Evolutionary Ecology Research, 2004, **6**: 261–283

Evolution of rumours that discriminate lying defectors

Mayuko Nakamaru[1]* and Masakado Kawata[2]

[1]*Department of Systems Engineering, Shizuoka University, 3-5-1 Johoku, Hamamatsu 432-8561 and* [2]*Division of Ecology and Evolutionary Biology, Graduate School of Life Sciences, Tohoku University, Aoba, Sendai 980-8578, Japan*

ABSTRACT

Discrimination of defectors is the key to the evolution of cooperation. In this paper, we examine the effect of rumours on the evolution of cooperation. 'Rumour' is defined as spreading or starting a reputation (concerned with cooperativeness of players), which is passed from one player to another. If players receive a rumour that a given player is a defector, they can avoid being defected by that player. However, to establish and maintain their cooperative relationship, players must also be able to distinguish incorrect rumours from correct ones because the rumour might be untrue. The speed of spreading an incorrect rumour is expected to affect its likelihood of detection and the consequences of detection. In computer simulations (individual-based simulation model), a pair (chosen randomly) plays the Prisoner's Dilemma game once. Each strategy consists of three rules: (1) a rule for the Prisoner's Dilemma game; (2) a rule for spreading a rumour; and (3) a rule for starting rumours. We consider 39 strategies in total. Then we classify strategies into several groups. The ADVISOR group of strategies, whose members are cooperative and start rumours about defectors, is not invaded by the LIAR group, whose members are defectors who tell lies, saying 'I'm cooperative'. ADVISOR (as a group) is not invaded by LIAR even though given pairs (one from each group) seldom meet more than once. If a correct rumour is seldom received, however, and if an incorrect rumour spreads quickly, members of ADVISOR (i.e. members that use strategies from the ADVISOR group) are confused by the incorrect rumour. As a result, ADVISOR is invaded by LIAR. As another group of strategies, we consider CONDITIONAL_ADVISOR, of which each representative member only spreads rumours received from other players who have cooperated with him. Unlike ADVISOR, CONDITIONAL_ADVISOR is able to refuse being invaded by LIAR. On the other hand, if the same conditions hold, but the cooperative strategy is instead based on the individual's own past experience, that group is invaded by LIAR. This implies that a rumour evolves if cooperators can, using that rumour, detect and punish cheaters – especially a lying defector.

Keywords: honest signalling, language, lying cheater, Prisoner's Dilemma game, punishment, rumour.

* Author to whom all correspondence should be addressed. e-mail: nakamaru@sys.eng.shizuoka.ac.jp

INTRODUCTION

The detection of cheaters is important for the evolution of cooperative behaviour. Genes might cheat other genes in the cell in the early evolutionary stage of life (Maynard Smith and Szathmary, 1995) by using shared resources to replicate themselves faster than others. Having a mechanism for detecting cheaters would enable the multiple genes to cooperate and succeed in proliferating together.

Punishment or policing is favoured to prevent cooperative individuals from being cheated and to discourage defection in some social animals such as red deer, social insects and primates (e.g. Clutton-Brock and Parker, 1995). In some hymenopteran societies where a queen mates with two or three males, workers eat the eggs produced by others. This is considered to be policing behaviour aimed at preventing other workers from producing eggs (e.g. Monnin and Ratnieks, 2001). In addition to theoretical work (e.g. Sigmund *et al.*, 2001; Boyd *et al.*, 2003), experimental studies (Fehr and Gachter, 2002) have shown that punishment of cheaters is effective in maintaining cooperation among humans. When cheaters are punished, their benefits are reduced, which might promote the evolution of cooperation. However, the magnitude of the cost that the punishers incur might affect the evolution of cooperation and punishment. For example, Boyd *et al.* (2003) showed that group selection promoted the evolution of altruistic punishment only when the cost of being punished was much greater than the cost of punishing.

In human societies, language is available as a cost-free method of transmitting information among individuals. Language can be used as an effective tool for punishment and signalling. Sober and Wilson (1998) and Wilson *et al.* (2000) indicated that gossip and rumours about cheaters might work as punishment, which can prevent individuals from being cheated. In their experiment, Wilson *et al.* (2000) showed that self-serving gossip was not approved of but gossip in response to norm violations was. Their results suggested that gossip might function to detect defectors. Dunbar (1996) discussed the evolution of language in early human populations, and suggested that gossip instead of grooming would have been needed to maintain cooperation in a large group, because each member could not groom all the other members. Thus language might have evolved to inform others of traitors.

The evolution of cooperation by indirect reciprocity promoted by reputation or signalling has been studied by several researchers (e.g. Pollock and Dugatkin, 1992; Wilson and Dugatkin, 1997; Nowak and Sigmund, 1998; Leimar and Hammerstein, 2001; Riolo *et al.*, 2001; Sigmund *et al.*, 2001; Milinski *et al.*, 2002). Nowak and Sigmund (1998), for instance, used image scoring to indicate the reputation of each player for cooperativeness. Leimar and Hammerstein (2001) showed that when they introduced errors of execution and perception, the 'good-standing' strategy beat (with strong gene flow) the image scoring strategies in the island model. Riolo *et al.* (2001) considered tags (representing the degree of similarity between players) as a kind of signalling. These studies assumed that each individual observed others' reputations, which, for each of the others, were the signals indicating its past behaviour or which were used just for self-promotion. They also assumed that these reputations were true and correct, although Leimar and Hammerstain (2001) studied the possibility that recipients of reputation might fail to recognize correct information. However, no studies have considered the possibility that reputation is used to cheat others.

A basic difference between signalling (or reputation) and rumour (or gossip) is that, with signalling, group members receive information about other members by meeting and seeing them directly, whereas with rumour, they receive information about others without directly seeing and meeting them.

We consider situations in which individuals pay nothing for promoting themselves or for spreading (or starting) a rumour or gossip, as, for example, about defectors. Moreover, a rumour produces no direct benefits or pay-offs, although it indirectly affects the future behaviour of the players and their pay-offs in social interaction.

Here, rumour (as activity within a social group) is defined as the spreading of a given player's reputation (about its cooperativeness), or the initiation of such spread, as a consequence of players passing pieces of information from one to another. We then propose a new framework for modelling the effect of rumour on the evolution of social interaction. We use the Prisoner's Dilemma game to examine social interaction through rumour exchange, which allows us to investigate the indirect effects of rumour on the evolution of cooperation. The model allows us to control the speed at which a rumour is spread relative to the number of Prisoner's Dilemma games played.

In the model, a lying defector is a cheater. He always defects in the Prisoner's Dilemma game but tells others that he is cooperative. If lies cannot be detected, lying defectors spread in the population and cooperation cannot be maintained. Hence, we ask which strategies are effective against a liar strategy and thereby promote the evolution of rumour and its role in detecting defectors. We also investigate how the speed at which a rumour is spread affects the defection of a lying defector and how the spreading of a rumour can change the cooperative relationships among individuals. We hypothesize that the activity of starting and spreading of rumours evolves if cooperators can, using rumour, detect and punish cheaters, especially liars.

THE MODEL

Structure of the model

Each player has three rules: (1) a rule for playing the Prisoner's Dilemma game (see Table 1); (2) a rule for spreading a rumour; and (3) a rule for starting a rumour. We consider rules (2) and (3) separately because, for a rumour to spread, it must first be started.

Table 1. The pay-off for a Prisoner's Dilemma game

	Partner	
Focal player	Cooperation	Defection
Cooperation	$B - C$	$-C$
Defection	B	0

Note: By cooperating with a partner, a player incurs a cost $-C$ ($C > 0$), and his partner receives a benefit B ($B > 0$). When both players cooperate, both receive the pay-off $B - C$; when both players defect, neither benefits and neither loses the game. Table 1 shows the pay-off received by the focal player of the game. We assume that B is 5 and C is 2, which satisfies the condition of a Prisoner's Dilemma game (Axelrod and Hamilton, 1981).

The population consists of N individuals (players). To provide sufficient capacity for complex interactions, each generation consists of T units of time. When each generation ends, the strategies (each represented by a group of players) are reproduced, each based on the accumulated scores of its members' outcomes in playing the Prisoner's Dilemma game. With population size N fixed, the number of players in the next generation with a particular strategy i is $(S_i/\sum_j S_j) \times N$, where S_i is the sum of the scores of all players using strategy i.

We assume that rumour exchange is independent of the Prisoner's Dilemma game for two reasons. First, to play the Prisoner's Dilemma game, two players are randomly (in a uniform way) chosen g times from the population within a given unit of time. Then, after all of the g Prisoner's Dilemma games have been played, two players are randomly chosen (in a uniform way) for rumour exchange. Such choices are repeated r times in each unit of time. Thus, there are $T \times g$ Prisoner's Dilemma games per generation and $T \times r$ rumour exchanges per generation (T = duration of one generation). By varying r and g, we can regulate the speed at which rumours spread (i.e. amount of spread per generation) (Fig. 1A). More specifically, $r/g > 1$ indicates that a rumour spreads quickly, and $r/g < 1$ that it spreads slowly.

In one generation (duration T), the expected number of Prisoner's Dilemma games for a particular pair of players is $w_g = T \times g/{}_N C_2$, and the expected number of rumour exchanges for a particular pair of players is $w_r = T \times r/{}_N C_2$. Our parameter set, then, is (N, T, g, r, w_g, w_r). By varying g, r and T independently, we examine how changes in these three parameters influence the evolution of cooperation. In particular, we devote our attention to the effect of r, a measure of the speed at which rumours spread.

An arbitrary focal player i ($i = 1, \ldots, N$), with respect to another player j, can have $4N$ internal state variables: $C_i(j)$, $D_i(j)$, $pc_i(j)$ and $pd_i(j)$ (see Fig. 1B). $C_i(j)$ is the number of times that player i has received a positive rumour about player j, and $D_i(j)$ a negative rumour about player j. The number of past instances of cooperation is $pc_i(j)$, and the number of past instances of defection is $pd_i(j)$.

We first determine whether a rumour-using strategy evolves due to an increase in its frequency (relative number of players using it) within a population. Each player represents a combination of traits, with each trait being a rule for interaction. To simplify the investigation of the effect of the rumour itself, in deciding whether to cooperate or defect with a particular partner in the Prisoner's Dilemma game, some players only use information from rumour exchange with the opponent that they are engaged with at that time. Other players choose cooperation or defection, depending on the information that they have stored regarding the previous behaviour of the opponent.

Prisoner's Dilemma game and rule 1

Two players, after having played the Prisoner's Dilemma game with each other once, receive pay-offs as shown in Table 1. During a generation, player i accumulates, with respect to player j, the number of past instances of cooperation ($pc_i(j)$) and the number of past instances of defection ($pd_i(j)$) that have occurred in playing the Prisoner's Dilemma game with that player (see Fig. 1B). For example, $pd_{10}(4) = 1$ means that player 10 has suffered defection by player 4 once in the Prisoner's Dilemma game. Thus, the number of times player i has played the game with player j equals the sum of $pc_i(j)$ and $pd_i(j)$. Initially, $pc_i(j)$ and $pd_i(j)$ are zero for each generation.

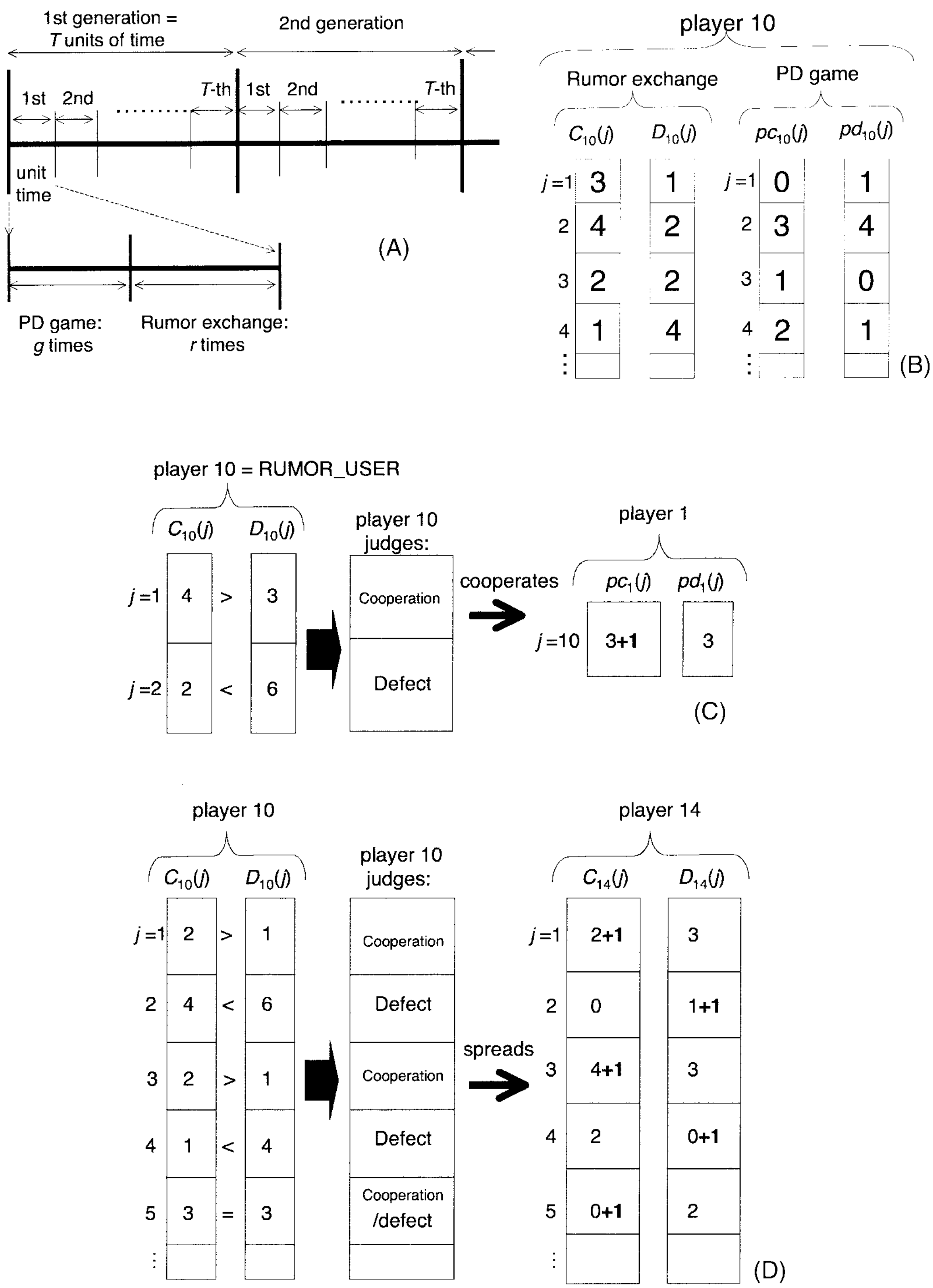
A player participating in the Prisoner's Dilemma game conducts himself based on having one of three traits, that of a RUMOR_USER, a DEFECTOR or a TFT_LIKE. RUMOR_USER chooses either cooperation or defection based on the rumours it has received (see Fig. 1C). Figure 1C shows that player 10 judges that player 1 is cooperative because $C_{10}(1) > D_{10}(1)$, and then he cooperates with player 1 when player 10 meets player 1 in the next game. After this Prisoner's Dilemma game, $pc_1(10)$ is increased by 1. DEFECTOR only defects its partner. TFT_LIKE, on the other hand, makes decisions based on its experience, in a manner such as tit-for-tat (TFT; Axelrod and Hamilton, 1981), in a repeated Prisoner's Dilemma game. We explain this in the Appendix.

Rumour exchange

Rumours are assumed to spread through personal one-to-one communication. Two players exchange rumours (as pieces of information) after being randomly chosen from the population. A rumour about a given player includes information about cooperation, resulting in either a positive outcome for that player (positive rumour – cooperative) or a negative outcome (negative rumour – defection). Each player keeps memorizing the number of positive and negative rumours he receives about each of the other players (Fig. 1B) and does so with 'mathematically' perfect recall. $C_i(j)$ is the number of times that player i has received a positive rumour about player j, and $D_i(j)$ is the number of times he has received a negative rumour about player j (see Fig. 1B). For example, $C_{10}(2) = 4$ means that player 10 has received positive rumours about player 2 four times. Initially, $C_i(j)$ and $D_i(j)$ are zero (for all i and j) for each generation. When $i = j$ (i.e. the other player is oneself), $C_i(j)$ and $D_i(j)$ are always zero, except for a player with the trait OWN_RUMOR_STARTER (one trait for starting rumours, rule 3), as we will explain later. We assume two rules for rumour exchange: a rule for starting rumours and a rule for spreading rumours. Both rules are applied simultaneously during a rumour exchange. In the following, we explain these rules in greater detail.

Rule 2: for spreading rumours

Suppose player i receives more positive rumours about player j (j is cooperative) than negative rumours (j is a defector). Then player i, presuming that player j is cooperative, spreads positive rumours about player j. If $C_i(j) < D_i(j)$, player i spreads a negative rumour about player j. In the case of $C_i(j) = D_i(j) > 0$, player i spreads either a positive or a negative rumour about player j with probability 0.5. However, if $C_i(j) = D_i(j) = 0$, then player i spreads no rumour about player j. As an example (see Fig. 1D), player 10 receives a positive rumour about player 1 two times ($C_{10}(1) = 2$) and a negative rumour once ($D_{10}(1) = 1$). Player 10 then supposes that player 1 is cooperative because $C_{10}(1) > D_{10}(1)$. If player 10 meets player 14 and uses a strategy of spreading all the rumours he received, then player 10 tells player 14 that player 1 is cooperative. Thus, player 14 receives a positive rumour about player 1, so $C_{14}(1)$ increases by 1 (Fig. 1D). We assume there are no perception errors, no transition error rates and no noise, implying that players who know that player i is cooperative never mistakenly spread rumours that player i is a defector, and that a player who receives a rumour that player i is cooperative never mistakenly communicates to another player anything inconsistent with what he received.


1st generation =
T units of time
2nd generation
1st
2nd
T-th
1st
2nd
T-th
unit
time
PD game:
g times
Rumor exchange:
r times
(A)
player 10
Rumor exchange
PD game
$C_{10}(j)$
$D_{10}(j)$
$pc_{10}(j)$
$pd_{10}(j)$
j =1 3 1 j =1 0 1
2 4 2 2 3 4
3 2 2 3 1 0
4 1 4 4 2 1
(B)
player 10 = RUMOR_USER
$C_{10}(j)$
$D_{10}(j)$
j =1 4 > 3
j =2 2 < 6
player 10
judges:
Cooperation
Defect
cooperates
player 1
$pc_1(j)$
$pd_1(j)$
j =10 3+1 3
(C)
player 10
$C_{10}(j)$
$D_{10}(j)$
j =1 2 > 1
2 4 < 6
3 2 > 1
4 1 < 4
5 3 = 3
player 10
judges:
Cooperation
Defect
Cooperation
Defect
Cooperation
/defect
spreads
player 14
$C_{14}(j)$
$D_{14}(j)$
j =1 2+1 3
2 0 1+1
3 4+1 3
4 2 0+1
5 0+1 2
(D)

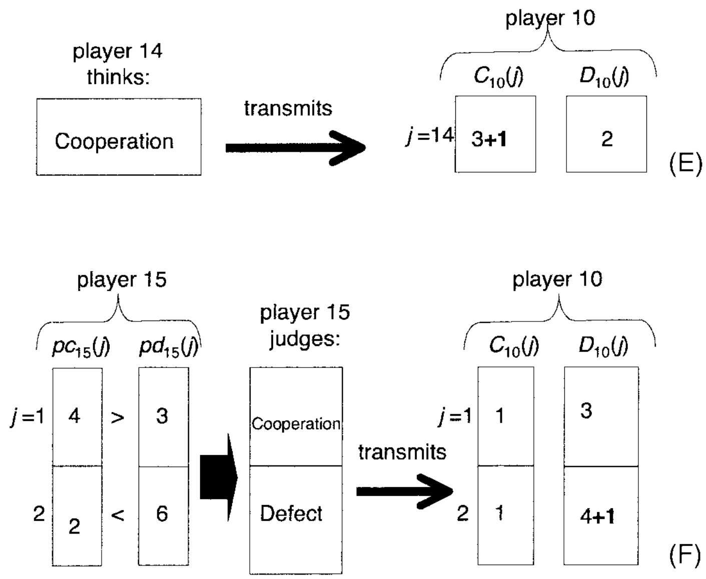


Fig. 1. The structure of the model. (A) For the time course of the model, each generation consists of T time units. In a single time unit, the Prisoner's Dilemma game occurs g times and rumour exchange occurs r times. An accumulated pay-off of each player at the end of each generation affects the number of players of each kind in the next generation. The ratio r/g is the index of the speed at which rumours spread. (B) Each individual has four internal states: $C_i(j)$, $D_i(j)$, $pc_i(j)$ and $pd_i(j)$. Two states, $C_i(j)$ and $D_i(j)$, are the outcomes of rumour exchange, and the other two states, $pc_i(j)$ and $pd_i(j)$, are the outcomes of the Prisoner's Dilemma game. (C) In a Prisoner's Dilemma game, a RUMOR_USER plays as follows. If player 10 judges that player 1 is cooperative because $C_{10}(1) > D_{10}(1)$, he then cooperates with player 1 when player 10 meets player 1 in the next game. After this Prisoner's Dilemma game, $pc_1(10)$ is increased by 1. In the figure, '+ 1' indicates an increment of $pc_1(10)$. (D) The spreading of an ALL_RUMOR_SPREADER. When player 10 meets player 14 in rumour exchange, player 10 spreads a rumour that players 1 and 3 are cooperative, because $C_{10}(1) > D_{10}(1)$. If player 10 also spreads a rumour that players 2 and 4 are defectors, because $C_{10}(2) < D_{10}(2)$, then $C_{14}(1)$ and $C_{14}(3)$ are increased by 1, and $D_{14}(2)$ and $D_{14}(4)$ are increased by 1, respectively. In the case of $C_{10}(5) = D_{10}(5)$, player 10 spreads either a positive or a negative rumour about player 5 with a probability of 0.5; if $C_i(j) = D_i(j) = 0$, however, then player i spreads no rumours about player j. (E) To illustrate the starting of the rumour 'I am cooperative' by an OWN_RUMOR_STARTER, suppose player 14 is either HONEST or LIAR. When player 14 meets player 10 in rumour exchange, player 14 transmits this self-promotional rumour to player 10, and $C_{10}(14)$ is increased by 1. (F) ADVISOR (player 15, for example) starts a rumour as follows after judging that player 2 is a defector because $pc_{15}(2) < pd_{15}(2)$. When player 15 meets player 10, he transmits that player 2 is a defector. As a result, $D_{10}(2)$ is increased by 1.

Here we explain five kinds of traits. They can be roughly classified into three types corresponding to simple behaviours that a player can have that reflect the player's effect on rumours and their spread. Players with the first type of trait spread all the rumours they have received except those about themselves (ALL_RUMOR_SPREADER); players with the second type of trait never spread rumours (NO_RUMOR_SPREADER); and players with the last type of trait spread rumours only under some conditions

(GOOD_RUMOR_SPREADER, BAD_RUMOR_SPREADER or CONDITIONAL_RUMOR_SPREADER). In the Appendix we explain each trait in detail.

Rule 3: for starting rumours

Rule 3 includes four possible traits, which fall into three types: players of the first type start self-promotional rumours (OWN_RUMOR_STARTER; see Fig. 1E); players of the second type start rumours based on a player's outcomes for the Prisoner's Dilemma game (GOOD_RUMOR_STARTER, BAD_RUMOR_STARTER); and players of the third type start no rumours (NO_RUMOR_STARTER). If a player is characterized by the OWN_RUMOR_STARTER trait (see Fig. 1E) and that player is in fact cooperative, the rumour is true and serves to signal correct information, honestly. However, if that player is a defector, the rumour is a lie and serves to signal false information, dishonestly. Figure 1E shows that player 14, when meeting with player 10 in rumour exchange, transmits such self-promotional rumour to player 10 and thus $C_{10}(14)$ increases by 1. Second, on the basis of the outcome of the Prisoner's Dilemma game (GOOD_RUMOR_STARTER, BAD_RUMOR_STARTER), suppose a player memorizes whether or not his partner cooperated; then the player can spread a rumour based on this memory. Figure 1F shows that player 15 (ADVISOR) judges that player 2 is a defector because $pc_{15}(2) < pd_{15}(2)$. When player 15 meets player 10, he tells player 10 that player 2 is a defector. As a result, $D_{10}(2)$ increases by 1. Here, we also assume no perception or transition errors (error rate = 0) and no noise. A detailed explanation of each trait can be found in the Appendix.

Strategies

As mentioned, a player's strategy consists of one trait for each of the three rule categories. Tables 2, 3 and 4 show some of the possible combinations, each combination constituting a strategy in this paper. In total, there are 60 possible strategies. We now consider the strategies listed in the tables, classified into three groups based on the three rules for the Prisoner's Dilemma game: that of the RUMOR_USER group (Table 2), that of the DEFECTOR group (Table 3) and that of the TFT_LIKE group (Table 4). For some key strategies, we use the following abbreviated names: LIAR, HONEST, ADVISOR, GOOD and TFT_LIKE. LIAR is a lying defector, HONEST is one who uses an 'honest signalling' strategy in terms of the classical signalling theory, ADVISOR is one who starts a negative rumour about a player who defected him, and GOOD is one who starts a positive rumour about a player who has cooperated with him. TFT_LIKE neither starts nor transmits rumours and plays the Prisoner's Dilemma game based on his own experiences.

RESULTS

Evolutionary dynamics

Focusing on the effect of spreading rumours, we used 39 of the 60 possible strategies. We excluded the strategies of NO_RUMOR_SPREADER in rule 2. In the following sections, the three most important strategies appear: two strategies of CONDITIONAL_RUMOR_SPREADER in rule 2 and the strategy of All-Defector – that is, always defect and never start or spread rumours. Using the 39 selected strategies (see Tables 2, 3 and 4),

Table 2. The possible strategies for players (group i) with the RUMOR_USER trait for rule category 1

Rule 2	Rule 3	Abbreviated name
(i) ALL_RUMOR_SPREADER	(i) OWN_RUMOR_STARTER	HONEST
	(ii) BAD_RUMOR_STARTER	ADVISOR
	(iii) GOOD_RUMOR_STARTER	GOOD
	(iv) NO_RUMOR_STARTER	
(ii) GOOD_RUMOR_SPREADER	(i) OWN_RUMOR_STARTER	
	(ii) BAD_RUMOR_STARTER	
	(iii) GOOD_RUMOR_STARTER	
	(iv) NO_RUMOR_STARTER	
(iii) BAD_RUMOR_SPREADER	(i) OWN_RUMOR_STARTER	
	(ii) BAD_RUMOR_STARTER	
	(iii) GOOD_RUMOR_STARTER	
	(iv) NO_RUMOR_STARTER	
(iv) CONDITIONAL_RUMOR_SPREADER	(i) OWN_RUMOR_STARTER	CONDITIONAL_HONEST
	(ii) BAD_RUMOR_STARTER	CONDITIONAL_ADVISOR

Table 3. The possible strategies for players (group ii) with the DEFECTOR trait for rule category 1

Rule 2	Rule 3	Abbreviated name
(i) ALL_RUMOR_SPREADER	(i) OWN_RUMOR_STARTER (ii) BAD_RUMOR_STARTER (iii) GOOD_RUMOR_STARTER (iv) NO_RUMOR_STARTER	LIAR
(ii) NO_RUMOR_SPREADER	(iv) NO_RUMOR_STARTER	All-Defect (AD)
(iii) GOOD_RUMOR_SPREADER	(i) OWN_RUMOR_STARTER (ii) BAD_RUMOR_STARTER (iii) GOOD_RUMOR_STARTER (iv) NO_RUMOR_STARTER	LIAR1
(iv) BAD_RUMOR_SPREADER	(i) OWN_RUMOR_STARTER (ii) BAD_RUMOR_STARTER (iii) GOOD_RUMOR_STARTER (iv) NO_RUMOR_STARTER	LIAR2

Table 4. The possible strategies for players (group iii) with the TFT_LIKE trait for rule category 1

Rule 2	Rule 3	Abbreviated name
(i) ALL_RUMOR_SPREADER	(i) OWN_RUMOR_STARTER (ii) BAD_RUMOR_STARTER (iii) GOOD_RUMOR_STARTER (iv) NO_RUMOR_STARTER	
(iii) GOOD_RUMOR_SPREADER	(i) OWN_RUMOR_STARTER (ii) BAD_RUMOR_STARTER (iii) GOOD_RUMOR_STARTER (iv) NO_RUMOR_STARTER	
(iv) BAD_RUMOR_SPREADER	(i) OWN_RUMOR_STARTER (ii) BAD_RUMOR_STARTER (iii) GOOD_RUMOR_STARTER (iv) NO_RUMOR_STARTER	

the results of the simulation showed that not only T but also the number of rumour exchanges (r) and the number of Prisoner's Dilemma games (g) per unit time greatly affected the success of RUMOR_USER. When $w_g < 1$, only the DEFECTOR group defeated others. Then, LIAR, LIAR1 and LIAR2 (see Table 3) of the DEFECTOR group were only able to win when w_g was near 1. When w_g was greater than 1 (w_g is around 2–3 in Fig. 2), cooperative players could defeat the DEFECTOR group. When w_g was sufficiently large (greater than about 3), so that each individual played the Prisoner's Dilemma game with every other individual many times within a generation, the simulation results no longer depended on r, and the model's inherent selection process always favoured the TFT_LIKE group and/or the RUMOR_USER group.

Upon varying r and g, however, the outcomes were different for two conditions, that of fixing the number of rumours exchanged ($T \times r$) and that of fixing, in addition, the number of Prisoner's Dilemma games ($T \times g$) per generation. We assumed $T \times r$ equals $T \times g$, which means that g equals r. Both the RUMOR_USER group and the TFT_LIKE group were favoured (by the selection process) when T was small and g and r were large (Fig. 2), but only the TFT_LIKE group was favoured when T became larger and g and r in w_g and w_r were small. This implies that with w_g and w_r both fixed, an increase in the fixed value for g promoted an increase in the eventual size of the RUMOR_USER group, because larger values of g enabled individuals in that group to obtain more accurate information, in that players could then play more Prisoner's Dilemma games with others before the next rumour exchange started.

The effects of rumour exchange rate and the number of Prisoner's Dilemma games

We expected that the qualitative results of the simulation would be independent of rule 2 – that is, independent of different traits within rule 2, which determined the way players spread rumours. Thus, to concentrate on the effect of rule 3 in further simulations, we restricted the population to a subset of the initial 39 strategies. For the DEFECTOR group, we used only LIAR (Table 3); for the RUMOR_USER group, we used only HONEST, ADVISOR and GOOD (see Table 2). These resulting populations had only the ALL_RUMOR_SPREADER trait for rule 2 of their strategies. If all the players only had either GOOD_RUMOR_SPREADER or BAD_RUMOR_SPREADER for rule 2, the kinds of rumours that occurred constituted a biased set, making examination of the role of rumours in discriminating defectors ineffective. CONDITIONAL_HONEST and CONDITIONAL_ADVISOR, having the trait CONDITIONAL_RUMOR_SPREADER for rule 2, were also used to consider the reliability of rumours.

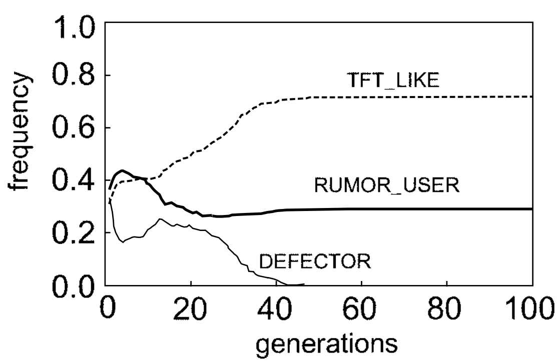


Fig. 2. Change (over generations) in frequencies of three groups of strategies. We carried out an individual-based simulation with a population that takes all 39 strategies into account. Each strategy is initially represented by five players, and thus the population has 195 individuals ($N = 195$). The bold, thin and dashed lines indicate, respectively, the RUMOR_USER, DEFECTOR and TFT_LIKE groups of strategies. Parameter values are $g = 5$, $r = 5$ and $T = 10{,}000$ ($w_g = 2.64$).

(1) *The results of the IBM model with only HONEST, ADVISOR, GOOD and LIAR*

Figure 3 shows that it was possible for the three forms of RUMOR_USER (HONEST, ADVISOR and GOOD) to drive LIAR to extinction only when r/g was low and w_g was larger than 1. When r/g was low (0.2), the three cooperative strategies of RUMOR_USER were more likely to be favoured and prevail than the LIAR strategy. On the other hand, when r/g was 0.4 or more, LIAR was favoured over any of the three RUMOR_USERs in all cases. When r/g was close to zero (no rumours), or more than 0.4, LIAR always won. To examine which parameters affected this result, we next conducted a simulation with only two strategies in the population, those of GOOD and LIAR.

(2) *A population with only GOOD and LIAR*

The simulations were repeated 100 times for each set of parameter values, with N and g set to 40 and 5, respectively, r ranging from 1 to 10 and T ranging from 100 to 2000. The initial frequency of GOOD was set at either 0.1 or 0.9. In such a population, LIAR always won irrespective of r/g or the initial frequency of GOOD, because GOOD only started rumours about cooperative players and could not distinguish between the true rumours he had started and false rumours started by LIAR.

(3) *A population with only HONEST and LIAR*

Other populations with only the strategies (i.e. groups) HONEST and LIAR were simulated. For such populations, when the parameter set was the same as that used for GOOD versus LIAR, LIAR always won again. This result showed that when a member of the HONEST group promoted itself as cooperative, another such player could benefit by cooperating. At the same time, however, a member of LIAR can deceive a member of HONEST in the same way that it can deceive a member of GOOD.

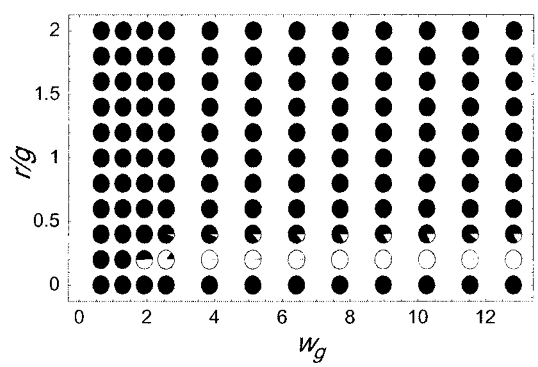


Fig. 3. Outcomes of competition among four strategies: HONEST, ADVISOR and GOOD of the RUMOR_USER group, and LIAR. The ratio r/g is the number of rumour exchanges per Prisoner's Dilemma game, and w_g is the expected number of Prisoner's Dilemma games per individual per generation. The black portion of each circle indicates the fraction of runs in which LIAR won in 100 simulations; the open portion represents the fraction of runs in which the RUMOR_USER group of strategies won. The population size (N) is 40, with 10 individuals for each strategy, initially. Parameter values are $g = 5$, $r = 0$–10 ($0 \le r/g \le 2$) and $T = 100$–2000 ($0.641 \le w_g \le 12.8$).

(4) *A population with only ADVISOR and LIAR*

For these simulations, only two strategies, ADVISOR and LIAR, compete in the population. Initially, the same parameter sets as defined in (2) and (3) above were used.

LIAR always won against ADVISOR when the initial density of ADVISOR was low (less than around 0.3 in $g = 5$). However, ADVISOR was able to win against LIAR in some parameter (r/g) regions, for example when the initial density of ADVISOR was not low (more than around 0.3). It is interesting that Fig. 4A shows that the slow speed of spreading a rumour ($r < g$) enabled ADVISOR to win against LIAR, even though the initial frequency of ADVISOR was not high (0.5). When $r = 0$, which means that no players started or spread rumours, ADVISOR was not favoured, also when w_g was small and r/g was large (Fig. 4), which suggests that ADVISOR was at a selective disadvantage when rumours spread very quickly. Otherwise, ADVISOR was favoured.

The open circles in Fig. 4B were interpreted as being an evolutionarily stable strategy (ESS) region for ADVISOR, which meant that the population, initially occupied by ADVISOR, could not be invaded by a few LIARs in the simulation. Even in those cases where w_g was small (less than 1 and more than 0.4), ADVISOR was still able to be an ESS.

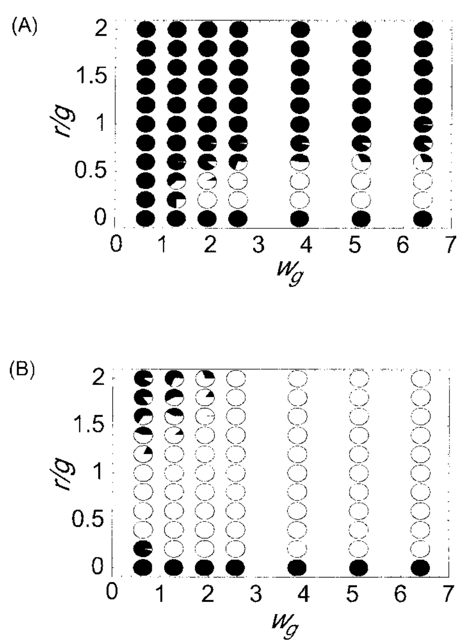


Fig. 4. Outcomes of competition between two strategies, ADVISOR and LIAR. The black portion of each circle indicates (for 100 simulations) the fraction of runs in which LIAR won; the open portion indicates the fraction of runs in which ADVISOR won. The ratio r/g is the number of rumour exchanges per Prisoner's Dilemma game, and w_g is the expected number of Prisoner's Dilemma games per individual per generation. Parameter values are $g = 5$, $N = 40$, $r = 0$–10. The initial frequency of ADVISOR is 0.5 for (A) and 0.9 for (B); $T = 100$–1000 ($w_g = 0.64$–6.41) for (A) and (B). The open circles in (B) are for the ESS region of ADVISOR. In (B), when $w_g < 0.4$, LIAR always won irrespective of r/g.

We also examined the cases when $g = 1$ and 10, and we found that the same value of r/g resulted in almost the same result, even though g was varied.

To study further the conditions under which ADVISOR was favoured (i.e. replaces LIAR), additional parameter values were explored. Figure 5 shows the results of varying w_g and the initial frequency of ADVISOR as well, when $r = 5$. These results indicate that if w_g was not small, a relatively high initial frequency enabled ADVISOR to win among these parameter values, ADVISOR being most likely to succeed when $r < g$: the higher the value of r, the less likely selection would favour ADVISOR, with $g = 5$ (Fig. 4). Figures 4 and 5 show in addition that, when r was small, the region in which ADVISOR could win against LIAR was large. This implies that even though the initial frequency of ADVISOR was not low, and r was small, rumours were corrected; as a result, ADVISOR avoided being deceived by LIAR.

From the results of (2)–(4), we concluded that only ADVISOR won against LIAR, and that HONEST and GOOD did not work at all in Fig. 3.

The reputation of a player j is taken to be $F(j) = \sum_i [REP_i(j)]$, where i is varied over all players and $REP_i(j)$ is 1 for $C_i(j) > D_i(j)$, 0 for $C_i(j) = D_i(j)$, and −1 for $C_i(j) < D_i(j)$. The average degree of reputation for cooperative behaviour of strategy l is taken to be

$$Z(l) = \frac{1}{N_A} \sum_{j}^{N_A} F(j)$$

where j is varied over all players with strategy l and N_A is the number of such players. A positive $Z(l)$ indicates that strategy l has a good reputation on average. If $Z(l) < 0$, then strategy l has a negative reputation and is considered to be a defector (on average). Figure 6A shows the results of averaging the reputation (Z) of all players using the LIAR strategy ($r = 1$ and the initial frequency of ADVISOR was 0.2). When the initial frequency of ADVISOR was 0.2, players had a tendency to be deceived by LIAR, resulting in the average fitness of ADVISOR being lower than that of LIAR. When the initial frequency of ADVISOR was high (0.9), correct rumours about LIAR spread (the average degree of reputation became less than 0 as T increased), and with time the average fitness of ADVISOR became higher than that of LIAR.

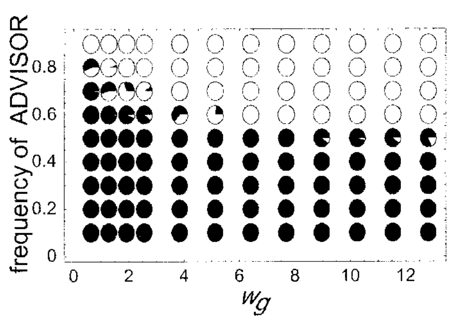


Fig. 5. The black portion of each circle indicates (for 100 simulations) the fraction of runs in which LIAR won; the open portion represents the fraction of runs in which ADVISOR won. ADVISOR's initial frequency is indicated on the vertical axis. The other parameters are $g = 5$, $r = 5$, $N = 40$ and $T = 100$–2000 ($w_g = 0.64$–12.82).

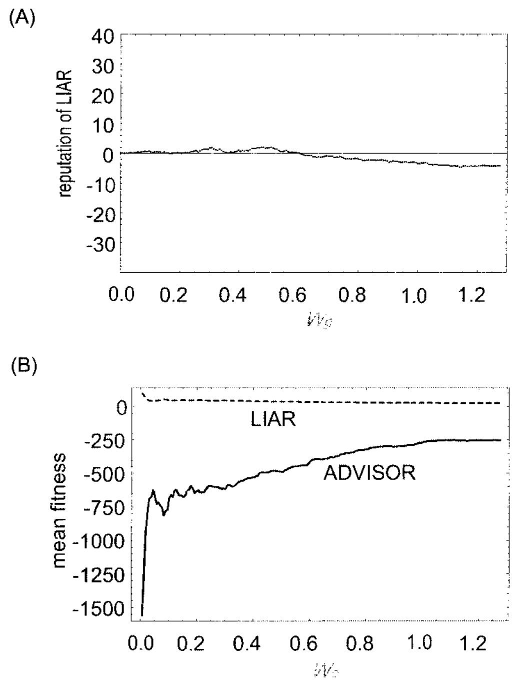


Fig. 6. (A) Change in the average number of positive (cooperative) or negative (defector) reputations about a LIAR member (Z). The initial frequency of ADVISOR is 0.2 and that of LIAR is 0.8. (B) Change over time in the average fitness of the two strategies, obtained by dividing the sum of scores for individuals having a certain strategy by w_g. Parameter values for (A) and (B) are $r = 1$, $g = 5$, $T = 0$–200 ($w_g = 0$–1.28).

(5) *A population with only TFT_LIKE and LIAR*

Two strategies, TFT_LIKE (choosing either cooperation or defection based on past experience with a partner) and LIAR (always defecting, and the same as AD in this simulation) competed for selection advantage in a population. This simulation, which served as a control, was conducted to determine how effectively the use of rumour could facilitate the evolution of cooperation. Thus, we used the same parameter sets as in (4) above.

When varying the initial frequency of TFT_LIKE and w_g, Fig. 7 shows (as the outcome for TFT_LIKE) that TFT_LIKE can be an ESS if $w_g > 1$; and if w_g is large enough, then TFT_LIKE, as a population, could invade and replace LIAR even though the initial frequency of TFT_LIKE was very low. Comparing Fig. 5 with Fig. 7, it is apparent that ADVISOR, but not TFT_LIKE, was able to beat LIAR when $w_g < 1$. This implies that the use of rumours was able to facilitate an increase in the frequency of ADVISOR relative to TFT_LIKE when any given pair of players (i.e. partners) played the Prisoner's Dilemma game less than once and when the rumour exchange rate was not too high.

The evolution of reliability

In the previous sections, we only examined strategies by which all the received rumours are spread. This assumption, in failing to consider the reliability of rumours, is unrealistic. We

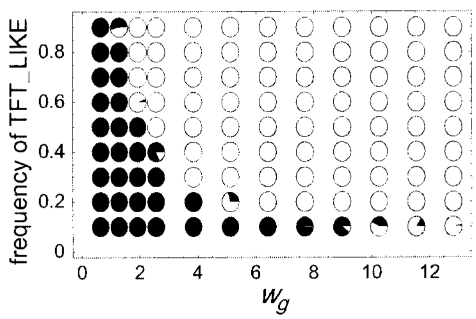


Fig. 7. Outcomes of competition between strategies LIAR and TFT_LIKE. The initial frequency of TFT_LIKE is shown on the vertical axis; w_g is the expected number of Prisoner's Dilemma games per individual per generation; $N = 40$. The black portion of each circle indicates (for 100 simulations) the fraction of runs in which LIAR won; the open portion indicates the fraction of runs in which TFT_LIKE won. The range for w_g was 0.64–12.82. When $w_g < 1.0$, LIAR always won irrespective of the initial frequency of TFT_LIKE.

therefore also examined strategies CONDITIONAL_HONEST and CONDITIONAL_ ADVISOR. A player of these strategies spreads the rumour that a partner is cooperative only when that partner has cooperated with the player. Because these two strategies spread reliable rumours, we studied whether they can succeed against LIAR (see Table 2).

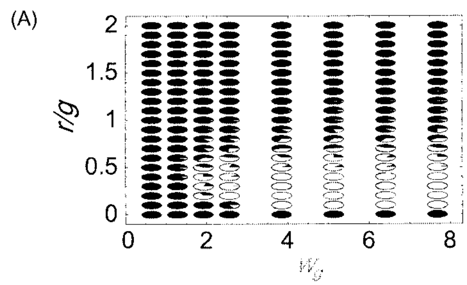
(1) *A population with CONDITIONAL_HONEST and LIAR only*

In these simulations, CONDITIONAL_HONEST and LIAR competed in the population. The initial frequency of CONDITIONAL_HONEST was varied, as were rumour exchange rate and $T \times g$. In all parameter ranges, LIAR always beat CONDITIONAL_HONEST, because CONDITIONAL_HONEST could prevent a lie from spreading, but did not start the rumour that LIAR was a defector. As a result, CONDITIONAL_HONEST always cooperated with LIAR.

(2) *A population with CONDITIONAL_ADVISOR and LIAR only*

Using the same parameters as (1), we studied the case including CONDITIONAL_ ADVISOR and LIAR only. The simulation results for CONDITIONAL_ADVISOR versus LIAR were the same as those for ADVISOR versus LIAR when r/g was less than 1. For $r/g > 1$, however, CONDITIONAL_ADVISOR was more likely to be an ESS against LIAR than was ADVISOR, even though w_g was not high (less than 2 in Fig. 8B). When r was higher and w_g was high, CONDITIONAL_ADVISOR was able to win against LIAR in lower initial frequencies of CONDITIONAL_ADVISOR (more than around 0.2; see legend of Fig. 8A as an example). This result indicates that the trait CONDITIONAL_ RUMOR_SPREADER for rule 2 allowed CONDITIONAL_ADVISOR to avoid spreading the false rumours started by LIAR and to avoid being deceived by those rumours – even when the rumour exchange rate was fast. When the initial frequency of LIAR was not so high, then (among all the other strategies that we considered) CONDITIONAL_ ADVISOR was the most likely to succeed against LIAR.

We were able to show that for the same value of r/g but with different g (1 and 5) resulted in almost the same result.



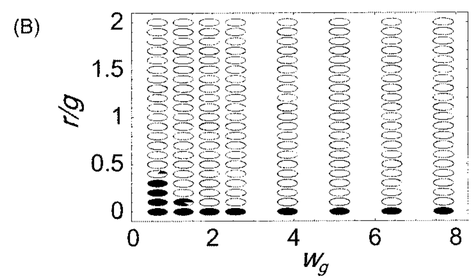


Fig. 8. Outcomes of competition between strategies CONDITIONAL_ADVISOR and LIAR. The parameter w_g is the expected number of Prisoner's Dilemma games per individual per generation; $g = 10$, $N = 40$ and $r = 0$–20 ($0 \leq r/g \leq 2$). The black portion of a circle indicates (for 100 simulations) the number of times LIAR wins; the open portion indicates the number of times CONDITIONAL_ADVISOR wins. The initial frequency of ADVISOR is 0.4 for (A) and 0.9 for (B); $T = 50$–600 ($w_g = 0.64$–7.69) for (A) and (B). The open circles in (B) designate the ESS region of CONDITIONAL_ADVISOR. In (B), when $w_g < 0.3$–0.4, LIAR always wins, irrespective of r/g.

DISCUSSION

We examined the hypothesis that the starting and spreading of rumours about a player's reputation can help in detecting defectors and liars and thereby facilitate the potential for cooperation-fostering traits to succeed evolutionarily as elements of strategy. When players can interact with others more than once in a population that includes players with various types of strategies, rumours can prevent defectors and liars from exploiting cooperators as a means of increasing their own fitness values at the expense of cooperative strategies.

Among the various strategies that make use of rumour, ADVISOR, whose players accuse defectors by starting rumours about them being defectors, is the most effective in preventing LIAR (a population of dishonest players) from exploiting cooperators. Even though HONEST and GOOD advertise their own cooperativeness – and that of others – and succeed in maintaining the cooperative relationship among cooperative players, they are easily deceived by LIAR. ADVISOR only has the advantage, however, when a rumour is not spread too quickly from one individual to the next, since if a dishonest or cooperation-damaging rumour is spread too quickly, correcting it becomes difficult. As a result, dishonest rumours spread by members of LIAR can easily control relationships among individuals, even though there are only few LIAR players in the population. Towards the other end of the spectrum, if dishonest rumours spread too slowly, then a rumour can itself cease to be effective in influencing the choice of cooperation or defection in the Prisoner's Dilemma game.

In contrast to the ADVISOR, a population of CONDITIONAL_ADVISOR, whose members only receive conditional rumours, cannot be invaded by LIAR even when rumours spread quickly. This is because CONDITIONAL_ADVISOR can neglect a dishonest rumour produced by a liar: when rumours spread quickly, CONDITIONAL_ADVISOR players receive correct rumours quickly. However, given the condition that players interact with the same partner less than once and that TFT_LIKE is invaded by LIAR, then both ADVISOR and CONDITIONAL_ADVISOR cannot be invaded by LIAR and can be evolutionarily stable strategies. Moreover, in the same condition, LIAR cannot be invaded by them. This holds because neither ADVISOR nor CONDITIONAL_ADVISOR can correct wrong rumours started by numerous LIAR members, when overwhelmed by them. Hence we have bistability.

By starting correct rumours about defectors, ADVISOR members accuse defectors and impose sanctions on them. Having the simple role of policing or punishment in societies (Clutton-Brock and Parker, 1995; Fehr and Gachter, 2002), rumour is a means of maintaining cooperation in the population. Although theoretical studies have assumed that players pay a cost for punishment, spreading or starting a rumour need not itself incur a cost. Hence, in this study we assumed that players pay nothing for spreading or starting a rumour; as a consequence, the behaviour of spreading and starting rumours, as a trait for discriminating and punishing defectors, is more likely to increase in frequency and effectiveness than costly punishment or policing.

In the case of individual interactions – for instance, to attract more females – males might exaggerate their signals such as conspicuous coloured ornaments. If the signals are not related to any quality of the males, the females would be cheated by the false signals. However, it has been considered that the cost of signalling prevents the spread of cheating signalling, and honest signals are likely to be maintained (e.g. Zahavi, 1975, 1977; Grafen, 1990; Godfray, 1991; Iwasa *et al.*, 1991). A recent paper (Lechmann *et al.*, 2001) showed that honest signalling is selected for even when the cost of such signalling is slight. Cheating, however, can evolve both within a species and between species. For example, the well-known behaviours of brood parasites (certain cuckoos and cowbirds) and Batesian mimicry (tephritid flies, burrowing owls and caterpillars of several moths) (e.g. Alcock, 1998) can be considered cheating. As cheated individuals evolve a mechanism for detecting cheaters, cheaters should continue to evolve more sophisticated tactics to cheat them, which leads to an arms race between cheated and cheaters.

No cost is incurred when using language, and in this sense our study is different from classical signalling theory. When telling a lie and punishing defectors results in no direct costs, an individual using an honest signalling strategy, such as HONEST, is cheated by a lying defector and is not favoured by selection, as in classical signalling theory. Our results indicate that ADVISOR was the most important strategy for the evolution of cooperation. Wilson *et al.* (2000) showed that gossip works for detecting a norm violation to maintain cooperation, which supports our results.

Pollock and Dugatkin (1992) assumed that players (called O-TFT) observe others playing a Prisoner's Dilemma game and use that information to individually construct reputations which they assign to other players. They showed that when a player doesn't interact with the same partner frequently, O-TFT can succeed (increase in frequency), even though TFT cannot. Our results also showed that cooperative players who start or spread rumours can also succeed under the condition that cooperative players who play the Prisoner's Dilemma game based on their individual past experiences cannot succeed.

Even if rumours spread relatively quickly, CONDITIONAL_ADVISOR can still succeed, because CONDITIONAL_ADVISOR can correct false rumours spread by LIAR. Rosnow (1991) showed that the spreading of rumours is strongly related to three factors: the anxiety that individuals feel, ambiguity of information and reliability of information. Our results also indicate that the reliability of information spread by rumours about defectors is important for the evolutionary success of rumours in fostering cooperative behaviour.

Other researchers have studied defecting defectors from the evolutionary point of view of the human society. For example, using the Wason Selection task with the sharing rule, Hiraishi and Hasegawa (2001) showed that humans have a strong tendency to try to detect free-riders (coming from the out-group). We have also shown that people have a trait such as ADVISOR. In addition, the commitment problem (Frank, 1988) can also be related to the evolution of a trait such as ADVISOR. When people find themselves deceived, they lose their temper, and when angry they start a bad rumour. Such emotional behaviour is indecent and irrational in some societies, but is adaptive from the viewpoint of the commitment problem (Frank, 1988), because others can get information about deceptive people and thus avoid interacting with them, thereby maintaining cooperative relationships.

Many studies on rumours in human society have been conducted in the field of social science, and several researchers have examined the relationship between a rumour's character and its speed of spread. For example, personal communication, such as word-of-mouth, promotes the spread of rumours, but the media does not (e.g. Herr *et al.*, 1991). Tybout *et al.* (1981) showed how to deny a negative rumour or negative reputation effectively. Theories in physics have also been applied to the phenomenon of rumours (e.g. Liu *et al.*, 2001). These studies have focused on the spread of a rumour in human society; no studies, however, have analysed the behaviour of starting and spreading a rumour from the viewpoint of the theory of evolution. We hoped that the present study would contribute to the study of rumours in social science.

In the present study, to derive general principles of the influence of rumours on the dynamics of cooperative behaviour, we adapted simplified rules for spreading and starting a rumour. The spreading, starting and receiving of rumours is much more complicated, detailed and elaborate in human society and, therefore, more realistic rules and strategies need to be considered to further expand our understanding of the evolution of cooperative behaviour. We think, however, that our simple and basic framework model can also be applied to these situations. For instance, players can be spiteful (e.g. Iwasa *et al.*, 1998) and say that 'he is a defector', even though he is cooperative. This strategy may cause the defection of a cooperative player and, as a result, make cooperative relationships hard to maintain. Nevertheless, even though bad rumours are received, individuals do have an ability to distinguish defectors and cooperators. Therefore, in future simulations, we will allow some players to play the Prisoner's Dilemma game using both the outcomes of previous games and the results of spreading rumours to determine if cooperation can be maintained in spite of the existence of spiteful players. Moreover, it may be interesting to allow players to observe other players' games and start a rumour, like O-TFT (Pollock and Dugatkin, 1992). Such strategies may play a role in enhancing the spread of correct rumours and in correcting information spread by bad rumours. Noise, spatial structures and social networks, like a small-world network and a power-law network, also need to be considered in future studies of the evolution of rumours (e.g. M. Nakamaru and A.S. Levin, in prep.).

ACKNOWLEDGEMENTS

This work was supported in part by two Grants-in-Aid for Scientific Research from the Ministry of Education, Science and Culture, Japan (No. 13304006 and No. 14740418) and by the CREST (Core Research for Evolutionary Science and Technology Risk Management Group) project of the Japan Science and Technology Corporation. We thank R. Axelrod, U. Dieckmann, M. Doebeli, P. Johnson, A. Sasaki, R. Riolo, K. Sigmund, T. Shimizu, K. Tsuji and P.M. Wolanin for their helpful comments.

REFERENCES

Alcock, J. 1998. *Animal Behavior*, 6th edn. Sunderland, MA: Sinauer Associates.

Axelrod, R. and Hamilton, W.D. 1981. The evolution of cooperation. *Science*, **211**: 1390–1396.

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

Clutton-Brock, T.H. and Parker, G.A. 1995. Punishment in animal societies. *Nature*, **373**: 209–216.

Dunbar, R.I.M. 1996. *Grooming, Gossip and the Evolution of Language*. Cambridge, MA: Harvard University Press.

Fehr, E. and Gachter, S. 2002. Altruistic punishment in humans. *Nature*, **415**: 137–140.

Frank, R.H. 1988. *Passions Within Reason*. New York: W.W. Norton.

Grafen, A. 1990. Biological signals as handicaps. *J. Theor. Biol.*, **144**: 517–546.

Godfray, H.C.J. 1991. Signalling of need by offspring to their parents. *Nature*, **352**: 328–330.

Herr, P.M., Kardes, F.R. and Kim, J. 1991. Effects of word-of-mouth and product-attribute information on persuasion: an accessibility-diagnosticity perspective. *J. Consumer Res.*, **17**: 454–462.

Hiraishi, K. and Hasegawa, T. 2001. Sharing-rule and detection of free-riders in cooperative groups: evolutionarily important deontic reasoning in the Wason Selection task. *Thinking and Reasoning*, **7**: 255–294.

Iwasa, Y., Pomiankowski, A. and Nee, S. 1991. The evolution of costly mate preferences. II. The ‘handicap’ principle, *Evolution*, **45**: 1431–1442.

Iwasa, Y., Nakamaru, M. and Levin, S.A. 1998. Allelopathy of bacteria in a lattice population: competition between colicin-sensitive and colicin-producing strains. *Evol. Ecol.*, **12**: 785–802.

Lechmann, M., Szamado, S. and Bergstrom, C.T. 2001. Cost and conflict in animal signals and human language. *Proc. Natl. Acad. Sci. USA*, **98**: 13189–13194.

Leimar, O. and Hammerstein, P. 2001. Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B*, **268**: 745–753.

Liu, Z., Luo, J. and Shao, C. 2001. Potts models for exaggeration of a simple rumour transmitted by recreant rumormongers. *Phys. Rev. E*, **64**: 046134–046142.

Maynard Smith, J. and Szathmary, E. 1995. *The Major Transitions in Evolution*. Oxford: Oxford University Press.

Milinski, M., Semmann, D. and Krambeck, H. 2002. Reputation helps solve the ‘tragedy of the commons’. *Nature*, **415:** 424–426.

Monnin, T. and Ratnieks, F.L. 2001. Policing in queenless ponerine ants. *Behav. Ecol.*, **50**: 97–108.

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

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

Riolo, R., Cohen, M.D. and Axelrod, R. 2001. Evolution of cooperation without reciprocity. *Nature*, **414**: 441–443.

Rosnow, R.L. 1991. Inside rumour: a personal journey. *Am. Psychol.*, **46**: 484–496.

Sigmund, K., Hauert, C. and Nowak, M.A. 2001. Reward and punishment. *Proc. Natl. Acad. Sci. USA*, **19**: 10757–10762.

Sober, E. and Wilson, D.S. 1998. *Undo Others: The Evolution and Psychology of Unselfish Behaviour*. Cambrigde, MA: Harvard University Press.

Tybout, A.M., Calder, B.J. and Sternthal, B. 1981. Using information processing theory to design marketing strategies. *J. Marketing Res.*, **18**: 73–79.

Wilson, D.S. and Dugatkin, L.A. 1997. Group selection and assortative interactions. *Am. Nat.*, **149**: 336–351.

Wilson, D.S., Wilczynski, C., Wells, A. and Weiser, L. 2000. Gossip and other aspects of language as group-level adaptations. In *The Evolution of Cognition* (C. Heyes and L. Huber, eds), pp. 347–365. Cambridge, MA: MIT Press.

Zahavi, A. 1975. Mate selection: a selection for a handicap. *J. Theor. Biol.*, **53**: 205–214.

Zahavi, A. 1977. The cost of honesty (further remarks on the handicap principle.) *J. Theor. Biol.*, **67**: 603–605.

APPENDIX

Three traits for rule 1: the Prisoner's Dilemma game

Trait 1: RUMOR_USER

A player with this trait chooses to either cooperate or defect based on the rumours received about a particular partner. For example, suppose player *i* has this trait and plays the Prisoner's Dilemma game with player *j* (see Fig. 1C). If player *i* has received more positive than negative rumours about player *j*, that is $C_i(j) > D_i(j)$, then player *i* cooperates with player *j*. Conversely, if $C_i(j) < D_i(j)$, then player *i* defects. When $C_i(j) = D_i(j) \neq 0$, player *i* chooses either cooperation or defection with a probability of 0.5. If player *i* has not yet received a rumour about player *j*, as when the model begins, for example, or later, then $C_i(j) = D_i(j) = 0$ and player *i* cooperates.

Trait 2: DEFECTOR

A player with this trait always defects against his partner in the Prisoner's Dilemma game.

Trait 3: TFT_LIKE

A player with this trait chooses to either cooperate or defect based on previous outcomes of the Prisoner's Dilemma game with a particular partner. That is, player *i* with this trait, instead of playing the Prisoner's Dilemma game with player *j* based on $C_i(j)$ and $D_i(j)$, chooses either cooperation or defection based upon his own past experience in the Prisoner's Dilemma game with player *j*. In short, a TFT_LIKE player uses only first-hand information (no second-hand information) – that is, he uses his own experience with another player in deciding to cooperate or defect when playing a Prisoner's Dilemma game with that player. If player *j* cooperated with player *i* in their latest contest or they are meeting for the first time, player *i* cooperates with player *j*. If player *j* defected against player *i* in their latest contest, player *i* defects player *j*.

Five traits for rule 2: spreading rumours

Trait 1: ALL_RUMOR_SPREADER

A player with this trait spreads rumours about all other players based on its accumulated rumours (Fig. 1D). Player *i* spreads the rumour that player *j* is cooperative if $C_i(j) > D_i(j)$, or that he is a defector if $C_i(j) < D_i(j)$. If $C_i(j) = D_i(j) \neq 0$ (e.g. $C_{10}(5) = D_{10}(5)$ in Fig. 1D), then player *i* spreads either a positive or a negative rumour (i.e. cooperative or defector) about player *j* with a probability of 0.5; but if $C_i(j) = D_i(j) = 0$, then player *i* spreads no rumour about player *j*.

Trait 2: NO_RUMOR_SPREADER

Players with this trait spread no rumours.

Trait 3: GOOD_RUMOR_SPREADER

A player with this trait spreads only positive rumours. That is, player *i* spreads the rumour that player *j* is cooperative if $C_i(j) > D_i(j)$. If $C_i(j) = D_i(j) \neq 0$, then player *i* spreads either a positive or a negative rumour about player *j* with a probability of 0.5; but if $C_i(j) < D_i(j)$, then no rumour is spread about player *j*.

Trait 4: BAD_RUMOR_SPREADER

A player with this trait spreads only negative rumours. That is, player *i* spreads the rumour that player *j* is a defector if $C_i(j) < D_i(j)$. If $C_i(j) = D_i(j) \neq 0$, then player *i* spreads either a positive or a negative rumour about player *j* with a probability of 0.5; but if $C_i(j) > D_i(j)$, then he spreads no rumours about player *j*.

Trait 5: CONDITIONAL_RUMOR_SPREADER

A player with this trait fosters the spread of positive rumours about other players and of rumours consistent with personally experienced cooperative conduct by other players. That is, player *i* with this trait spreads the rumour that player *j* is cooperative if (a) player *i* has received more positive than negative rumours about player *j* ($C_i(j) > D_i(j)$) and (b) player *j* has cooperated with player *i* more times than defected ($pc_i(j) > pd_i(j)$). If the number of positive rumours about player *j* equals the number of negative rumours ($C_i(j) = D_i(j)$), player *i* spreads a positive or a negative rumour about player *j* with a probability of 0.5. If the number of positive rumours about player *j* is less than the number of negative rumours ($C_i(j) < D_i(j)$), or if the number of times player *j* has cooperated with player *i* is less than or equal to the number of times player *j* has defected ($pc_i(j) \leq pd_i(j)$), player *i* spreads no rumours about player *j*.

Four traits for rule 3: starting rumours

Trait 1: OWN_RUMOR_STARTER

Player *i* with this trait, assuming that $C_i(i) = 1$ and $D_i(i) = 0$, starts a rumour 'I'm cooperative' (Fig. 1E). More specifically, if player *j* receives this rumour about player *i* (who has this trait), then $C_j(i)$ is increased by 1. If player *i* has the trait RUMOR_USER or TFT_LIKE for its rule 1, then player *i* has an honest strategy; but if he has the DEFECTOR trait, then he has a dishonest strategy (a liar), which can deceive his partner (player *j*). In such a case, player *i*, by spreading this untrue self-promotional rumour, benefits from the Prisoner's Dilemma game, resulting in $C_j(i)$ increasing by 1 despite player *i* having the DEFECTOR trait.

Trait 2: BAD_RUMOR_STARTER

Player *i* with this trait starts a negative rumour about any partner *j* who has defected more often than cooperated ($pc_i(j) < pd_i(j)$) (see Fig. 1F). If player *i* communicates this rumour to a third player *k*, then $D_k(j)$ is increased by 1. If player *j* cooperated and defected an equal (but non-zero) number of times (i.e. $pc_i(j) = pd_i(j)$), player *i* starts a negative rumour about *j* with a probability of 0.5.

Trait 3: GOOD_RUMOR_STARTER

A player with this trait starts a positive rumour about any cooperative partners, and does so in a manner exactly converse to that of a player with the BAD_RUMOR_STARTER trait.

Trait 4: NO_RUMOR_STARTER

A player with this trait starts no rumours.