observation and the number of indirect observers) reveal that it is advantageous to lean on indirect observation even if information transfer is much noisier than for direct observation, which is because to follow the majorities' opinion suppresses information noise meaningfully.

© 2010 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Evolutionary origin and stability of indirect reciprocity where the return of altruistic aid is expected from someone other than the recipient of the aid is one that is very characteristic of human nature. This behavior can only be explained if the actions are observed and classified by the members of society with the help of a social norm (Trivers, 1985; Alexander, 1987). Knowing the score (reputation) of a potential recipient (and the donor) and the norm followed by the potential donor, she can decide whether her recipient is worth donating to or not. If free-riders are excluded effectively from the interaction by this norm then indirect reciprocity can be maintained.

By studying this concept within a mathematical framework Nowak and Sigmund (1998a,b) have shown that cooperation by indirect reciprocity is maintained by a norm called image scoring. In their model individuals score increases by one if she donates to a recipient and decreases by one if she refuses donation. Individuals who follow image scoring help only those individuals whose score is above a threshold, so individuals that were altruistic enough in the past are favored.

These keystone papers catalyzed a series of studies, including experimental works (e.g. Wedekind and Milinski, 2000; Fehr and Fischbacher, 2003), and a range of analytical and numerical investigations (e.g. Leimar and Hammerstein, 2001; Brandt and Sigmund, 2004; Panchanathan and Boyd, 2003, 2004; Ohtsuki and Iwasa, 2004; Chalub et al., 2006; Pacheco et al., 2006). Leimar and Hammerstein (2001) pointed out that image scoring is not an evolutionary stable strategy if group structure of human populations and inherent decision stochasticity is taken into account. They found that a so-called standing strategy which offers help if its score is below a critical level, can overcome image scoring strategy. Despite

* Correspondence address: Department of Plant Taxonomy and Ecology, Research Group of Theoretical Biology and Ecology, Eötvös Loránd University and the Hungarian Academy of Sciences, H-1117, Pázmány Péter sétány, 1/c, Hungary.
E-mail address: shieazsf@ludens.elte.hu.

that image scoring is not an evolutionary stable strategy, subjects follow image scoring rather than standing strategy according to empirical studies (Milinski et al., 2001). Milinski et al. (2001) argued that errors in perception and limited working memory leads to the subjects adopting image scoring strategy.

Comprehensive theoretical investigations on the success of social norms in indirect reciprocity game were addressed later by (Brandt and Sigmund, 2004; Ohtsuki and Iwasa, 2004, 2006; Ohtsuki et al., 2009). Ohtsuki and Iwasa (2004, 2006), similar to most recent studies, assumed that individuals are either in *Good* or in *Bad* reputation, and they are reliably informed about the reputation state of every one. They considered all the possible third order norms, that is when an observer makes decision according to the donor and recipient reputation state and the action of the donor. (For first order norms, only the actions of the donor are taken into account, for the second order norms the reputation of either the recipient or the donor and the action also contribute to determine the new reputation value of the donor.) Thus, there are $2^4$ different action strategies and $2^8$ possible norms. They assumed that except for some small error individuals that are well-informed about the actions, further individuals can make some errors during execution of intended actions. They found eight reputation systems among the possible $2^4 \times 2^8 = 4096$, which are ESS and maintain a high level of cooperation. The common nature of these so-called "leading-eight" reputation systems is that all of them are nice (maintenance of cooperation), retaliatory (detection and punishment of defection, and justification of punishment), apologetic, and forgiving (Ohtsuki and Iwasa, 2006). In a parallel work, Brandt and Sigmund (2004) studied the evolution of only 14 different reputation systems among the possible 4096, but they studied the invasion and coexistence of strategies in a group structured individual-based model. Their main conclusion is that the standing strategy is generally superior to image scoring strategy, but standing, image scoring and a judging norms (see below) are typically in stable coexistence.

Recently Ohtsuki et al. (2009) extended their previous models (Ohtsuki and Iwasa, 2004, 2006) by studying the role of costly punishment in the indirect reciprocity framework. They used second order norms, but beside cooperation and defection, punishment, as a third possible action was available. They were interested again the evolutionary stable norms which maintain high level of cooperation. They assumed that there is an inherent error in assigning reputation, but all individuals have the same opinion to a given person. They found that costly punishment is more efficient than non-punishing defection towards bad individual only at a narrow range of parameters. They studied this situation as well when everyone has a private list of the reputation of the others, and found that even small interpretation error can destroy cooperation. However, if there is a communication phase among the actions, where individuals sample each others opinion about a third party, then cooperation is maintained.

Most studies emphasized that ancient human populations lived in small interacting groups (e.g. Leimar and Hammerstein, 2001; Brandt and Sigmund, 2004; Pacheco et al., 2006; Chalub et al., 2006; Scheuring, 2009), thus cooperative norms and social institutions are evolved through cultural group selection (Bowles et al., 2003; Bowles and Gintis, 2003; Pacheco et al., 2006; Chalub et al., 2006). The group selection mechanism is widely debated by arguing that group (or multilevel) selection can be explained by kin selection mechanism as well (Traulsen and Nowak, 2006; Taylor and Nowak, 2007; Lehmann et al., 2007; Traulsen et al., 2008). However, the group structure of hunter-gatherer societies is obvious (Ember, 1978; Richerson et al., 2001; Soltis et al., 1995), thus the multilevel selection perspective is natural in our case.

By using multilevel selection Pacheco et al. (2006) found that evolution leads to a second (despite that individuals can follow

third order norms), named "stern-judging". Under stern-judging norm giving help to a good individual and refusing help to a bad individual lead to a good reputation, while refusing help to a good and giving help to a bad one lead to bad reputation. Stern-judging is among the leading-eight norms found by Ohtsuki and Iwasa (2004, 2006), although stern-judging is the most successful in a multi-level selection process while leading-eight norms are only stable against the invasion of rare non-cooperative strategies under individual level selection. We note here that Chalub et al. (2006) had the same conclusion by using a similar model framework.

In a recent paper Scheuring (2009) extended the definition of norm by using stochastic norms in his model. Individuals following a stochastic norm consider an action to be good with probability $p$ (and bad with probability $1 − p$). He focused on the question whether cooperative norms can evolve when populations used non-cooperative norms initially. By using a similar modeling framework that was used by Pacheco et al. (2006), and assuming that mutants can differ in the probabilities defining the norm, it is found that the evolution of norms lead to a population which follow a so-called "generous stern-judging strategy" on average. According to this strategy giving help to a good individual and refusing help to a bad individual is considered to be good with a high probability, but to help a good individual is classified to be a better action than refusing a bad. Refusing help to a good individual is a bad action with a high probability under generous judging, but with a small, but definite probability it is valued to be good. This kind of generous norm system is effectively maintained by cooperation in a system where norm polymorphism and social noise is present (Scheuring, 2009).

In previous models individuals are either well informed about the social status of others (Ohtsuki and Iwasa, 2004, 2006) or only some social noise can cause misinterpretations for previous actions (Nowak and Sigmund, 1998b; Pacheco et al., 2006; Scheuring, 2009; Ohtsuki et al., 2009). Since it unlikely that individuals observe every previous (or at least the most) interaction, the information transfer (gossiping) among individuals forms the opinion of the behavior of others in the population (Ohtsuki et al., 2009). However information transfer is not necessarily reliable (Hess and Hagen, 2006), what seems to be even more plausible is that giving false or negative information about others increases the relative fitness of the gossiping people (McAndrew and Milenkovic, 2002), thus being an advantageous strategy. Consequently, to understand the origin of reliable and honest social information transfer in the context of evolution of cooperative norms is a crucial problem.

The central questions of this paper are: how a reliable communication system can emerge in indirect reciprocity game? Can social norm evolve to maintain high level of indirect cooperation if information is based on indirect observation? I study these problems by extending my previous model (Scheuring, 2009), by assuming that the potential donor (partly or totally) is informed by other observers, and that individuals use information transfer as an adaptive (stochastic) strategy. In this model we can study the evolution of information strategies and the social norms, and even the evolution of information network. In the following, I introduce the model and then results of numerical simulations are presented and discussed.

## 2. The Model

### 2.1. Indirect Reciprocity Game, Reputation System

Individuals play the indirect reciprocity game. The actor can give a help to the recipient, which decreases its fitness by $c$, while the fitness of the recipient increases by $b$, where $b > c$ (Nowak and Sigmund, 1998b). (For convenience, we fixed $c = 1$.) If a selfish actor