Five Rules for the Evolution of Cooperation

Martin A. Nowak

Authors

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.
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.

**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 \( f = (b - c)i/N \).

**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:

\[
    r > c/b
\]

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
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.

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:

\[ w > \frac{1}{b} \]  

**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:

$$ q > c/b $$

**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:

$$ b/c > k $$

**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,” 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

\[
1/n \ll 1 + (m/n)
\]  

(5)

**Evolutionary Success**

Before proceeding to a comparative analysis of the five mechanisms, let me introduce some measures of evolutionary success. Suppose a game between two strategies, cooperators \( C \) and defectors \( D \), is given by the payoff matrix

\[
\begin{array}{cc}
 C & D \\
 C & \alpha & \beta \\
 D & \gamma & \delta \\
\end{array}
\]

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
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 $a > \gamma$ 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 $2 \times 2$ payoff matrix (Table 1). From this matrix, we can derive the relevant condition for evolution of cooperation.
Each mechanism can be described by a simple $2 \times 2$ 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 $H = \frac{(b - c)k - 2c}{(k + 1)(k - 2)}$. All conditions can be expressed as the benefit-to-cost ratio exceeding a critical value. See (53) for further explanations of the underlying calculations.

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 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

<table>
<thead>
<tr>
<th>Kin selection</th>
<th>Payoff matrix</th>
<th>Cooperation is...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C$</td>
<td>$D$</td>
</tr>
<tr>
<td></td>
<td>$(b - c)k - 2c$</td>
<td>$b - c$</td>
</tr>
<tr>
<td>Direct reciprocity</td>
<td>$(b - c)k - 2c$</td>
<td>$b - c$</td>
</tr>
<tr>
<td>Indirect reciprocity</td>
<td>$(b - c)k - 2c$</td>
<td>$b - c$</td>
</tr>
<tr>
<td>Network reciprocity</td>
<td>$(b - c)k - 2c$</td>
<td>$b - c$</td>
</tr>
<tr>
<td>Group selection</td>
<td>$(b - c)k - 2c$</td>
<td>$b - c$</td>
</tr>
</tbody>
</table>

Table 1.
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 \( \times \) 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.

**Supporting Online Material**

[www.sciencemag.org/cgi/content/full/314/5805/1560/DC1](http://www.sciencemag.org/cgi/content/full/314/5805/1560/DC1)

**References and Notes**


2. A. Grafen, in *Oxford Surveys in Evolutionary Biology* vol. 2 R. Dawkins, M


53. See supporting material on Science Online.


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.

View Abstract