Five Rules for the Evolution of Cooperation

Martin A. Nowak

Cooperation is needed for evolution to construct new levels of organization. Genomes, cells, multicellular organisms, social insects, and human society are all based on cooperation. Cooperation means that selfish replicators forgo some of their reproductive potential to help one another. But natural selection implies competition and therefore opposes cooperation unless a specific mechanism is at work. Here I discuss five mechanisms for the evolution of cooperation: kin selection, direct reciprocity, indirect reciprocity, network reciprocity, and group selection. For each mechanism, a simple rule is derived that specifies whether natural selection can lead to cooperation.

Evolution is based on a fierce competition between individuals and should therefore reward only selfish behavior. Every gene, every cell, and every organism should be designed to promote its own evolutionary success at the expense of its competitors. Yet we observe cooperation on many levels of biological organization. Genes cooperate in genomes. Chromosomes cooperate in eukaryotic cells. Cells cooperate in multicellular organisms. There are many examples of cooperation among animals. Humans are the champions of cooperation: From hunter-gatherer societies to nation-states, cooperation is the decisive organizing principle of human society. No other life form on Earth is engaged in the same complex games of cooperation and defection. The question of how natural selection can lead to cooperative behavior has fascinated evolutionary biologists for several decades.

A cooperator is someone who pays a cost, c, for another individual to receive a benefit, b. A defector has no cost and does not deal out benefits. Cost and benefit are measured in terms of fitness. Reproduction can be genetic or cultural. In any mixed population, defectors have a higher average fitness than cooperators (Fig. 1). Therefore, selection acts to increase the relative abundance of defectors. After some time, cooperators vanish from the population. Remarkably, however, a population of only cooperators has the highest average fitness, whereas a population of only defectors has the lowest. Thus, natural selection constantly reduces the average fitness of the population. Fisher's fundamental theorem, which states that average fitness increases under constant selection, does not apply here because selection is frequency-dependent: The fitness of individuals depends on the frequency (= relative abundance) of cooperators in the population. We see that natural selection in well-mixed populations needs help for establishing cooperation.

Kin Selection

When J. B. S. Haldane remarked, "I will jump into the river to save two brothers or eight cousins," he anticipated what became later known as Hamilton's rule (*1*). This ingenious idea is that natural selection can favor cooperation if the donor and the recipient of an altruistic act are genetic relatives. More precisely, Hamilton's rule states that the coefficient of relatedness, r, must exceed the cost-to-benefit ratio of the altruistic act:

$$r > c/b \qquad (1)$$

Relatedness is defined as the probability of sharing a gene. The probability that two brothers share the same gene by descent is 1/2; the same probability for cousins is 1/8. Hamilton's theory became widely known as "kin selection" or "inclusive fitness" (*2–7*). When evaluating the fitness of the behavior induced by a certain gene, it is important to include the behavior's effect on kin who might carry the same gene. Therefore, the "extended phenotype" of cooperative behavior is the consequence of "selfish genes" (*8, 9*).

Direct Reciprocity

It is unsatisfactory to have a theory that can explain cooperation only among relatives. We also observe cooperation between unrelated individuals or even between members of different species. Such considerations led Trivers (*10*) to propose another mechanism for the evolution of cooperation, direct reciprocity. Assume that there are repeated encounters between the same two individuals. In every round, each player has a choice between cooperation and defection. If I cooperate now, you may cooperate later. Hence, it might pay off to cooperate. This game theoretic framework is known as the repeated Prisoner's Dilemma.

But what is a good strategy for playing this game? In two computer tournaments, Axelrod (*11*) discovered that the "winning strategy" was the simplest of all, tit-for-tat. This strategy always starts with a cooperation, then it does whatever the other player has done in the previous round: a cooperation for a cooperation, a defection for a defection. This simple concept captured the fascination of all enthusiasts of the repeated Prisoner's Dilemma. Many empirical and theoretical studies were inspired by Axelrod's groundbreaking work (*12–14*).

But soon an Achilles heel of the world champion was revealed: If there are erroneous moves caused by "trembling hands" or "fuzzy minds," then the performance of tit-for-tat declines (*15, 16*). Tit-for-tat cannot correct mistakes, because an accidental defection leads to a long sequence of retaliation. At first, tit-for-tat was replaced by generous-tit-for-tat (*17*), a strategy that cooperates whenever you cooperate, but sometimes cooperates although you have defected [with probability $1 - (c/b)$]. Natural selection can promote forgiveness.

Subsequently, tit-for-tat was replaced by win-stay, lose-shift, which is the even simpler idea of repeating your previous move whenever you are doing well, but changing otherwise (*18*). By various measures of success, win-stay, lose-shift is more robust than either tit-for-tat or generous-tit-for-tat (*15, 18*). Tit-for-tat is an efficient catalyst of cooperation in a society where nearly everybody is a defector, but once cooperation is established, win-stay, lose-shift is better able to maintain it.
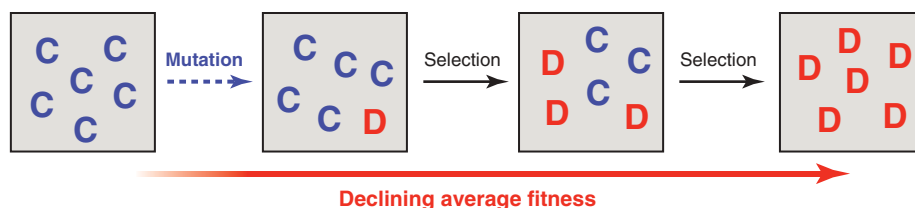
Fig. 1. Without any mechanism for the evolution of cooperation, natural selection favors defectors. In a mixed population, defectors, D, have a higher payoff (= fitness) than cooperators, C. Therefore, natural selection continuously reduces the abundance, i, of cooperators until they are extinct. The average fitness of the population also declines under natural selection. The total population size is given by N. If there are i cooperators and $N - i$ defectors, then the fitness of cooperators and defectors, respectively, is given by $f_C = [b(i - 1)/(N - 1)] - c$ and $f_D = bi/(N - 1)$. The average fitness of the population is given by $\bar{f} = (b - c)i/N$.

Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, and Department of Mathematics, Harvard University, Cambridge, MA 02138, USA. E-mail: martin_nowak@harvard.edu

The number of possible strategies for the repeated Prisoner's Dilemma is unlimited, but a simple general rule can be shown without any difficulty. Direct reciprocity can lead to the evolution of cooperation only if the probability, w, of another encounter between the same two individuals exceeds the cost-to-benefit ratio of the altruistic act:

$$w > c/b \qquad (2)$$

Indirect Reciprocity

Direct reciprocity is a powerful mechanism for the evolution of cooperation, but it leaves out certain aspects that are particularly important for humans. Direct reciprocity relies on repeated encounters between the same two individuals, and both individuals must be able to provide help, which is less costly for the donor than it is beneficial for the recipient. But often the interactions among humans are asymmetric and fleeting. One person is in a position to help another, but there is no possibility for a direct reciprocation. We help strangers who are in need. We donate to charities that do not donate to us. Direct reciprocity is like a barter economy based on the immediate exchange of goods, whereas indirect reciprocity resembles the invention of money. The money that fuels the engines of indirect reciprocity is reputation.

Helping someone establishes a good reputation, which will be rewarded by others. When deciding how to act, we take into account the possible consequences for our reputation. We feel strongly about events that affect us directly, but we also take a keen interest in the affairs of others, as demonstrated by the contents of gossip.

In the standard framework of indirect reciprocity, there are randomly chosen pairwise encounters where the same two individuals need not meet again. One individual acts as donor, the other as recipient. The donor can decide whether or not to cooperate. The interaction is observed by a subset of the population who might inform others. Reputation allows evolution of cooperation by indirect reciprocity (19). Natural selection favors strategies that base the decision to help on the reputation of the recipient. Theoretical and empirical studies of indirect reciprocity show that people who are more helpful are more likely to receive help (20–28).

Although simple forms of indirect reciprocity can be found in animals (29), only humans seem to engage in the full complexity of the game. Indirect reciprocity has substantial cognitive demands. Not only must we remember our own interactions, we must also monitor the ever-changing social network of the group. Language is needed to gain the information and spread the gossip associated with indirect reciprocity. Presumably, selection for indirect reciprocity and human language has played a decisive role in the evolution of human intelligence (28). Indirect

reciprocity also leads to the evolution of morality (30) and social norms (21, 22).

The calculations of indirect reciprocity are complicated and only a tiny fraction of this universe has been uncovered, but again a simple rule has emerged (19). Indirect reciprocity can only promote cooperation if the probability, q, of knowing someone's reputation exceeds the cost-to-benefit ratio of the altruistic act:

$$q > c/b \qquad (3)$$

Network Reciprocity

The argument for natural selection of defection (Fig. 1) is based on a well-mixed population, where everybody interacts equally likely with everybody else. This approximation is used by all standard approaches to evolutionary game dynamics (31–34). But real populations are not well mixed. Spatial structures or social networks imply that some individuals interact more often than others. One approach of capturing this effect is evolutionary graph theory (35), which allows us to study how spatial structure affects evolutionary and ecological dynamics (36–39).

The individuals of a population occupy the vertices of a graph. The edges determine who interacts with whom. Let us consider plain cooperators and defectors without any strategic complexity. A cooperator pays a cost, c, for each neighbor to receive a benefit, b. Defectors have no costs, and their neighbors receive no benefits. In this setting, cooperators can prevail by forming network clusters, where they help each other. The resulting "network reciprocity" is a generalization of "spatial reciprocity" (40).

Games on graphs are easy to study by computer simulation, but they are difficult to analyze mathematically because of the enormous number of possible configurations that can arise. Nonetheless, a surprisingly simple rule determines whether network reciprocity can favor cooperation (41). The benefit-to-cost ratio must exceed the average number of neighbors, k, per individual:

$$b/c > k \qquad (4)$$

Group Selection

Selection acts not only on individuals but also on groups. A group of cooperators might be more successful than a group of defectors. There have been many theoretical and empirical studies of group selection, with some controversy, and recently there has been a renaissance of such ideas under the heading of "multilevel selection" (42–50).

A simple model of group selection works as follows (51). A population is subdivided into groups. Cooperators help others in their own group. Defectors do not help. Individuals reproduce proportional to their payoff. Offspring are added to the same group. If a group reaches

a certain size, it can split into two. In this case, another group becomes extinct in order to constrain the total population size. Note that only individuals reproduce, but selection emerges on two levels. There is competition between groups because some groups grow faster and split more often. In particular, pure cooperator groups grow faster than pure defector groups, whereas in any mixed group, defectors reproduce faster than cooperators. Therefore, selection on the lower level (within groups) favors defectors, whereas selection on the higher level (between groups) favors cooperators. This model is based on "group fecundity selection," which means that groups of cooperators have a higher rate of splitting in two. We can also imagine a model based on "group viability selection,"
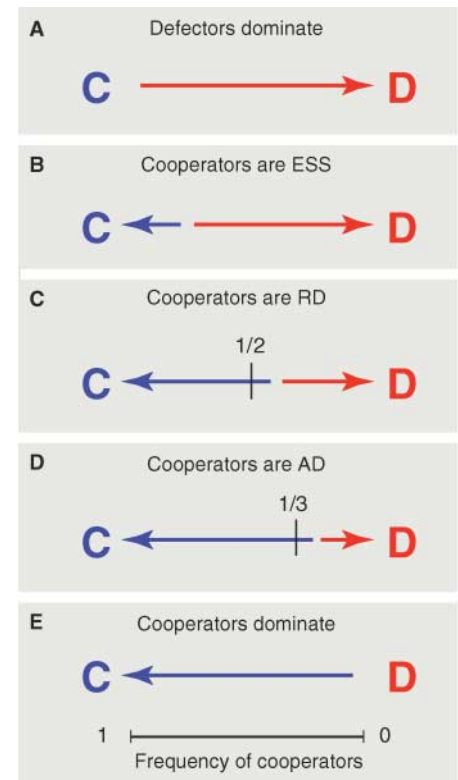
Fig. 2. Evolutionary dynamics of cooperators and defectors. The red and blue arrows indicate selection favoring defectors and cooperators, respectively. (**A**) Without any mechanism for the evolution of cooperation, defectors dominate. A mechanism for the evolution of cooperation can allow cooperators to be the evolutionarily stable strategy (ESS), risk-dominant (RD), or advantageous (AD) in comparison with defectors. (**B**) Cooperators are ESS if they can resist invasion by defectors. (**C**) Cooperators are RD if the basin of attraction of defectors is less than 1/2. (**D**) Cooperators are AD if the basin of attraction of defectors is less than 1/3. In this case, the fixation probability of a single cooperator in a finite population of defectors is greater than the inverse of the population size (for weak selection). (**E**) Some mechanisms allow cooperators to dominate defectors.

where groups of cooperators are less likely to go extinct.

In the mathematically convenient limit of weak selection and rare group splitting, we obtain a simple result (*51*): If n is the maximum group size and m is the number of groups, then group selection allows evolution of cooperation, provided that

$$b/c > 1 + (n/m) \qquad (5)$$

Evolutionary Success

Before proceeding to a comparative analysis of the five mechanisms, let me introduce some

Fig. 3. Five mechanisms for the evolution of cooperation. Kin selection operates when the donor and the recipient of an altruistic act are genetic relatives. Direct reciprocity requires repeated encounters between the same two individuals. Indirect reciprocity is based on reputation; a helpful individual is more likely to receive help. Network reciprocity means that clusters of cooperators outcompete defectors. Group selection is the idea that competition is not only between individuals but also between groups.

measures of evolutionary success. Suppose a game between two strategies, cooperators C and defectors D, is given by the payoff matrix

$$
\begin{array}{ccc}
 & C & D \\
C & \alpha & \beta \\
D & \gamma & \delta
\end{array}
$$

The entries denote the payoff for the row player. Without any mechanism for the evolution of cooperation, defectors dominate cooperators, which means $\alpha < \gamma$ and $\beta < \delta$. A mechanism for the evolution of cooperation can change these inequalities.

1) If $\alpha > \gamma$, then cooperation is an evolutionarily stable strategy (ESS). An infinitely large population of cooperators cannot be invaded by defectors under deterministic selection dynamics (*32*).

2) If $\alpha + \beta > \gamma + \delta$, then cooperators are risk-dominant (RD). If both strategies are ESS, then the risk-dominant strategy has the bigger basin of attraction.
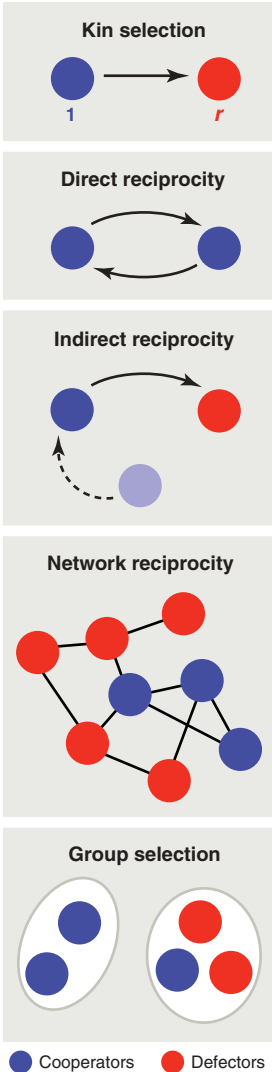
3) If $\alpha + 2\beta > \gamma + 2\delta$, then cooperators are advantageous (AD). This concept is important for stochastic game dynamics in finite populations. Here, the crucial quantity is the fixation probability of a strategy, defined as the probability that the lineage arising from a single mutant of that strategy will take over the entire population consisting of the other strategy. An AD strategy has a fixation probability greater than the inverse of the population size, $1/N$. The condition can also be expressed as a 1/3 rule: If the fitness of the invading strategy at a frequency of 1/3 is greater than the fitness of the resident, then the fixation probability of the invader is greater than $1/N$. This condition holds in the limit of weak selection (*52*).

A mechanism for the evolution of cooperation can ensure that cooperators become ESS, RD, or AD (Fig. 2). Some mechanisms even allow cooperators to dominate defectors, which means $\alpha > \gamma$ and $\beta > \delta$.

Comparative Analysis

We have encountered five mechanisms for the evolution of cooperation (Fig. 3). Although the mathematical formalisms underlying the five mechanisms are very different, at the center of each theory is a simple rule. I now present a coherent mathematical framework that allows the derivation of all five rules. The crucial idea is that each mechanism can be presented as a game between two strategies given by a 2×2 payoff matrix (Table 1). From this matrix, we can derive the relevant condition for evolution of cooperation.

For kin selection, I use the approach of inclusive fitness proposed by Maynard Smith (*31*). The relatedness between two players is r. Therefore, your payoff multiplied by r is added to mine. A second method, shown in (*53*), leads to a different matrix but the same result. For direct reciprocity, the cooperators use tit-for-tat while the defectors use "always-defect." The expected number of rounds is $1/(1 - w)$. Two tit-for-tat players cooperate all the time. Tit-for-tat versus always-defect cooperates only in the first move and then defects. For indirect reciprocity, the probability of knowing someone's reputation is given by q. A cooperator helps unless the reputation of the other person indicates a defector. A defector never helps. For network reciprocity, it can be shown that the expected frequency of cooperators is described by a standard replicator equation with a transformed payoff matrix (*54*). For group selection, the payoff matrices of the two games—within

Table 1. Each mechanism can be described by a simple 2×2 payoff matrix, which specifies the interaction between cooperators and defectors. From these matrices we can directly derive the necessary conditions for evolution of cooperation. The parameters c and b denote, respectively, the cost for the donor and the benefit for the recipient. For network reciprocity, we use the parameter $H = [(b - c)k - 2c]/[(k + 1)(k - 2)]$. All conditions can be expressed as the benefit-to-cost ratio exceeding a critical value. See (*53*) for further explanations of the underlying calculations.

| | | Payoff matrix | | Cooperation is… | | | |
		C	D	ESS	RD	AD	
Kin selection	C	$(b-c)(1+r)$	$br-c$	$\dfrac{b}{c} > \dfrac{1}{r}$	$\dfrac{b}{c} > \dfrac{1}{r}$	$\dfrac{b}{c} > \dfrac{1}{r}$	r…genetic relatedness
	D	$b-rc$	0				
Direct reciprocity	C	$(b-c)/(1-w)$	$-c$	$\dfrac{b}{c} > \dfrac{1}{w}$	$\dfrac{b}{c} > \dfrac{2-w}{w}$	$\dfrac{b}{c} > \dfrac{3-2w}{w}$	w…probability of next round
	D	b	0				
Indirect reciprocity	C	$b-c$	$-c(1-q)$	$\dfrac{b}{c} > \dfrac{1}{q}$	$\dfrac{b}{c} > \dfrac{2-q}{q}$	$\dfrac{b}{c} > \dfrac{3-2q}{q}$	q…social acquaintanceship
	D	$b(1-q)$	0				
Network reciprocity	C	$b-c$	$H-c$	$\dfrac{b}{c} > k$	$\dfrac{b}{c} > k$	$\dfrac{b}{c} > k$	k…number of neighbors
	D	$b-H$	0				
Group selection	C	$(b-c)(m+n)$	$(b-c)m-cn$	$\dfrac{b}{c} > 1+\dfrac{n}{m}$	$\dfrac{b}{c} > 1+\dfrac{n}{m}$	$\dfrac{b}{c} > 1+\dfrac{n}{m}$	n…group size m…number of groups
	D	bn	0				

and between groups—can be added up. The details of all these arguments and their limitations are given in (*53*).

For kin selection, the calculation shows that Hamilton's rule, $r > c/b$, is the decisive criterion for all three measures of evolutionary success: ESS, RD, and AD. Similarly, for network reciprocity and group selection, we obtain the same condition for all three evaluations, namely $b/c > k$ and $b/c > 1 + (n/m)$, respectively. The reason is the following: If these conditions hold, then cooperators dominate defectors. For direct and indirect reciprocity, we find that the ESS conditions lead to $w > c/b$ and $q > c/b$, respectively. Slightly more stringent conditions must hold for cooperation to be RD or AD.

Conclusion

Each of the five possible mechanisms for the evolution of cooperation—kin selection, direct reciprocity, indirect reciprocity, network reciprocity and group selection—can be described by a characteristic 2 × 2 payoff matrix, from which we can directly derive the fundamental rules that specify whether cooperation can evolve (Table 1). Each rule can be expressed as the benefit-to-cost ratio of the altruistic act being greater than some critical value. The payoff matrices can be imported into standard frameworks of evolutionary game dynamics. For example, we can study replicator equations for games on graphs (*54*), for group selection, and for kin selection. This creates interesting new possibilities for the theory of evolutionary dynamics (*55*).

I have not discussed all potential mechanisms for the evolution of cooperation. An interesting possibility is offered by "green beard" models where cooperators recognize each other via arbitrary labels (*56–58*). Another way to obtain cooperation is making the game voluntary rather than obligatory: If players can choose to cooperate, defect, or not play at all, then some level of cooperation usually prevails in dynamic oscillations (*59*). Punishment is an important factor that can promote cooperative behavior in some situations (*60–64*), but it is not a mechanism for the evolution of cooperation. All evolutionary models of punishment so far are based on underlying mechanisms such as indirect reciprocity (*65*), group selection (*66*, *67*), or network reciprocity (*68*). Punishment can enhance the level of cooperation that is achieved in such models.

Kin selection has led to mathematical theories (based on the Price equation) that are more general than just analyzing interactions between genetic relatives (*4*, *5*). The interacting individuals can have any form of phenotypic correlation. Therefore, kin selection theory also provides an approach to compare different mechanisms for the evolution of cooperation (*69*, *70*).

The two fundamental principles of evolution are mutation and natural selection. But evolution is constructive because of cooperation. New levels of organization evolve when the competing units on the lower level begin to cooperate. Cooperation allows specialization and thereby promotes biological diversity. Cooperation is the secret behind the open-endedness of the evolutionary process. Perhaps the most remarkable aspect of evolution is its ability to generate cooperation in a competitive world. Thus, we might add "natural cooperation" as a third fundamental principle of evolution beside mutation and natural selection.

References and Notes

1. W. D. Hamilton, *J. Theor. Biol.* **7**, 1 (1964).
2. A. Grafen, in *Oxford Surveys in Evolutionary Biology*, vol. 2, R. Dawkins, M. Ridley, Eds. (Oxford Univ. Press, Oxford, 1985), pp. 28–89.
3. P. D. Taylor, *Evol. Ecol.* **6**, 352 (1992).
4. D. C. Queller, *Am. Nat.* **139**, 540 (1992).
5. S. A. Frank, *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, NJ, 1998).
6. S. A. West, I. Pen, A. S. Griffin, *Science* **296**, 72 (2002).
7. K. R. Foster, T. Wenseleers, F. L. W. Ratnieks, *Trends Ecol. Evol.* **21**, 57 (2006).
8. R. Dawkins, *The Selfish Gene* (Oxford Univ. Press, Oxford, 1976).
9. E. O. Wilson, *Sociobiology* (Harvard Univ. Press, Cambridge, 1975).
10. R. Trivers, *Q. Rev. Biol.* **46**, 35 (1971).
11. R. Axelrod, *The Evolution of Cooperation* (Basic Books, New York, 1984).
12. R. Axelrod, W. D. Hamilton, *Science* **211**, 1390 (1981).
13. M. Milinski, *Nature* **325**, 434 (1987).
14. L. A. Dugatkin, *Cooperation Among Animals* (Oxford Univ. Press, Oxford, 1997).
15. D. Fudenberg, E. Maskin, *Am. Econ. Rev.* **80**, 274 (1990).
16. R. Selten, P. Hammerstein, *Behav. Brain Sci.* **7**, 115 (1984).
17. M. A. Nowak, K. Sigmund, *Nature* **355**, 250 (1992).
18. M. A. Nowak, K. Sigmund, *Nature* **364**, 56 (1993).
19. M. A. Nowak, K. Sigmund, *Nature* **393**, 573 (1998).
20. C. Wedekind, M. Milinski, *Science* **288**, 850 (2000).
21. H. Ohtsuki, Y. Iwasa, *J. Theor. Biol.* **231**, 107 (2004).
22. H. Brandt, K. Sigmund, *J. Theor. Biol.* **231**, 475 (2004).
23. O. Leimar, P. Hammerstein, *Proc. R. Soc. London Ser. B* **268**, 745 (2001).
24. M. Milinski, D. Semmann, H. J. Krambeck, *Nature* **415**, 424 (2002).
25. M. A. Fishman, *J. Theor. Biol.* **225**, 285 (2003).
26. M. D. Hauser, M. K. Chen, F. Chen, E. Chuang, *Proc. R. Soc. London Ser. B* **270**, 2363 (2003).
27. K. Panchanathan, R. Boyd, *Nature* **432**, 499 (2004).
28. M. A. Nowak, K. Sigmund, *Nature* **437**, 1291 (2005).
29. R. Bshary, A. S. Grutter, *Nature* **441**, 975 (2006).
30. R. D. Alexander, *The Biology of Moral Systems* (de Gruyter, New York, 1987).
31. J. Maynard Smith, *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
32. J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, 1998).
33. J. Hofbauer, K. Sigmund, *Bull. Am. Math. Soc.* **40**, 479 (2003).
34. M. A. Nowak, K. Sigmund, *Science* **303**, 793 (2004).
35. E. Lieberman, C. Hauert, M. A. Nowak, *Nature* **433**, 312 (2005).
36. R. Durrett, S. A. Levin, *Theor. Popul. Biol.* **46**, 363 (1994).
37. M. P. Hassell, H. N. Comins, R. M. May, *Nature* **370**, 290 (1994).
38. C. Hauert, M. Doebeli, *Nature* **428**, 643 (2004).
39. R. M. May, *Trends Ecol. Evol.* **21**, 394 (2006).
40. M. A. Nowak, R. M. May, *Nature* **359**, 826 (1992).
41. H. Ohtsuki, C. Hauert, E. Lieberman, M. A. Nowak, *Nature* **441**, 502 (2006).
42. G. C. Williams, D. C. Williams, *Evolution* **11**, 32 (1957).
43. D. S. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 143 (1975).
44. P. D. Taylor, D. S. Wilson, *Evolution* **42**, 193 (1988).
45. A. R. Rogers, *Am. Nat.* **135**, 398 (1990).
46. R. E. Michod, *Darwinian Dynamics* (Princeton Univ. Press, Princeton, NJ, 1999).
47. L. Keller, Ed., *Levels of Selection in Evolution* (Princeton Univ. Press, Princeton, NJ, 1999).
48. J. Paulsson, *Genetics* **161**, 1373 (2002).
49. P. B. Rainey, K. Rainey, *Nature* **425**, 72 (2003).
50. E. O. Wilson, B. Hölldobler, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13367 (2005).
51. A. Traulsen, M. A. Nowak, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10952 (2006).
52. M. A. Nowak, A. Sasaki, C. Taylor, D. Fudenberg, *Nature* **428**, 646 (2004).
53. See supporting material on *Science* Online.
54. H. Ohtsuki, M. A. Nowak, *J. Theor. Biol.* **243**, 86 (2006).
55. M. A. Nowak, *Evolutionary Dynamics* (Harvard Univ. Press, Cambridge, MA, 2006).
56. R. L. Riolo, M. D. Cohen, R. Axelrod, *Nature* **414**, 441 (2001).
57. A. Traulsen, H. G. Schuster, *Phys. Rev. E* **68**, 046129 (2003).
58. V. A. Jansen, M. van Baalen, *Nature* **440**, 663 (2006).
59. C. Hauert, S. De Monte, J. Hofbauer, K. Sigmund, *Science* **296**, 1129 (2002).
60. T. Yamagishi, *J. Pers. Soc. Psychol.* **51**, 110 (1986).
61. E. Fehr, S. Gaechter, *Nature* **415**, 137 (2002).
62. E. Fehr, U. Fischbacher, *Nature* **425**, 785 (2003).
63. C. F. Camerer, E. Fehr, *Science* **311**, 47 (2006).
64. Ö. Gürerk, B. Irlenbusch, B. Rockenbach, *Science* **312**, 108 (2006).
65. K. Sigmund, C. Hauert, M. A. Nowak, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10757 (2001).
66. R. Boyd, H. Gintis, S. Bowles, P. J. Richerson, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3531 (2003).
67. S. Bowles, H. Gintis, *Theor. Popul. Biol.* **65**, 17 (2004).
68. M. Nakamaru, Y. Iwasa, *Evol. Ecol. Res.* **7**, 853 (2005).
69. L. Lehmann, L. Keller, *J. Evol. Biol.* **19**, 1365 (2006).
70. J. A. Fletcher, M. Zwick, *Am. Nat.* **168**, 252 (2006).
71. Supported by the John Templeton Foundation and the NSF-NIH joint program in mathematical biology (NIH grant 1R01GM078986-01). The Program for Evolutionary Dynamics at Harvard University is sponsored by J. Epstein.

Five rules for the evolution of cooperation
Supporting Online Material

Martin A. Nowak

Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology,
Department of Mathematics, Harvard University, Cambridge, MA 02138, USA
Email: martin_nowak@harvard.edu

The standard payoff matrix between cooperators, C, and defectors, D, is given by

$$
\begin{array}{cc}
 & \begin{array}{cc} C & \quad D \end{array} \\
\begin{array}{c} C \\ D \end{array} &
\begin{pmatrix} b-c & -c \\ b & 0 \end{pmatrix}
\end{array}
\tag{1}
$$

The entries in the payoff matrix refer to the 'row player'. For each interaction, a cooperator pays a cost, c. Interacting with a cooperator leads to a benefit, b. Thus, the payoff for C versus C is $b-c$; the payoff for C versus D is $-c$; the payoff for D versus C is b; the payoff for D versus D is 0. Usually, we assume that $b > c$, otherwise the payoff for two cooperators is less than the payoff for two defectors, and cooperation becomes nonsensical. I will now discuss how to derive the five 2×2 matrices of Table 1.

1. Kin selection

A simple way to study games between relatives was proposed by Maynard Smith for the Hawk-Dove game (S1). I will use this method to analyze the interaction between cooperators and defectors. Consider a population where the average relatedness between individuals is given by r, which is a number between 0 and 1. The concept of inclusive fitness implies that the payoff received by a relative is added to my own payoff multiplied by r. Therefore, we obtain the modified matrix

$$
\begin{array}{cc}
 & \begin{array}{cc} C & \quad\quad D \end{array} \\
\begin{array}{c} C \\ D \end{array} &
\begin{pmatrix} (b-c)(1+r) & br-c \\ b-rc & 0 \end{pmatrix}
\end{array}
\tag{2}
$$

For this payoff matrix, cooperators dominate defectors if $b/c > 1/r$. In this case, cooperators are also evolutionarily stable (ESS), risk-dominant (RD) and advantageous (AD); see main text for the definition of ESS, RD and AD.

Another method to describe games among relatives was proposed by Grafen (S2) also in the context of the Hawk Dove game. Let us assume that interactions are more likely between relatives. Each individual has a fraction, r, of its interactions with its own relatives, who use the same strategy, and a fraction $1 - r$ with random individuals from the population, who could use the same or a different strategy. Let x denote the frequency of cooperators. The frequency of defectors is given by $1 - x$. The fitness of a cooperator is $F_C(x) = r(b-c) + (1-r)[(b-c)x - c(1-x)]$. The fitness of a defector is $F_D(x) = (1-r)bx$. These linear fitness function can be described by the payoff matrix

$$
\begin{array}{c}
\\
C \\
D
\end{array}
\begin{array}{cc}
C & D \\
\left(\begin{array}{cc} F_C(1) & F_C(0) \\ F_D(1) & F_D(0) \end{array} \right)
\end{array}
=
\begin{array}{c}
\\
C \\
D
\end{array}
\begin{array}{cc}
C & D \\
\left(\begin{array}{cc} b-c & br-c \\ b(1-r) & 0 \end{array} \right)
\end{array}
\tag{3}
$$

Again we find that cooperators dominate defectors if $b/c > 1/r$. Therefore, both approaches give the same answer, which turns out to be Hamilton's rule (S3). Note that the exact population genetics of sexually reproducing, diploid individuals require more complicated calculations (S4).

2. Direct reciprocity

In order to derive a necessary condition for the evolution of cooperation in the repeated Prisoner's Dilemma, we can study the interaction between 'always-defect' (ALLD) and tit-for-tat (TFT). If TFT cannot hold itself against ALLD then no cooperative strategy can. TFT starts with cooperation and then does whatever the opponent has done in the previous move. We ignore erroneous moves. In this setting, TFT playing ALLD will cooperate in the first round and defect afterwards. Therefore, the payoff for TFT versus ALLD is $-c$. The payoff for ALLD versus TFT is b. Only the first round leads to a payoff, while all subsequent rounds consist of mutual defection and produce zero payoff for both players. The payoff for ALLD versus ALLD is 0. The payoff for TFT versus TFT is $(b-c)/(1-w)$. The parameter w denotes the probability of playing another round between the same two players. The average number of rounds is given by $1/(1-w)$. Hence, we obtain the payoff matrix

$$
\begin{array}{c}
\\
C \\
D
\end{array}
\begin{array}{cc}
C & D \\
\left(\begin{array}{cc} (b-c)/(1-w) & -c \\ b & 0 \end{array} \right)
\end{array}
\tag{4}
$$

From this matrix we immediately obtain the three conditions for ESS, RD and AD that are shown in Table 1. For cooperators (using TFT) to be ESS in comparison with ALLD,

we need $b/c > 1/w$. Slightly more stringent conditions are required for cooperators to be RD or AD. Note that ALLD is always an ESS, and hence cooperators cannot dominate defectors in the framework of direct reciprocity.

The calculations for exploring the interactions of larger sets of probabilistic strategies of the repeated Prisoner's Dilemma in the presence of noise (S5) are more complicated (S6, S7). Often there are cycles between ALLD, TFT and unconditional cooperators (ALLC) (S8). The point is that $b/c > 1/w$ is a necessary condition for the evolution of cooperation. This argument is related to the Folk theorem which states that certain trigger strategies can achieve cooperation if there are enough rounds of the repeated Prisoner's Dilemma (S9, S10).

3. Indirect reciprocity

Indirect reciprocity describes the interaction between a donor and a recipient. The donor can either cooperate or defect. The basic idea of indirect reciprocity is that cooperation increases ones own reputation, while defection reduces it. The fundamental question is whether natural selection can lead to strategies that base their decision to cooperate (at least to some extent) on the reputation of the recipient.

A strategy for indirect reciprocity consists of an action rule and an assessment norm. The action rule determines whether to cooperate or to defect in a particular situation depending on the recipient's reputation (image score) and ones own. The assessment norm determines how to evaluate an interaction between two other people as an observer. Most analytic calculations of indirect reciprocity assume binary image scores: the reputation of someone is either 'good' or 'bad'. Nobody so far has succeeded to formulate an exact analysis for the realistic situation where the image scores are more gradual and different people have different image scores of the same person as a consequence of private and incomplete information.

In order to derive a necessary condition for the evolution of cooperation by indirect reciprocity, let us study the interaction between the two basic strategies: (i) defectors and (ii) cooperators who cooperate unless they know the reputation of the other person to indicate a defector. The parameter q denotes the probability to know the reputation of another person. A cooperator always helps another cooperator. A cooperator helps a

3

defector with probability $1 - q$. Defectors never help. Hence, we obtain the payoff matrix

$$
\begin{array}{cc}
 & \begin{array}{cc} C & \quad\quad D \end{array} \\
\begin{array}{c} C \\ D \end{array} &
\left(\begin{array}{cc}
b - c & -c(1 - q) \\
b(1 - q) & 0
\end{array} \right)
\end{array}
\tag{5}
$$

We have assumed that in a pairwise interaction both individuals are donor and recipient. If only one of them is donor and the other recipient, then all entries are multiplied by $1/2$, which makes no difference. Note that the payoff matrices (4) and (5) are identical (up to a factor) if we set $w = q$. Hence, indirect reciprocity leads to the same three conditions for ESS, RD and AD as direct reciprocity with q instead of w (see Table 1).

4. Network reciprocity

Spatial games can lead to cooperation in the absence of any strategic complexity (S11): unconditional cooperators can coexist with and sometimes outcompete unconditional defectors. This effect is called 'spatial reciprocity'. Spatial games are usually played on regular lattices such as square, triangular or hexagonal lattices. Network reciprocity is a generalization of spatial reciprocity to graphs. Individuals occupy the vertices of a graph. The edges denotes who interacts with whom. In principle, there can be two different graphs. The 'interaction graph' determines who plays with whom. The 'replacement graph' determines who competes with whom for reproduction, which can be genetic or cultural. Here we assume that the interaction and replacement graphs are identical. Evolutionary graph theory (S12) is a general approach to study the effect of population structure or social networks on evolutionary dynamics.

We consider a 'two coloring' of the graph: each vertex can be either a cooperator or a defector. A cooperator pays a cost, c, for each neighbor to receive a benefit, b. Defectors pay no cost and distribute no benefits. According to this simple rule the payoff, P, for each individual is evaluated. The fitness of an individual is given by $1 - \omega + \omega P$ where $\omega \in [0, 1]$ denotes the intensity of selection. Weak selection means that ω is much smaller than 1. Evolutionary updating works as follows: in each time step a random individual is chosen to die; the neighbors compete for the empty site proportional to their fitness.

We want to calculate the 'fixation probabilities', ρ_C, that a single cooperator starting in a random position on the graph takes over an entire population of defectors, and ρ_D, that a single defector starting in a random position on the graph takes over an entire population of cooperators. The fixation probability of a neutral mutant is $1/N$ where N is

the population size. If $\rho_C > 1/N$ then selection favors the fixation of cooperators; in this case cooperation is an advantageous strategy (AD).

For regular graphs, where each individual has exactly k neighbors, a calculation using pair-approximation (S13) leads to a surprisingly simple result: if $b/c > k$ then $\rho_C > 1/N > \rho_D$ for weak selection and large N. Numerical simulations show that this result is also an excellent approximation for non-regular graphs such as random graphs and scale free networks (S13).

The pair approximation calculation (for $k \geq 3$) also leads to a deterministic differential equation which describes how the expected frequency of cooperators (and defectors) changes over time (S14). This differential equation turns out to be a standard replicator equation (S15,S16) with a modified payoff matrix. For the interaction between cooperators and defectors on a graph with average degree k this modified payoff matrix is of the form

$$
\begin{array}{cc}
 & \begin{array}{cc} C & \quad D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} b-c & H-c \\ b-H & 0 \end{array} \right)
\end{array}
\tag{6}
$$

where

$$
H = \frac{(b-c)k - 2c}{(k+1)(k-2)}.
$$

It is easy to see that the payoff matrix (6) leads to the condition $b/c > k$ for cooperators to dominate defectors. In this case, cooperators are also ESS, RD and AD.

5. Group selection

Many models of group selection have been proposed over the years (S17-S29). It is difficult to formulate a model which is so simple that it can be studied analytically. One such model is the following (S30). A population is subdivided into m groups. The maximum size of a group is n. Individuals interact with others in the same group. Cooperators pay a cost c for each other member of the group to receive a benefit b. Defectors pay no costs and distribute no benefits. The fitness of an individual is $1 - \omega + \omega P$, where P is the payoff and ω the intensity of selection. At each time step, an individual from the entire population is chosen for reproduction proportional to fitness. The offspring is added to the same group. If the group reaches the maximum size, it can split into two groups with a certain probability, p. In this case, a randomly selected group dies to prevent the population from exploding. The maximum population size is mn. With probability $1 - p$ the group does not divide. In this case, a random individual of that group is chosen to die.

5

For small p, the fixation probability of a single cooperator in the entire population is given by the fixation probability of a single cooperator in a group times the fixation probability of that group.

For the fixation probability of one cooperator in a group of $n-1$ defectors we obtain $\phi_C = [1/n][1 - (b + cn - c)\omega/2]$. For the fixation probability of one cooperator group in a population of $m-1$ defector groups we obtain $\Phi_C = [1/m][1 + (b - c)(m - 1)\omega/2]$. Both results hold for weak selection (small ω). Note that the lower level selection within a group is frequency dependent and opposes cooperators, while the higher level selection between groups is constant and favors cooperators.

In the case of rare group splitting, the fixation probability of a single cooperator in the entire population, is given by the product $\rho_C = \phi_C \Phi_C$. It is easy to see that $\rho_C > 1/(nm)$ leads to $b/c > 1 + [n/(m - 2)]$. If this inequality holds, then cooperators are advantageous (and defectors disadvantageous) once both levels of selection are combined.

For a large number of groups, $m \gg 1$, we obtain the simplified condition $b/c > 1 + n/m$. The benefit to cost ratio of the altruistic act must exceed one plus the ratio of group size over the number of groups. The model can also be extended to include migration, which can be seen as 'noise' of group selection. In this case, the relevant criterion is $b/c > 1 + \mu + n/m$, where μ is the average number of migrants accepted over the life-time of a group (S30).

Now comes a surprising move that allows us to reduce the evolutionary dynamics on two levels of selection to a single two-person game on one level of selection. The payoff matrix that describes the interactions within a group is given by

$$
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \begin{pmatrix} b - c & -c \\ b & 0 \end{pmatrix}
\end{array}
\tag{7}
$$

Between groups there is no game dynamical interaction in our model, but groups divide at rates that are proportional to the average fitness of individuals in that group. Therefore one can say that cooperator groups have a constant payoff $b - c$, while defector groups have a constant payoff 0. Hence, in a sense the following 'game' between groups is happening

$$
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \begin{pmatrix} b - c & b - c \\ 0 & 0 \end{pmatrix}
\end{array}
\tag{8}
$$

Remember also that the 'fitness' of a group is $1 - \omega + \omega P$ where P is its 'payoff'. We can now multiply the first matrix by the group size, n, and the second matrix by the number

6

of groups, m, and add the two matrices. The result is

$$
\begin{array}{cc}
 & \begin{array}{cc} C & \qquad\quad D \end{array} \\
\begin{array}{c} C \\ D \end{array} &
\left(\begin{array}{cc}
(b-c)(n+m) & bm - c(m+n) \\
bn & 0
\end{array} \right)
\end{array}
\tag{9}
$$

In this simple 2×2 game, cooperators dominate defectors if $b/c > 1 + (n/m)$. In this case, cooperators are also ESS, RD and AD.

Interestingly, the method also gives the right answer for two arbitrary payoff matrices describing the games on the two levels. The intuition for adding the two matrices multiplied with the respective population size is as follows. For fixation of a new strategy in a homogeneous population using the other strategy, first the game dynamics within one group (of size n) have to be won and then the game dynamics between m groups have to be won. For weak selection and large m and n, the overall fixation probability is the same as the fixation probability in the single game using the combined matrix (9) and population size, mn. The stochastic process on two levels can be studied by a standard replicator equation using the combined matrix.

Finally, note that payoff matrix (9) for group selection is structurally identical to the payoff matrix (3) for kin selection if we set $r = m/(m + n)$ pointing to yet another interesting relationship between kin selection and group selection (S31).

References

S1. J. Maynard Smith, *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, UK, 1982).

S2. A. Grafen, *Anim. Behav.* **27**, 905 (1979).

S3. W. D. Hamilton, *J. Theor. Biol.* **7**, 1 (1964).

S4. L. L. Cavalli-Sforza, M. W. Feldman, *Theor. Popul. Biol.* **14**, 268 (1978).

S5. R. M. May, *Nature* **327**, 15 (1987).

S6. D. Fudenberg, E. Maskin, *Am. Econ. Rev.* **80**, 274 (1990).

S7. M. A. Nowak, K. Sigmund, *Acta Appl. Math.* **20**, 247 (1990).

S8. L. A. Imhof, D. Fudenberg, M. A. Nowak, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10797 (2005).

S9. D. Fudenberg, E. Maskin, *Econometrica* **50**, 533 (1986).

S10. K. Binmore, *Fun and Games* (Heath, Lexington, MA, 1992).

S11. M. A. Nowak, R. M. May, *Nature* **359**, 826 (1992).

S12. E. Lieberman, C. Hauert, M. A. Nowak, *Nature* **433**, 312 (2005).

S13. H. Ohtsuki, C. Hauert, E. Lieberman, M. A. Nowak, *Nature*, **441** (2006).

S14. H. Ohtsuki, M. A. Nowak, *J. Theor. Biol.* **243**, 86 (2006).

S15. P. D. Taylor, L. B. Jonker *Math. Biosci.* **40**, 145 (1978).

S16. J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, UK, 1998).

S17. B. R. Levin, W. L. Kilmer, *Evolution* **28**, 527 (1974).

S18. J. Maynard Smith, *Q. Rev. Biol.* **51**, 277 (1976).

S19. M. Slatkin, M. J. Wade, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3531 (1978).

S20. M. Uyenoyama, M. W. Feldman, *Theor. Popul. Biol.* **17**, 380 (1980).

S21. J. F. Crow, K. Aoki, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2628 (1982).

S22. R. G. Leigh, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2985 (1983).

S23. D. S. Wilson, *Annu. Rev. Ecol. Syst.* **14**, 159 (1983).

S24. L. Nunney, *Am. Nat.* **126**, 212 (1985).

S25. H. Harpending, A. Rogers, *J. Theor. Biol.* **127**, 51 (1986).

S26. E. Szathmáry, L. Demeter, *J. Theor. Biol.* **128**, 463 (1987).

S27. R. Boyd, P. J. Richerson, *J. Theor. Biol.* **215**, 287 (2002).

S28. B. Kerr, P. Godfrey-Smith, *Biol. Philos.* **17**, 477 (2002).

S29. J. Fletcher, M. Zwick, *J. Theor. Biol.* **228**, 303 (2004).

S30. A. Traulsen, M. A. Nowak, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10952 (2006).

S31. S. A. Frank, *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, NJ, 1998).

8

Supporting Online Material for

Five Rules for the Evolution of Cooperation

Martin A. Nowak

This PDF file includes:

SOM Text

References

Five rules for the evolution of cooperation
Supporting Online Material

Martin A. Nowak

Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology,
Department of Mathematics, Harvard University, Cambridge, MA 02138, USA
Email: martin_nowak@harvard.edu

The standard payoff matrix between cooperators, C, and defectors, D, is given by

$$
\begin{array}{c}
\begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array}
\left(\begin{array}{cc} b-c & -c \\ b & 0 \end{array} \right)
\end{array}
\tag{1}
$$

The entries in the payoff matrix refer to the 'row player'. For each interaction, a cooperator
pays a cost, c. Interacting with a cooperator leads to a benefit, b. Thus, the payoff for
C versus C is $b-c$; the payoff for C versus D is $-c$; the payoff for D versus C is b; the
payoff for D versus D is 0. Usually, we assume that $b > c$, otherwise the payoff for two
cooperators is less than the payoff for two defectors, and cooperation becomes nonsensical.
I will now discuss how to derive the five 2×2 matrices of Table 1.

1. Kin selection

A simple way to study games between relatives was proposed by Maynard Smith for
the Hawk-Dove game (S1). I will use this method to analyze the interaction between
cooperators and defectors. Consider a population where the average relatedness between
individuals is given by r, which is a number between 0 and 1. The concept of inclusive
fitness implies that the payoff received by a relative is added to my own payoff multiplied
by r. Therefore, we obtain the modified matrix

$$
\begin{array}{c}
\begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array}
\left(\begin{array}{cc} (b-c)(1+r) & br-c \\ b-rc & 0 \end{array} \right)
\end{array}
\tag{2}
$$

For this payoff matrix, cooperators dominate defectors if $b/c > 1/r$. In this case, coopera-
tors are also evolutionarily stable (ESS), risk-dominant (RD) and advantageous (AD); see
main text for the definition of ESS, RD and AD.

1

Another method to describe games among relatives was proposed by Grafen (S2) also in the context of the Hawk Dove game. Let us assume that interactions are more likely between relatives. Each individual has a fraction, r, of its interactions with its own relatives, who use the same strategy, and a fraction $1 - r$ with random individuals from the population, who could use the same or a different strategy. Let x denote the frequency of cooperators. The frequency of defectors is given by $1 - x$. The fitness of a cooperator is $F_C(x) = r(b-c) + (1-r)[(b-c)x - c(1-x)]$. The fitness of a defector is $F_D(x) = (1-r)bx$. These linear fitness function can be described by the payoff matrix

$$
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} F_C(1) & F_C(0) \\ F_D(1) & F_D(0) \end{array} \right)
\end{array}
=
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} b - c & br - c \\ b(1-r) & 0 \end{array} \right)
\end{array}
\tag{3}
$$

Again we find that cooperators dominate defectors if $b/c > 1/r$. Therefore, both approaches give the same answer, which turns out to be Hamilton's rule (S3). Note that the exact population genetics of sexually reproducing, diploid individuals require more complicated calculations (S4).

2. Direct reciprocity

In order to derive a necessary condition for the evolution of cooperation in the repeated Prisoner's Dilemma, we can study the interaction between 'always-defect' (ALLD) and tit-for-tat (TFT). If TFT cannot hold itself against ALLD then no cooperative strategy can. TFT starts with cooperation and then does whatever the opponent has done in the previous move. We ignore erroneous moves. In this setting, TFT playing ALLD will cooperate in the first round and defect afterwards. Therefore, the payoff for TFT versus ALLD is $-c$. The payoff for ALLD versus TFT is b. Only the first round leads to a payoff, while all subsequent rounds consist of mutual defection and produce zero payoff for both players. The payoff for ALLD versus ALLD is 0. The payoff for TFT versus TFT is $(b-c)/(1-w)$. The parameter w denotes the probability of playing another round between the same two players. The average number of rounds is given by $1/(1-w)$. Hence, we obtain the payoff matrix

$$
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} (b-c)/(1-w) & -c \\ b & 0 \end{array} \right)
\end{array}
\tag{4}
$$

From this matrix we immediately obtain the three conditions for ESS, RD and AD that are shown in Table 1. For cooperators (using TFT) to be ESS in comparison with ALLD,

we need $b/c > 1/w$. Slightly more stringent conditions are required for cooperators to be RD or AD. Note that ALLD is always an ESS, and hence cooperators cannot dominate defectors in the framework of direct reciprocity.

The calculations for exploring the interactions of larger sets of probabilistic strategies of the repeated Prisoner's Dilemma in the presence of noise (S5) are more complicated (S6, S7). Often there are cycles between ALLD, TFT and unconditional cooperators (ALLC) (S8). The point is that $b/c > 1/w$ is a necessary condition for the evolution of cooperation. This argument is related to the Folk theorem which states that certain trigger strategies can achieve cooperation if there are enough rounds of the repeated Prisoner's Dilemma (S9, S10).

3. Indirect reciprocity

Indirect reciprocity describes the interaction between a donor and a recipient. The donor can either cooperate or defect. The basic idea of indirect reciprocity is that cooperation increases ones own reputation, while defection reduces it. The fundamental question is whether natural selection can lead to strategies that base their decision to cooperate (at least to some extent) on the reputation of the recipient.

A strategy for indirect reciprocity consists of an action rule and an assessment norm. The action rule determines whether to cooperate or to defect in a particular situation depending on the recipient's reputation (image score) and ones own. The assessment norm determines how to evaluate an interaction between two other people as an observer. Most analytic calculations of indirect reciprocity assume binary image scores: the reputation of someone is either 'good' or 'bad'. Nobody so far has succeeded to formulate an exact analysis for the realistic situation where the image scores are more gradual and different people have different image scores of the same person as a consequence of private and incomplete information.

In order to derive a necessary condition for the evolution of cooperation by indirect reciprocity, let us study the interaction between the two basic strategies: (i) defectors and (ii) cooperators who cooperate unless they know the reputation of the other person to indicate a defector. The parameter q denotes the probability to know the reputation of another person. A cooperator always helps another cooperator. A cooperator helps a

3

defector with probability $1 - q$. Defectors never help. Hence, we obtain the payoff matrix

$$
\begin{array}{cc}
 & \begin{array}{cc} C & \quad\quad D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} b - c & -c(1 - q) \\ b(1 - q) & 0 \end{array} \right)
\end{array}
\tag{5}
$$

We have assumed that in a pairwise interaction both individuals are donor and recipient. If only one of them is donor and the other recipient, then all entries are multiplied by $1/2$, which makes no difference. Note that the payoff matrices (4) and (5) are identical (up to a factor) if we set $w = q$. Hence, indirect reciprocity leads to the same three conditions for ESS, RD and AD as direct reciprocity with q instead of w (see Table 1).

4. Network reciprocity

Spatial games can lead to cooperation in the absence of any strategic complexity (S11): unconditional cooperators can coexist with and sometimes outcompete unconditional defectors. This effect is called 'spatial reciprocity'. Spatial games are usually played on regular lattices such as square, triangular or hexagonal lattices. Network reciprocity is a generalization of spatial reciprocity to graphs. Individuals occupy the vertices of a graph. The edges denotes who interacts with whom. In principle, there can be two different graphs. The 'interaction graph' determines who plays with whom. The 'replacement graph' determines who competes with whom for reproduction, which can be genetic or cultural. Here we assume that the interaction and replacement graphs are identical. Evolutionary graph theory (S12) is a general approach to study the effect of population structure or social networks on evolutionary dynamics.

We consider a 'two coloring' of the graph: each vertex can be either a cooperator or a defector. A cooperator pays a cost, c, for each neighbor to receive a benefit, b. Defectors pay no cost and distribute no benefits. According to this simple rule the payoff, P, for each individual is evaluated. The fitness of an individual is given by $1 - \omega + \omega P$ where $\omega \in [0, 1]$ denotes the intensity of selection. Weak selection means that ω is much smaller than 1. Evolutionary updating works as follows: in each time step a random individual is chosen to die; the neighbors compete for the empty site proportional to their fitness.

We want to calculate the 'fixation probabilities', ρ_C, that a single cooperator starting in a random position on the graph takes over an entire population of defectors, and ρ_D, that a single defector starting in a random position on the graph takes over an entire population of cooperators. The fixation probability of a neutral mutant is $1/N$ where N is

4

the population size. If $\rho_C > 1/N$ then selection favors the fixation of cooperators; in this case cooperation is an advantageous strategy (AD).

For regular graphs, where each individual has exactly k neighbors, a calculation using pair-approximation (S13) leads to a surprisingly simple result: if $b/c > k$ then $\rho_C > 1/N > \rho_D$ for weak selection and large N. Numerical simulations show that this result is also an excellent approximation for non-regular graphs such as random graphs and scale free networks (S13).

The pair approximation calculation (for $k \geq 3$) also leads to a deterministic differential equation which describes how the expected frequency of cooperators (and defectors) changes over time (S14). This differential equation turns out to be a standard replicator equation (S15,S16) with a modified payoff matrix. For the interaction between cooperators and defectors on a graph with average degree k this modified payoff matrix is of the form

$$
\begin{array}{cc}
 & \begin{array}{cc} C & \quad D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} b-c & H-c \\ b-H & 0 \end{array} \right)
\end{array}
\tag{6}
$$

where

$$
H = \frac{(b-c)k - 2c}{(k+1)(k-2)}.
$$

It is easy to see that the payoff matrix (6) leads to the condition $b/c > k$ for cooperators to dominate defectors. In this case, cooperators are also ESS, RD and AD.

5. Group selection

Many models of group selection have been proposed over the years (S17-S29). It is difficult to formulate a model which is so simple that it can be studied analytically. One such model is the following (S30). A population is subdivided into m groups. The maximum size of a group is n. Individuals interact with others in the same group. Cooperators pay a cost c for each other member of the group to receive a benefit b. Defectors pay no costs and distribute no benefits. The fitness of an individual is $1 - \omega + \omega P$, where P is the payoff and ω the intensity of selection. At each time step, an individual from the entire population is chosen for reproduction proportional to fitness. The offspring is added to the same group. If the group reaches the maximum size, it can split into two groups with a certain probability, p. In this case, a randomly selected group dies to prevent the population from exploding. The maximum population size is mn. With probability $1 - p$ the group does not divide. In this case, a random individual of that group is chosen to die.

5

For small p, the fixation probability of a single cooperator in the entire population is given by the fixation probability of a single cooperator in a group times the fixation probability of that group.

For the fixation probability of one cooperator in a group of $n-1$ defectors we obtain $\phi_C = [1/n][1 - (b + cn - c)\omega/2]$. For the fixation probability of one cooperator group in a population of $m-1$ defector groups we obtain $\Phi_C = [1/m][1 + (b - c)(m - 1)\omega/2]$. Both results hold for weak selection (small ω). Note that the lower level selection within a group is frequency dependent and opposes cooperators, while the higher level selection between groups is constant and favors cooperators.

In the case of rare group splitting, the fixation probability of a single cooperator in the entire population, is given by the product $\rho_C = \phi_C \Phi_C$. It is easy to see that $\rho_C > 1/(nm)$ leads to $b/c > 1 + [n/(m - 2)]$. If this inequality holds, then cooperators are advantageous (and defectors disadvantageous) once both levels of selection are combined.

For a large number of groups, $m \gg 1$, we obtain the simplified condition $b/c > 1 + n/m$. The benefit to cost ratio of the altruistic act must exceed one plus the ratio of group size over the number of groups. The model can also be extended to include migration, which can be seen as 'noise' of group selection. In this case, the relevant criterion is $b/c > 1 + \mu + n/m$, where μ is the average number of migrants accepted over the life-time of a group (S30).

Now comes a surprising move that allows us to reduce the evolutionary dynamics on two levels of selection to a single two-person game on one level of selection. The payoff matrix that describes the interactions within a group is given by

$$
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \begin{pmatrix} b - c & -c \\ b & 0 \end{pmatrix}
\end{array}
\tag{7}
$$

Between groups there is no game dynamical interaction in our model, but groups divide at rates that are proportional to the average fitness of individuals in that group. Therefore one can say that cooperator groups have a constant payoff $b - c$, while defector groups have a constant payoff 0. Hence, in a sense the following 'game' between groups is happening

$$
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \begin{pmatrix} b - c & b - c \\ 0 & 0 \end{pmatrix}
\end{array}
\tag{8}
$$

Remember also that the 'fitness' of a group is $1 - \omega + \omega P$ where P is its 'payoff'. We can now multiply the first matrix by the group size, n, and the second matrix by the number

6

of groups, m, and add the two matrices. The result is

$$
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} &
\left(\begin{array}{cc}
(b-c)(n+m) & bm - c(m+n) \\
bn & 0
\end{array} \right)
\end{array}
\tag{9}
$$

In this simple 2×2 game, cooperators dominate defectors if $b/c > 1 + (n/m)$. In this case, cooperators are also ESS, RD and AD.

Interestingly, the method also gives the right answer for two arbitrary payoff matrices describing the games on the two levels. The intuition for adding the two matrices multiplied with the respective population size is as follows. For fixation of a new strategy in a homogeneous population using the other strategy, first the game dynamics within one group (of size n) have to be won and then the game dynamics between m groups have to be won. For weak selection and large m and n, the overall fixation probability is the same as the fixation probability in the single game using the combined matrix (9) and population size, mn. The stochastic process on two levels can be studied by a standard replicator equation using the combined matrix.

Finally, note that payoff matrix (9) for group selection is structurally identical to the payoff matrix (3) for kin selection if we set $r = m/(m + n)$ pointing to yet another interesting relationship between kin selection and group selection (S31).

References

S1. J. Maynard Smith, *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, UK, 1982).

S2. A. Grafen, *Anim. Behav.* **27**, 905 (1979).

S3. W. D. Hamilton, *J. Theor. Biol.* **7**, 1 (1964).

S4. L. L. Cavalli-Sforza, M. W. Feldman, *Theor. Popul. Biol.* **14**, 268 (1978).

S5. R. M. May, *Nature* **327**, 15 (1987).

S6. D. Fudenberg, E. Maskin, *Am. Econ. Rev.* **80**, 274 (1990).

S7. M. A. Nowak, K. Sigmund, *Acta Appl. Math.* **20**, 247 (1990).

S8. L. A. Imhof, D. Fudenberg, M. A. Nowak, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10797 (2005).

S9. D. Fudenberg, E. Maskin, *Econometrica* **50**, 533 (1986).

S10. K. Binmore, *Fun and Games* (Heath, Lexington, MA, 1992).

S11. M. A. Nowak, R. M. May, *Nature* **359**, 826 (1992).

S12. E. Lieberman, C. Hauert, M. A. Nowak, *Nature* **433**, 312 (2005).

S13. H. Ohtsuki, C. Hauert, E. Lieberman, M. A. Nowak, *Nature*, **441** (2006).

S14. H. Ohtsuki, M. A. Nowak, *J. Theor. Biol.* **243**, 86 (2006).

S15. P. D. Taylor, L. B. Jonker *Math. Biosci.* **40**, 145 (1978).

S16. J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, UK, 1998).

S17. B. R. Levin, W. L. Kilmer, *Evolution* **28**, 527 (1974).

S18. J. Maynard Smith, *Q. Rev. Biol.* **51**, 277 (1976).

S19. M. Slatkin, M. J. Wade, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3531 (1978).

S20. M. Uyenoyama, M. W. Feldman, *Theor. Popul. Biol.* **17**, 380 (1980).

S21. J. F. Crow, K. Aoki, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2628 (1982).

S22. R. G. Leigh, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2985 (1983).

S23. D. S. Wilson, *Annu. Rev. Ecol. Syst.* **14**, 159 (1983).

S24. L. Nunney, *Am. Nat.* **126**, 212 (1985).

S25. H. Harpending, A. Rogers, *J. Theor. Biol.* **127**, 51 (1986).

S26. E. Szathmáry, L. Demeter, *J. Theor. Biol.* **128**, 463 (1987).

S27. R. Boyd, P. J. Richerson, *J. Theor. Biol.* **215**, 287 (2002).

S28. B. Kerr, P. Godfrey-Smith, *Biol. Philos.* **17**, 477 (2002).

S29. J. Fletcher, M. Zwick, *J. Theor. Biol.* **228**, 303 (2004).

S30. A. Traulsen, M. A. Nowak, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10952 (2006).

S31. S. A. Frank, *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, NJ, 1998).