

Evolutionary dynamics of cooperation

Martin A. Nowak

Abstract. Cooperation means a donor pays a cost, c , for a recipient to get a benefit b . In evolutionary biology, cost and benefit are measured in terms of fitness. While mutation and selection represent the main ‘forces’ of evolutionary dynamics, cooperation is a fundamental principle that is required for every level of biological organization. Individual cells rely on cooperation among their components. Multi-cellular organisms exist because of cooperation among their cells. Social insects are masters of cooperation. Most aspects of human society are based on mechanisms that promote cooperation. Whenever evolution constructs something entirely new (such as multi-cellularity or human language), cooperation is needed. Evolutionary construction is based on cooperation. I will present five basic principles for the evolution of cooperation, which arise in the theories of kin selection, direct reciprocity, indirect reciprocity, graph selection and group selection.

Mathematics Subject Classification (2000). 92B05.

Keywords. Mathematical biology, evolutionary dynamics, kin selection, evolutionary graph theory, indirect reciprocity, Prisoner’s Dilemma.

1. Kin selection

In a pub conversation, J. B. S. Haldane, one of the founding fathers of a mathematical approach to biology, once remarked: ‘I will jump into the river to save two brothers or eight cousins.’ This insight was precisely formulated by William Hamilton many years later. He wrote a PhD thesis on this topic, submitted a long paper to the *Journal of Theoretical Biology*, disappeared into the Brazilian jungle and was world famous when he returned a few years later (Hamilton 1964a, b). The theory was termed ‘kin selection’ by John Maynard Smith (1964). The crucial equation is the following. Cooperation among relatives can be favored by natural selection if the coefficient of genetic relatedness, r , between the donor and the recipient exceeds the cost to benefit ratio of the altruistic act

$$r > c/b. \quad (1)$$

Kin selection theory has been tested in numerous experimental studies. Many cooperative acts among animals occur between close kin (Frank 1998, Hamilton 1998). The exact relationship between kin selection and other mechanisms such as group selection and spatial reciprocity, however, remains unclear. A recent study even suggests that much of cooperation in social insects is due to group selection rather than kin selection (Wilson & Hölldobler 2005).

2. Direct reciprocity

In 1971, Robert Trivers published a landmark paper entitled ‘The evolution of reciprocal altruism’ (Trivers 1971). Trivers analyzed the question of how natural selection could lead to cooperation between unrelated individuals. He discusses three biological examples: cleaning symbiosis in fish, warning calls in birds and human interactions. Trivers cites Luce & Raiffa (1957) and Rapoport & Chammah (1965) for the Prisoner’s Dilemma, which is a game where two players have the option to cooperate or to defect. If both cooperate they receive the ‘reward’, R . If both defect they receive the ‘punishment’, P . If one cooperates and the other defects, then the cooperator receives the ‘sucker’s payoff’, S , while the defector receives the ‘temptation’, T . The PD is defined by the ranking $T > R > P > S$.

Would you cooperate or defect? Assuming the other person will cooperate it is better to defect, because $T > R$. Assuming the other person will defect it is also better to defect, because $P > S$. Hence, no matter what the other person will do it is best to defect. If both players analyze the game in this ‘rational’ way then they will end up defecting. The dilemma is that they both could have received a higher payoff if they had chosen to cooperate. But cooperation is ‘irrational’.

We can also imagine a population of cooperators and defectors and assume that the payoff for each player is determined by many random interactions with others. Let x denote the frequency of cooperators and $1-x$ the frequency of defectors. The expected payoff for a cooperator is $f_C = Rx + S(1-x)$. The expected payoff for a defector is $f_D = Tx + P(1-x)$. Therefore, for any x , defectors have a higher payoff than cooperators. In evolutionary game theory, payoff is interpreted as fitness. Successful strategies reproduce faster and outcompete less successful ones. Reproduction can be cultural or genetic. In the non-repeated PD, in a well mixed population, defectors will outcompete cooperators. Natural selection favors defectors.

Cooperation becomes an option if the game is repeated. Suppose there are m rounds. Let us compare two strategies, ‘always defect’ (ALLD), and GRIM, which cooperates on the first move, then cooperates as long as the opponent cooperates, but permanently switches to defection if the opponent defects once. The expected payoff for GRIM versus GRIM is nR . The expected payoff for ALLD versus GRIM is $T + (m-1)P$. If $nR > T + (m-1)P$ then ALLD cannot spread in a GRIM population when rare. This is an argument of evolutionary stability. Interestingly, Trivers (1971) quotes ‘Hamilton (pers. commun.)’ for this idea.

A small problem of the above analysis is that given a known number of rounds it is best to defect in the last round and by backwards induction it is also best to defect in the penultimate round and so on. Therefore, it is more natural to consider a repeated game with a probability w of having another round. In this case, the expected number of rounds is $1/(1-w)$, and GRIM is stable against invasion by ALLD provided $w > (T-R)/(T-P)$.

We can also formulate the PD as follows. The cooperator helps at a cost, c , and the other individual receives a benefit b . Defectors do not help. Therefore we have

$T = b$, $R = b - c$, $P = 0$ and $S = -c$. The family of games that is described by the parameters b and c is a subset of all possible Prisoner's Dilemma games as long as $b > c$. For the repeated PD, we find that ALLD cannot invade GRIM if

$$w > c/b. \quad (2)$$

The probability of having another round must exceed the cost to benefit ratio of the altruistic act (Axelrod & Hamilton 1981, Axelrod 1984).

Thus, the repeated PD allows cooperation, but the question arises – what is a good strategy for playing this game? This question was posed by the political scientist, Robert Axelrod. In 1979, he decided to conduct a tournament of computer programs playing the repeated PD. He received 14 entries, from which the surprise winner was tit-for-tat (TFT), the simplest of all strategies that were submitted. TFT cooperates in the first move, and then does whatever the opponent did in the previous round. TFT cooperates if you cooperate, TFT defects if you defect. It was submitted by the game theorist Anatol Rapoport (who is also the co-author of the book Rapoport & Chamah, 1965). Axelrod analyzed the events of the tournament, published a detailed account and invited people to submit strategies for a second championship. This time he received 63 entries. John Maynard Smith submitted 'tit-for-two-tats', a variant of TFT which defects only after the opponent has defected twice in a row. Only one person, Rapoport, submitted TFT, and it won again. At this time, TFT was considered to be the undisputed world champion in the heroic world of the repeated PD.

But one weakness became apparent very soon (Molander 1985, May 1987). TFT cannot correct mistakes. The tournaments were conducted without strategic noise. In the real world, 'trembling hands' and 'fuzzy minds' cause erroneous moves. If two TFT players interact with each other, a single mistake leads to a long sequence of alternating defection and cooperation. In the long run two TFT players get the same low payoff as two players who flip coins for every move in order to decide whether to cooperate or to defect. Errors destroy TFT.

In 1989, we began to conduct 'evolutionary tournaments' (Nowak & Sigmund 1992). Instead of inviting experts to submit programs, we asked mutation and selection to explore (some portion of) the strategy space of the repeated PD in the presence of noise. The initial random ensemble of strategies was quickly dominated by ALLD. If the opposition is nonsensical, it is best to defect. A large portion of the population began to adopt the ALLD strategy and everything seemed lost. But after some time, a small cluster of players adopted a strategy very close to TFT. If this cluster is sufficiently large, then it can increase in abundance, and the entire population swings from ALLD to TFT. Reciprocity (and therefore cooperation) has emerged. We can show that TFT is the best catalyst for the emergence of cooperation. But TFT's moment of glory was brief and fleeting. In all cases, TFT was rapidly replaced by another strategy. On close inspection, this strategy turned out to be 'generous-tit-for-tat' (GTFT) which always cooperates if the opponent has cooperated on the previous move, but sometimes (probabilistically) even cooperates when the opponent has defected. Natural selection had discovered 'forgiveness'.

After many generations, however, GTFT is undermined by unconditional cooperators, ALLC. In a society, where everybody is nice (using GTFT), there is almost no need to remember how to retaliate against a defection. A biological trait which is not used is likely to be lost by random drift. Birds that escape to islands without predators lose the ability to fly. Similarly, a GTFT population is softened and turns into an ALLC population.

Once most people play ALLC, there is an open invitation for ALLD to seize power. This is precisely what happens. The evolutionary dynamics run in cycles: from ALLD to TFT to GTFT to ALLC and back to ALLD. These oscillations of cooperative and defecting societies are a fundamental part of all our observations regarding the evolution of cooperation. Most models of cooperation show such oscillations. Cooperation is never a final state of evolutionary dynamics. Instead it is always lost to defection after some time and has to be re-established. These oscillations are also reminiscent of alternating episodes of war and peace in human history (Figure 1).

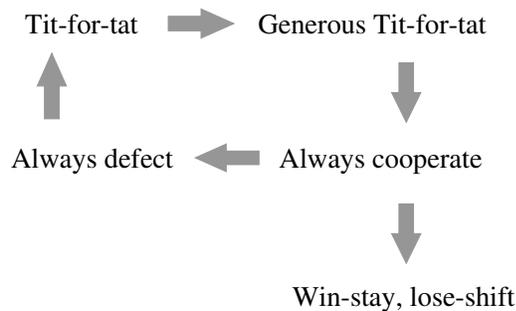


Figure 1. Evolutionary cycles of cooperation and defection. A small cluster of tit-for-tat (TFT) players or even a lineage starting from a single TFT player in a finite population can invade 'always defect' (ALLD). In fact, TFT is the most efficient catalyst for the first emergence of cooperation in an ALLD population. But in a world of 'fuzzy minds' and 'trembling hands', TFT is soon replaced by generous-tit-for-tat (GTFT), which can re-establish cooperation after occasional mistakes. If everybody uses GTFT, then 'always cooperate' (ALLC) is a neutral variant. Random drift leads to ALLC. An ALLC population invites invasion by ALLD. But ALLC is also dominated by 'win-stay, lose-shift' (WSLS), which leads to more stable cooperation than tit-for-tat-like strategies.

A subsequent set of simulations, exploring a larger strategy space, led to a surprise (Nowak & Sigmund 1993). The fundamental oscillations were interrupted by another strategy which seems to be able to hold its ground for a very long period of time. Most surprisingly, this strategy is based on the extremely simple principle of win-stay, lose-shift (WSLS). If my payoff is R or T then I will continue with the same move next round. If I have cooperated then I will cooperate again, if I have defected then I will defect again. If my payoff is only S or P then I will switch to the other move next round. If I have cooperated then I will defect, if I have defected then I will cooperate (Figure 2). If two WSLS strategists play each other, they cooperate

Win-stay			
C	(3)	C	
C		D	(5)
C		D	D
Lose-shift			
C	(0)	D	
D		D	(1)
D		C	(probabilistic)
D		D	

Figure 2. ‘Win-stay, lose-shift’ (WSLS) embodies a very simple principle. If you do well then continue with what you are doing. If you are not doing well, then try something else. Here we consider the Prisoner’s Dilemma (PD) payoff values $R = 3$, $T = 5$, $P = 1$ and $S = 0$. If both players cooperate, you receive 3 points, and you continue to cooperate. If you defect against a cooperater, you receive 5 points, and you continue to defect. But if you cooperate with a defector, you receive 0 points, and therefore you will switch from cooperation to defection. If, on the other hand, you defect against a defector, you receive 1 point, and you will switch to cooperation. Your aspiration level is 3 points. If you get at least 3 points then you consider it a ‘win’ and you will ‘stay’ with your current choice. If you get less than 3 points, you consider it a ‘loss’ and you will ‘shift’ to another move. If $R > (T + P)/2$ (or $b/c > 2$) then WSLS is stable against invasion by ALLD. If this inequality does not hold, then our evolutionary simulations lead to a stochastic variant of WSLS, which cooperates after a DD move only with a certain probability. This stochastic variant of WSLS is then stable against invasion by ALLD.

most of the time. If a defection occurs accidentally, then in the next move both will defect. Hereafter both will cooperate again. WSLS is a simple deterministic machine to correct stochastic noise. While TFT cannot correct mistakes, both GTFT and WSLS can correct mistakes. But WSLS has an additional ace in its hand. When WSLS plays ALLC it will discover after some time that ALLC does not retaliate. After an accidental defection, WSLS will switch to permanent defection. Therefore, a population of WSLS players does not drift to ALLC. Cooperation based on WSLS is more stable than cooperation based on tit-for-tat-like strategies. The repeated PD is mostly known as a story of tit-for-tat, but win-stay, lose-shift is a superior strategy in an evolutionary scenario with errors, mutation and many generations (Fudenberg & Maskin 1990, Nowak & Sigmund 1993).

Incidentally, WSLS is stable against invasion by ALLD if $b/c > 2$. If instead $1 < b/c < 2$ then a stochastic variant of WSLS dominates the scene; this strategy cooperates after a mutual defection only with a certain probability. Of course, all strategies of direct reciprocity, such as TFT, GTFT or WSLS can only lead to the evolution of cooperation if the fundamental inequality (2) is fulfilled.

3. Indirect reciprocity

While direct reciprocity embodies the idea ‘You scratch my back and I scratch yours’, indirect reciprocity suggests ‘You scratch my back and I scratch someone else’s’. Why should this work? Presumably I will not get scratched if it becomes known that I scratch nobody. Indirect reciprocity, in this view, is based on reputation (Nowak & Sigmund 1998a, b, 2005). But why should you care about what I do to a third person?

The main reason why economists and social scientists are interested in indirect reciprocity is because one-shot interactions between anonymous partners in a global market become increasingly common and tend to replace the traditional long-lasting associations and long-term interactions between relatives, neighbors, or members of the same village. A substantial part of our life is spent in the ‘company of strangers’, and many transactions are no longer face-to-face. The growth of e-auctions and other forms of e-commerce is based, to a considerable degree, on reputation and trust. The potential to exploit such trust raises what economists call moral hazards. How effective is reputation, especially if information is only partial?

Evolutionary biologists, on the other hand, are interested in the emergence of human societies, which constitutes the last (up to now) of the major transitions in evolution. In contrast to other eusocial species, such as bees, ants or termites, humans display a high degree of cooperation between non-relatives (Fehr & Fischbacher 2003). A considerable part of human cooperation is based on moralistic emotions, such as anger directed towards cheaters or the ‘warm inner glow’ felt after performing an altruistic action. Intriguingly, humans not only feel strongly about interactions which involve them directly, they also judge actions between third parties as evidenced by the contents of gossip. There are numerous experimental studies of indirect reciprocity based on reputation (Wedekind & Milinski 2000, Milinski et al. 2002, Wedekind & Braithwaite 2002, Seinen & Schramm 2006).

A simple model of indirect reciprocity (Nowak & Sigmund 1998a, b) assumes that, within a well-mixed population, individuals meet randomly, one in the role of the potential donor, the other as the potential recipient. Each individual experiences several rounds of this interaction in each role, but never with the same partner twice. A player can follow either an unconditional strategy, such as always cooperate or always defect, or a conditional strategy, which discriminates among the potential recipients according to their past interactions. In a simple example, a discriminating donor helps a recipient if her score exceeds a certain threshold. A player’s score is 0 at birth, increases whenever that player helps and decreases whenever the player withholds help. Individual-based simulations and direct calculations show that cooperation based on indirect reciprocity can evolve provided the probability, p , of knowing the social score of another person exceeds the cost-to-benefit ratio of the altruistic act,

$$p > c/b. \quad (3)$$

The role of genetic relatedness that is crucial for kin selection is replaced by social acquaintanceship. In a fluid population, where most interactions are anonymous and

people have no possibility of monitoring the social score of others, indirect reciprocity has no chance. But in a socially viscous population, where people know one another’s reputation, cooperation by indirect reciprocity can thrive (Nowak & Sigmund 1998a).

In a world of binary moral judgments (Nowak & Sigmund 1998b, Leimar & Hammerstein 2001, Panchanathan & Boyd 2003, Fishman 2003, Brandt & Sigmund 2004, 2005), there are four ways of assessing donors in terms of ‘first-order assessment’: always consider them as good, always consider them as bad, consider them as good if they refuse to give, or consider them as good if they give. Only this last option makes sense. Second-order assessment also depends on the score of the receiver; for example, it can be deemed good to refuse help to a bad person. There are 16 second-order rules. Third-order assessment also depends on the score of the donor; for example, a good person refusing to help a bad person may remain good, but a bad person refusing to help a bad person remains bad. There are 256 third-order assessment rules. We display three of them in Figure 3. Using the Scoring assessment rule,

Three assessment rules

Reputation of donor and recipient

		GG	GB	BG	BB	
Action of donor	C	G	G	G	G	Scoring
	D	B	B	B	B	
	C	G	G	G	G	Standing
	D	B	G	B	B	
	C	G	B	G	B	Judging
	D	B	G	B	B	

Reputation of donor
after the action

Figure 3. Assessment rules specify how an observer judges an interaction between a potential donor and a recipient. Here we show three examples of assessment rules in a world of binary reputation, good (G) and bad (B). For ‘Scoring’, cooperation (C) earns a good reputation and defection (D) earns a bad reputation. ‘Standing’ is very similar to Scoring, the only difference is that a ‘good’ donor can defect against a ‘bad’ recipient without losing his ‘good’ reputation. Note that Scoring is associated with costly punishment (Fehr & Gaechter 2002, Sigmund et al. 2001), whereas for Standing punishment of ‘bad’ recipients is cost-free. For ‘Judging’ it is ‘bad’ to help a ‘bad’ recipient.

cooperation, C, always leads to a good reputation, G, whereas defection, D, always leads to a bad reputation, B. Standing (Sugden 1986) is like Scoring, but it is not bad if a good donor defects against a bad recipient . With Judging, in addition, it is bad to cooperate with a bad recipient.

An action rule for indirect reciprocity prescribes giving or not giving, depending on the scores of both donor and recipient. For example, you may decide to help if the recipient’s score is good or your own score is bad. Such an action might increase your

own score and therefore increase the chance of receiving help in the future. There are 16 action rules.

If we view a strategy as the combination of an action rule and an assessment rule, we obtain 4096 strategies. In a remarkable calculation, Ohtsuki & Iwasa (2004, 2005) analyzed all 4096 strategies and proved that only eight of them are evolutionarily stable under certain conditions and lead to cooperation (Figure 4). Both Standing

Ohtsuki & Iwasa's 'Leading eight'

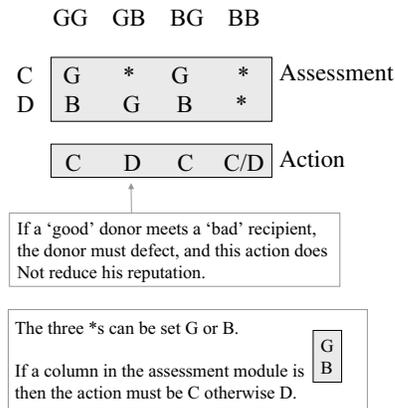


Figure 4. Ohtsuki & Iwasa (2004, 2005) analyzed the combination of $2^8 = 256$ assessment modules with $2^4 = 16$ action modules. This is a total of 4096 strategies. They found that 8 of these strategies can be evolutionarily stable and lead to cooperation, provided that everybody agrees on each other's reputation. (In general, uncertainty and incomplete information might lead to private lists of the reputation of others.) The three asterisks in the assessment module indicate a free choice between G and B. There are therefore $2^3 = 8$ different assessment rules which make up the 'leading eight'. The action module is built as follows: if the column in the assessment module is G and B, then the corresponding action is C, otherwise the action is D. Note that Standing and Judging are members of the leading eight, but neither Scoring nor Shunning is included.

and Judging belong to the leading eight, but not Scoring. We expect, however, that Scoring has a similar role in indirect reciprocity to that of tit-for-tat in direct reciprocity. Neither strategy is evolutionarily stable, but their ability to catalyze cooperation in adverse situations and their simplicity constitute their strength. In extended versions of indirect reciprocity in which donors can sometimes deceive others about the reputation of the recipient, Scoring is the 'foolproof' concept of 'I believe what I see'. Scoring judges the action and ignores the stories. There is also experimental evidence that in certain situations humans follow scoring rather than standing (Milinski et al. 2001).

In human evolution, there must have been a tendency to move from the simple cooperation promoted by kin or group selection to the strategic subtleties of direct and indirect reciprocity. Direct reciprocity requires precise recognition of individual

people, a memory of the various interactions one had with them in the past, and enough brain power to conduct multiple repeated games simultaneously. Indirect reciprocity, in addition, requires the individual to monitor interactions among other people, possibly judge the intentions that occur in such interactions, and keep up with the ever changing social network of the group. Reputation of players may not only be determined by their own actions, but also by their associations with others.

We expect that indirect reciprocity has coevolved with human language. On one hand, it is helpful to have names for other people and to receive information about how a person is perceived by others. On the other hand, a complex language is especially necessary if there are intricate social interactions. The possibilities for games of manipulation, deceit, cooperation and defection are limitless. It is likely that indirect reciprocity has provided the very selective scenario that led to cerebral expansion in human evolution.

4. Graph selection (or network reciprocity)

Game theory was invented by von Neumann and Morgenstern (1944) as a mathematical approach to understanding the strategic and economic decisions of humans. Hamilton (1967), Trivers (1971) and Maynard Smith & Price (1973) brought game theory to biology. Instead of analyzing the interaction between two rational players, evolutionary game theory explores the dynamics of a population of players under the influence of natural selection (Maynard Smith 1982). In the classical setting of the replicator equation, the population size is infinite and interactions are equally likely between any two individuals (Taylor & Jonker 1978, Hofbauer et al. 1979, Zeeman 1980). Each individual obtains an average payoff which is interpreted as biological fitness: strategies reproduce according to their payoff. Successful strategies spread and eliminate less successful ones. The payoff depends on the frequency of strategies in the population. Hence, natural selection is frequency dependent. The replicator equation is deeply connected to the concept of an evolutionarily stable strategy (ESS) or Nash equilibrium. In the framework of the replicator equation, an ESS cannot be invaded by any mutant strategy (Hofbauer & Sigmund 1998). For recent books on game theory and evolutionary game theory we refer to Fudenberg & Tirole 1991, Binmore 1994, Weibull 1995, Samuelson 1997, Fudenberg & Levine 1998, Hofbauer & Sigmund 1998, Gintis 2000, Cressman 2003. Recent reviews of evolutionary game dynamics are Hofbauer & Sigmund (2003) and Nowak & Sigmund (2004).

The traditional model of evolutionary game dynamics assumes that populations are well-mixed, which means that interactions between any two players are equally likely. More realistically, however, the interactions between individuals are governed by spatial effects or social networks. Let us therefore assume that the individuals of a population occupy the vertices of a graph (Nakamaru et al. 1997, 1998, Skyrms & Pemantle 2000, Abramson & Kuperman 2001, Ebel & Bornholdt 2002, Lieberman et al. 2005, Nakamaru & Iwasa 2005, Santos et al. 2005, Santos & Pacheco 2005).

The edges of the graph determine who interacts with whom (Figure 5). Consider a population of N individuals consisting of cooperators and defectors. A cooperator

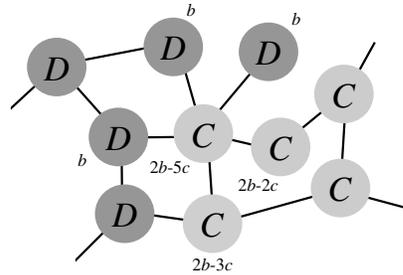


Figure 5. Games on graphs. The members of a population occupy the vertices of a graph (or social network). The edges denote who interacts with whom. Here we consider the specific example of cooperators, C , competing with defectors, D . A cooperator pays a cost, c , for every link. Each neighbor of a cooperator receives a benefit, b . The payoffs of some individuals are indicated in the figure. The fitness of each individual is a constant, denoting the ‘baseline fitness’, plus the payoff of the game. For evolutionary dynamics, we assume that in each round a random player is chosen to die, and the neighbors compete for the empty site in a mode proportional to their fitness. A simple rule emerges: if $b/c > k$ then selection favors cooperators over defectors. Here k is the average number of neighbors per individual.

helps all individuals to whom it is connected. If a cooperator is connected to k other individuals and i of those are cooperators, then its payoff is $bi - ck$. A defector does not provide any help, and therefore has no costs, but it can receive the benefit from neighboring cooperators. If a defector is connected to k other individuals and j of those are cooperators, then its payoff is bj . Evolutionary dynamics are described by an extremely simple stochastic process: at each time step, a random individual adopts the strategy of one of its neighbors proportional to their fitness.

We note that stochastic evolutionary game dynamics in finite populations is sensitive to the intensity of selection. In general, the reproductive success (fitness) of an individual is given by a constant, denoting the baseline fitness, plus the payoff that arises from the game under consideration. Strong selection means that the payoff is large compared to the baseline fitness; weak selection means the payoff is small compared to the baseline fitness. It turns out that many interesting results can be proven for weak selection, which is an observation also well known in population genetics.

The traditional, well-mixed population of evolutionary game theory is represented by the complete graph, where all vertices are connected, which means that all individuals interact equally often. In this special situation, cooperators are always opposed by natural selection. This is the fundamental intuition of classical evolutionary game theory. But what happens on other graphs?

We need to calculate the probability, ρ_C , that a single cooperator, starting in a random position, turns the whole population from defectors into cooperators. If selection neither favors nor opposes cooperation, then this probability is $1/N$, which

is the fixation probability of a neutral mutant. If the fixation probability ρ_C is greater than $1/N$, then selection favors the emergence of cooperation. Similarly, we can calculate the fixation probability of defectors, ρ_D .

A surprisingly simple rule determines whether selection on graphs favors cooperation. If

$$b/c > k, \quad (4)$$

then cooperators have a fixation probability greater than $1/N$ and defectors have a fixation probability less than $1/N$. Thus, for graph selection to favor cooperation, the benefit-to-cost ratio of the altruistic act must exceed the average degree, k , which is given by the average number of links per individual. This relationship can be shown with the method of pair-approximation for regular graphs, where all individuals have exactly the same number of neighbors (Ohtsuki et al. 2006). Regular graphs include cycles, all kinds of spatial lattices and random regular graphs. Moreover, computer simulations suggest that the rule $b/c > k$ also holds for non-regular graphs such as random graphs and scale free networks. The rule holds in the limit of weak selection and $k \ll N$. For the complete graph, $k = N$, we always have $\rho_D > 1/N > \rho_C$.

The basic idea is that natural selection on graphs (in structured populations) can favor unconditional cooperation without any need of strategic complexity, reputation or kin selection.

Games on graphs grew out of the earlier tradition of spatial evolutionary game theory (Nowak & May 1992, Herz 1994, Killingback & Doebeli 1996, Mitteldorf & Wilson 2000, Hauert et al. 2002, Le Galliard et al 2003, Hauert & Doebeli 2004, Szabo & Vukov 2004) and investigations of spatial models in ecology (Durrett & Levin 1994a, b, Hassell et al. 1994, Tilman & Kareiva 1997, Neuhauser 2001) and spatial models in population genetics (Wright 1931, Fisher & Ford 1950, Maruyama 1970, Slatkin 1981, Barton 1993, Pulliam 1988, Whitlock 2003).

5. Group selection

The enthusiastic approach of early group selectionists to explain the evolution of cooperation entirely from this one perspective (Wynne-Edwards 1962) has met with vigorous criticism (Williams 1966) and has led to a denial of group selection for decades. Only an embattled minority of scientists defended the approach (Eshel 1972, Wilson 1975, Matessi & Jayakar 1976, Wade 1976, Uyenoyama & Feldman 1980, Slatkin 1981, Leigh 1983, Szathmary & Demeter 1987). Nowadays, however, it seems clear that group selection acts as a powerful mechanism for the promotion of cooperation (Sober & Wilson 1998, Keller 1999, Michod 1999, Swenson et al. 2000, Kerr & Godfrey-Smith 2002, Paulsson 2002, Boyd & Richerson 2002, Bowles & Gintis 2004, Traulsen et al. 2005). We only have to make sure that its basic requirements are fulfilled in a particular situation (Maynard Smith 1976). We would like to illustrate exactly what these requirements are through the use of a simple model (Traulsen & Nowak 2006).

Imagine a population of individuals subdivided into groups. For simplicity, we assume the number of groups is constant and given by m . Each group contains between one and n individuals. The total population size can fluctuate between the bounds m and nm . Again, there are two types of individuals, cooperators and defectors. Individuals interact with others in their group and thereby receive a payoff. At each time step a random individual from the entire population is chosen proportional to payoff in order to reproduce. The offspring is added to the same group. If the group size is less than or equal to n nothing else happens. If the group size, however, exceeds n then with probability q the group splits into two. In this case, a random group is eliminated (in order to maintain a constant number of groups). With probability $1 - q$, the group does not divide, but instead a random individual from that group is eliminated (Figure 6). This minimalist model of multi-level selection has some interesting fea-

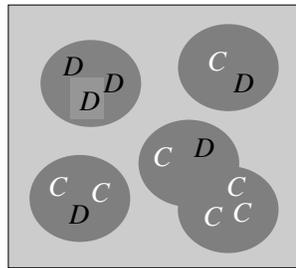


Figure 6. A simple model of group selection. A population consists of m groups of maximum size n . Individuals interact with others in their group in the context of an evolutionary game. Here we consider the game between cooperators, C , and defectors, D . For reproduction, individuals are chosen from the entire population with a probability proportional to their payoff. The offspring is added to the same group. If a group reaches the maximum size, n , then it either splits in two or a random individual from that group is eliminated. If a group splits, then a random group dies, in order to keep the total population size constant. This meta-population structure leads to the emergence of two levels of selection, although only individuals reproduce.

tures. Note that the evolutionary dynamics are entirely driven by individual fitness. Only individuals are assigned payoff values. Only individuals reproduce. Groups can stay together or split (divide) when reaching a certain size. Groups that contain fitter individuals reach the critical size faster and therefore split more often. This concept leads to selection among groups, even though only individuals reproduce. The higher level selection emerges from lower level reproduction. Remarkably, the two levels of selection can oppose each other.

As before, we can compute the fixation probabilities, ρ_C and ρ_D , of cooperators and defectors in order to check whether selection favors one or the other. If we add a single cooperator to a population of defectors, then this cooperator must first take over a group. Subsequently the group of cooperators must take over the entire population. The first step is opposed by selection, the second step is favored by selection. Hence,

we need to find out if the overall fixation probability is greater to or less than what we would obtain for a neutral mutant. An analytic calculation is possible in the limit $q \ll 1$ where individuals reproduce much more rapidly than groups divide. In this case, most of the groups are at their maximum size and hence the total population size is almost constant and given by $N = nm$. We find that selection favors cooperators and opposes defectors, $\rho_C > 1/N > \rho_D$, if

$$\frac{b}{c} > 1 + \frac{n}{m-2}. \quad (5a)$$

This result holds for weak selection. Smaller group sizes and larger numbers of competing groups favor cooperation. We also notice that the number of groups, m , must exceed two. There is an intuitive reason for this threshold. Consider the case of $m = 2$ groups with $n = 2$ individuals. In a mixed group, the cooperator has payoff $-c$ and the defector has payoff b . In a homogeneous group, two cooperators have payoff $b - c$, while two defectors have payoff 0. Thus the disadvantage for cooperators in mixed groups cannot be compensated for by the advantage they have in homogeneous groups. Interestingly, however, for larger splitting probabilities, q , we find that cooperators can be favored even for $m = 2$ groups. The reason is the following: for very small q , the initial cooperator must reach fixation in a mixed group; but for larger q , a homogeneous cooperator group can also emerge if a mixed group splits giving rise to a daughter group that has only cooperators. Thus, larger splitting probabilities make it easier for cooperation to emerge.

Let us also consider the effect of migration between groups. The average number of migrants accepted by a group during its life-time is denoted by z . We find that selection favors cooperation provided

$$\frac{b}{c} > 1 + z + \frac{n}{m}. \quad (5b)$$

In order, to derive this condition we have assumed weak selection and $q \ll 1$, as before, but also that both the numbers of groups, m , and the maximum group size, n , are much large than one.

Group selection (or multi-level selection) is a powerful mechanism for the evolution of cooperation if there are a large number of relatively small groups and migration between groups is not too frequent.

6. Conclusion

I have presented five simple (Equations 1–5) rules that determine whether particular mechanisms can promote the evolution of cooperation. In all five theories, b is the benefit for the recipient and c the cost for the donor of an altruistic act. The comparison of the five rules enables us to understand the crucial quantities that are responsible

for the natural selection of cooperation by the various mechanisms that have been proposed.

1. Kin selection leads to cooperation if $b/c > 1/r$, where r is the coefficient of genetic relatedness between donor and recipient (Hamilton 1964a).

2. Direct reciprocity leads to cooperation if $b/c > 1/w$, where w is the probability of playing another round in the repeated Prisoner's Dilemma (Axelrod & Hamilton 1981).

3. Indirect reciprocity leads to cooperation if $b/c > 1/q$, where q is the probability to know the reputation of a recipient (Nowak & Sigmund 1998a).

4. Graph selection (or 'network reciprocity') leads to cooperation if $b/c > k$, where k is the degree of the graph, that is the average number of neighbors (Ohtsuki et al. 2006).

5. Group selection leads to cooperation if $b/c > 1 + z + n/m$, where z is the number of migrants accepted by a group during its life-time, n is the group size and m is the number of groups (Traulsen & Nowak 2006).

References

- [1] Abramson, G., and M. Kuperman. 2001. Social games in a social network. *Phys. Rev. E* **63** (3), 030901R.
- [2] Axelrod, R. M. 1984. *The evolution of cooperation*. Basic Books, New York; reprint, Penguin, Harmondsworth 1989.
- [3] Axelrod, R., and W. D. Hamilton. 1981. The evolution of cooperation. *Science* **211**, 1390–1396.
- [4] Barton, N. 1993. The probability of fixation of a favoured allele in a subdivided population. *Genet. Res.* **62**, 149–158.
- [5] Binmore, K. 1994. *Game theory and the social contract*, Vol. 1. *Playing fair*. MIT Press, Cambridge.
- [6] Bowles, S., and H. Gintis. 2004. The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoret. Population Biol.* **65**, 17–28.
- [7] Boyd, R., and P. J. Richerson. 2002. Group beneficial norms can spread rapidly in a structured population. *J. Theoret. Biol.* **215**, 287–296.
- [8] Brandt, H., and K. Sigmund. 2004. The logic of reprobation: Assessment and action rules for indirect reciprocity. *J. Theoret. Biol.* **231**, 475–486.
- [9] Brandt, H., and K. Sigmund. 2005. Indirect reciprocity, image scoring, and moral hazard. *Proc. Natl. Acad. Sci. USA* **102**, 2666–2670.
- [10] Cressman, R. 2003. *Evolutionary dynamics and extensive form games*. MIT Press, Cambridge.
- [11] Durrett, R., and S. A. Levin. 1994a. The importance of being discrete (and spatial). *Theoret. Population Biol.* **46**, 363–394.
- [12] Durrett, R., and S. A. Levin. 1994b. Stochastic spatial models: A user's guide to ecological applications. *Philos. Trans. Roy. Soc. B* **343**, 329–350.

- [13] Ebel, H., and S. Bornholdt. 2002. Coevolutionary games on networks. *Phys. Rev. E* **66** (5), 056118.
- [14] Eshel, I. 1972. Neighbor effect and the evolution of altruistic traits. *Theoret. Population Biol.* **3**, 258–277.
- [15] Fehr, E., and U. Fischbacher. 2003. The nature of human altruism. *Nature* **425**, 785–791.
- [16] Fisher, R. A., and E. B. Ford. 1950. The Sewall Wright effect. *Heredity* **4**, 117–119.
- [17] Fishman, M. A. 2003. Indirect reciprocity among imperfect individuals. *J. Theoret. Biol.* **225**, 285–292.
- [18] Frank, S. A. 1998. *Foundations of social evolution*. Princeton University Press, Princeton, NJ.
- [19] Fudenberg, D., and D. K. Levine. 1998. *The Theory of Learning in Games*. MIT Press, Cambridge.
- [20] Fudenberg, D., and E. Maskin. 1990. Evolution and cooperation in noisy repeated games. *Amer. Econ. Rev.* **80**, 274–279.
- [21] Fudenberg, D., and J. Tirole. 1991. *Game theory*. MIT Press, Cambridge.
- [22] Gintis, H. 2000. *Game theory evolving*. Princeton University Press, Princeton, NJ.
- [23] Hamilton, W. D. 1964a. The genetical evolution of social behaviour I. *J. Theoret. Biol.* **7**, 1–16.
- [24] Hamilton, W. D. 1964b. The genetical evolution of social behaviour II. *J. Theoret. Biol.* **7**, 17–52.
- [25] Hamilton, W. D. 1967. Extraordinary sex ratios. *Science* **156**, 477–488.
- [26] Hamilton, W. D. 1998. *Narrow roads of gene land: The collected papers of W. D. Hamilton Volume 1: Evolution of social behaviour*. Oxford University Press, New York.
- [27] Hassell, M. P., H. N. Comins, and R. M. May. 1994. Species coexistence and self-organizing spatial dynamics. *Nature* **370**, 290–292.
- [28] Hauert, C., S. De Monte, J. Hofbauer, and K. Sigmund. 2002. Volunteering as red queen mechanism for cooperation in public goods games. *Science* **296**, 1129–1132.
- [29] Hauert, C., and M. Doebeli. 2004. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* **428**, 643–646.
- [30] Herz, A. V. M. 1994. Collective phenomena in spatially extended evolutionary games. *J. Theoret. Biol.* **169**, 65–87.
- [31] Hofbauer, J., P. Schuster, and K. Sigmund. 1979. A note on evolutionarily stable strategies and game dynamics. *J. Theoret. Biol.* **81**, 609–612.
- [32] Hofbauer, J., and K. Sigmund. 1998. *Evolutionary games and population dynamics*. Cambridge University Press, Cambridge.
- [33] Hofbauer, J., and K. Sigmund. 2003. Evolutionary game dynamics. *Bull. Amer. Math. Soc.* **40**, 479–519.
- [34] Keller, L., ed. 1999. *Levels of selection in evolution*. Princeton University Press, Princeton, NJ.
- [35] Kerr, B., and P. Godfrey-Smith. 2002. Individualist and multi-level perspectives on selection in structured populations. *Biol. Philos.* **17**, 477–517.

- [36] Killingback, T., and M. Doebeli. 1996. Spatial evolutionary game theory: Hawks and Doves revisited. *Proc. Royal Soc. London B* **263**, 1135–1144.
- [37] Le Galliard, J.-F., R. Ferrière, and U. Dieckmann. 2003. The adaptive dynamics of altruism in spatially heterogeneous populations. *Evolution* **57**, 1–17.
- [38] Leigh, E. G. 1983. When does the good of the group override the advantage of the individual? *Proc. Natl. Acad. Sci. USA* **80**, 2985–2989.
- [39] Leimar, O., and P. Hammerstein. 2001. Evolution of cooperation through indirect reciprocity. *Proc. Royal Soc. London B* **268**, 745–753.
- [40] Lieberman, E., C. Hauert, and M. A. Nowak. 2005. Evolutionary dynamics on graphs. *Nature* **433**, 312–316.
- [41] Luce, R. D., and H. Raiffa. 1957. *Games and Decisions*. John Wiley, New York, NY.
- [42] Maruyama, T. 1970. Effective number of alleles in a subdivided population. *Theoret. Population Biol.* **1**, 273–306.
- [43] Matessi, C., and S. D. Jayakar. 1976. Conditions for the evolution of altruism under Darwinian selection. *Theoret. Population Biol.* **9**, 360–387.
- [44] May, R. M. 1987. More evolution of cooperation. *Nature* **327**, 15–17.
- [45] Maynard Smith, J. 1964. Group selection and kin selection. *Nature* **63**, 20–29.
- [46] Maynard Smith, J. 1976. Group selection. *Quart. Rev. Biol.* **201**, 145–147.
- [47] Maynard Smith, J. 1982. *Evolution and the theory of games*. Cambridge University Press, Cambridge.
- [48] Maynard Smith, J., and G. R. Price. 1973. The logic of animal conflict. *Nature* **246**, 15–18.
- [49] Michod, R. E. 1999. *Darwinian dynamics: Evolutionary transitions in fitness and individuality*. Princeton University Press, Princeton, NJ.
- [50] Milinski, M., D. Semmann, T. C. M. Bakker, and H.-J. Krambeck. 2001. Cooperation through indirect reciprocity: Image scoring or standing strategy? *Proc. Royal Soc. London B* **268**, 2495–2501.
- [51] Milinski, M., D. Semmann, and H.-J. Krambeck. 2002. Reputation helps solve the ‘tragedy of the commons’. *Nature* **415**, 424–426.
- [52] Mitteldorf, J., and D. S. Wilson. 2000. Population viscosity and the evolution of altruism. *J. Theoret. Biol.* **204**, 481–496.
- [53] Molander, P. 1985. The optimal level of generosity in a selfish, uncertain environment. *J. Conflict Resolut.* **29**, 611–618.
- [54] Nakamaru, M., and Y. Iwasa. 2005. The evolution of altruism by costly punishment in lattice-structured populations: Score-dependent viability versus score-dependent fertility. *Evol. Ecol. Res.*, in press.
- [55] Nakamaru, M., H. Matsuda, and Y. Iwasa. 1997. The evolution of cooperation in a lattice-structured population. *J. Theoret. Biol.* **184**, 65–81.
- [56] Nakamaru, M., H. Nogami, and Y. Iwasa. 1998. Score-dependent fertility model for the evolution of cooperation in a lattice. *J. Theoret. Biol.* **194**, 101–124.
- [57] Neuhauser, C. 2001. Mathematical challenges in spatial ecology. *Notices Amer. Math. Soc.* **48**, 1304–1314.
- [58] Nowak, M. A., and R. M. May. 1992. Evolutionary games and spatial chaos. *Nature* **359**, 826–829.

- [59] Nowak, M. A., and K. Sigmund. 1992. Tit for tat in heterogeneous populations. *Nature* **355**, 250–253.
- [60] Nowak, M., and K. Sigmund. 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* **364**, 56–58.
- [61] Nowak, M. A., and K. Sigmund. 1998a. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577.
- [62] Nowak, M. A., and K. Sigmund. 1998b. The dynamics of indirect reciprocity. *J. Theoret. Biol.* **194**, 561–574.
- [63] Nowak M. A., and K. Sigmund. 2004. Evolutionary dynamics of biological games. *Science* **303**, 793–799.
- [64] Nowak, M. A., and K. Sigmund. 2005. Evolution of indirect reciprocity. *Nature*, in press.
- [65] Ohtsuki, H., C. Hauert, E. Lieberman, and M. A. Nowak. 2006. A simple rule for the evolution of cooperation on graphs. *Nature*, in press.
- [66] Ohtsuki, H., and Y. Iwasa. 2004. How should we define goodness? Reputation dynamics in indirect reciprocity. *J. Theoret. Biol.* **231**, 107–120.
- [67] Ohtsuki, H., and Y. Iwasa. 2005. The leading eight: Social norms that can maintain cooperation by indirect reciprocity *J. Theoret. Biol.*, in press.
- [68] Panchanathan, K., and R. Boyd. 2003. A tale of two defectors: The importance of standing for evolution of indirect reciprocity. *J. Theoret. Biol.* **224**, 115–126.
- [69] Paulsson, J. 2002. Multileveled selection on plasmid replication. *Genetics* **161**, 1373–1384.
- [70] Pulliam, H. R. 1988. Sources, sinks, and population regulation. *Amer. Nat.* **132**, 652–661.
- [71] Rapoport, A., and A. M. Chammah. 1965. *Prisoner's dilemma*. University of Michigan Press, Ann Arbor, MI.
- [72] Samuelson, L. 1997. *Evolutionary games and equilibrium selection*. MIT Press, Cambridge.
- [73] Santos, F. C., and J. M. Pacheco. 2005. Scale-free networks provide a unifying framework for the emergence of cooperation. *Phys. Rev. Lett.* **95**, 098104.
- [74] Santos, F. C., J. F. Rodrigues, and J. M. Pacheco. 2005. Graph topology plays a determinant role in the evolution of cooperation. *Proc. Royal Soc. London B* **273**, 51–55.
- [75] Seinen, I., and A. Schram. 2006. Social status and group norms: Indirect reciprocity in a repeated helping experiment. *Eur. Econ. Rev.*, in press.
- [76] Skyrms, B., and R. Pemantle. 2000. A dynamic model of social network formation. *Proc. Natl. Acad. Sci. USA* **97**, 9340–9346.
- [77] Slatkin, M. 1981. Fixation probabilities and fixation times in a subdivided population. *Evolution* **35**, 477–488.
- [78] Slatkin, M., and M. J. Wade. 1978. Group selection on a quantitative character. *Proc. Natl. Acad. Sci. USA* **75**, 3531–3534.
- [79] Sober, E., and D. S. Wilson. 1998. *Unto others: The evolution and psychology of unselfish behavior*. Harvard University Press, Cambridge, MA.
- [80] Sugden, R. 1986. *The economics of rights, co-operation and welfare*. Blackwell, Oxford.
- [81] Swenson, W., D. S. Wilson, and R. Elias. 2000. Artificial ecosystem selection. *Proc. Natl. Acad. Sci. USA* **97**, 9110–9114.

- [82] Szabó, G., and J. Vukov. 2004. Cooperation for volunteering and partially random partnerships. *Phys. Rev. E* **69** (3), 036107.
- [83] Szathmáry, E., and L. Demeter. 1987. Group selection of early replicators and the origin of life. *J. Theoret. Biol.* **128**, 463–486.
- [84] Takahashi, N., and R. Mashima. 2003. The emergence of indirect reciprocity: Is the standing strategy the answer? Center for the Study of cultural and ecological foundations of the mind, Hokkaido University, Japan, Working paper series No. 29.
- [85] Taylor, P. D., and L. B. Jonker. 1978. Evolutionary stable strategies and game dynamics. *Math. Biosci.* **40**, 145–156.
- [86] Tilman, D., and P. Karieva, eds. 1997. *Spatial ecology: The role of space in population dynamics and interspecific interactions*. Princeton University Press Monographs in Population Biology, Princeton University Press, Princeton, NJ.
- [87] Traulsen, A., and M. A. Nowak. 2006. Emerging multi-level selection. Preprint.
- [88] Traulsen, A., A. M. Sengupta, and M. A. Nowak. 2005. Stochastic evolutionary dynamics on two levels. *J. Theoret. Biol.* **235**, 393–401.
- [89] Trivers, R. L. 1971. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57.
- [90] Uyenoyama, M., and M.W. Feldman. 1980. Theories of kin and group selection: A population genetics perspective. *Theoret. Population Biol.* **17**, 380–414.
- [91] von Neumann, J., and O. Morgenstern. 1944. *Theory of games and economic behavior*. Princeton University Press, Princeton, NJ.
- [92] Wade, M. J. 1976. Group selection among laboratory populations of *Tribolium*. *Proc. Natl. Acad. Sci. USA* **73**, 4604–4607.
- [93] Wedekind, C., and V. A. Braithwaite. 2002. The long term benefits of human generosity in indirect reciprocity. *Curr. Biol.* **12**, 1012–1015.
- [94] Wedekind, C., and M. Milinski. 2000. Cooperation through image scoring in humans. *Science* **288**, 850–852.
- [95] Weibull, J. 1995. *Evolutionary game theory*. MIT Press, Cambridge.
- [96] Whitlock, M. 2003. Fixation probability and time in subdivided populations. *Genetics* **164**, 767–779.
- [97] Williams, G. C. 1966. *Adaptation and natural selection*. Princeton University Press, Princeton, NJ.
- [98] Wilson, E. O. 1975. *Sociobiology*. Harvard University Press, Cambridge, MA.
- [99] Wilson, E. O., and B. Hölldobler. 1976. Eusociality: Origin and consequences. *Proc. Natl. Acad. Sci. USA* **102**, 13367–13371.
- [100] Wright, S. 1931. Evolution in Mendelian populations. *Genetics* **16**, 97–159.
- [101] Wynn-Edwards, V. C. 1962. *Animal dispersion in relation to social behavior*. Oliver and Boyd, London.
- [102] Zeeman, E. C. 1980. Population dynamics from game theory. In *Proceedings of an international conference on global theory of dynamical systems* (ed. by A. Nitecki and C. Robinson), Lecture Notes in Math. 819, Springer-Verlag, Berlin.

Program for Evolutionary Dynamics, Department of Mathematics, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, U.S.A.

E-mail: martin_nowak@harvard.edu