Chapter 22

Cooperation

J. MCKENZIE ALEXANDER

The Darwinian problem of cooperation is the following: according to the theory of natural selection, behaviors which serve to increase an individual's fitness will be favored over behaviors which decrease an individual's fitness; yet since cooperative behavior generally results in an individual's fitness being lower than what it could have been, had he or she acted otherwise, how is it that cooperative behavior persists? Natural selection, it would seem, should select against cooperative behavior – because of the reduced individual fitness – thereby driving it out of the population and promoting uncooperative behavior.

Closely related to the problem of cooperation is the problem of altruism, which was identified by E. O. Wilson as the "central theoretical problem of socio-biology" (Wilson, 1975, p.3). An altruistic behavior, in the evolutionary sense, causes the donor to incur a fitness cost while conferring a fitness benefit to the recipient (Sober & Wilson, 2000, p.185). According to these definitions, although altruistic behaviors are considered cooperative, the converse need not be true. If all individuals begin with a common baseline fitness and benefits are distributed equally, altruistic individuals have lower fitness than selfish individuals: an altruistic individual incurs both a personal fitness cost (due to his action) while receiving the common fitness benefit (from other altruists in the population), whereas a selfish individual only receives the common fitness benefit. The altruist's fitness is thus lower than what it could have been, had he acted otherwise, and is therefore a cooperative behavior. However, cooperative behavior need not be altruistic because it is possible for a co-operator to fail to maximize his or her individual fitness without incurring an explicit fitness cost. That is, altruistic behavior imposes explicit and actual fitness *penalties* upon individuals, whereas cooperative behavior requires only that the truth of a counterfactual obtain. In the following, this difference between altruistic and cooperative behavior will generally be suppressed.

Historically, attitudes regarding the extent to which evolution is compatible with cooperation have ranged between two extremes represented by Thomas Henry Huxley and Prince Petr Kropotkin in their writings on evolutionary theory in the nineteenth century. Huxley, arguing for the incompatibility of cooperative behavior and evolution, explicitly invoked Hobbesian imagery in his characterization of natural selection:

Cooperate Defect

	Cooperate	Defect
Cooperate	(R,R)	(S,T)
Defect	(T,S)	(P,P)

Figure 22.1 The Prisoner's Dilemma. Payoffs listed for (row, column), where values indicate relative changes in individual fitness, and $T > R > P > S$ and $\dfrac{T+S}{2} < R$

> the weakest and the stupidest went to the wall, while the toughest and the shrewdest, those who were best fitted to cope with their circumstances, but not the best in any other way, survived. Life was a continuous free fight, and . . . a war of each against all was the normal state of existence. (Huxley, 1888)

Kropotkin, on the other hand, noted how the structures produced by the social insects would have been impossible without a high degree of cooperation:

> The ants and the termites have renounced the "Hobbesian War" and they are the better for it. Their wonderful nests, their buildings superior in size relative to man . . . all of these are the normal outcome of the mutual aid which they practice at every stage of their busy and laborious lives. (Kropotkin, 1902)

The problem of cooperation is compelling because a great deal of cooperative and altruistic behavior clearly exists in nature. Female vampire bats (*Desmodus rotundus*) regurgitate blood obtained during successful feeding runs to other bats that have been less successful in obtaining food (Wilkinson, 1984). Such cooperation is essential to survival, since individual bats can starve to death in 60 hours without food. House sparrows (*Passer domesticus*) emit calls which attract other birds to newly discovered food sources (Summers-Smith, 1963). Indeed, extreme examples of altruistic behavior, such as the existence of sterile workers among the social insects, and the problem they posed for the theory of natural selection, were well known to Darwin. In *The Origin of Species*, he asked, "how is it possible to reconcile this case with the theory of natural selection?" (Darwin 1985 [1859], p.258). The apparent incompatibility, he proposed, "disappears, when it is remembered that selection may be applied to the family, as well as the individual, and may thus gain the desired end" (ibid.: p.258).

In general, the solution to the Darwinian problem of cooperation proceeds by identifying additional features of the evolutionary process which facilitate the emergence and persistence of cooperative behavior, the primary mechanisms being kin selection, reciprocity, and group selection. (One should note that the latter has engendered some controversy [SEE THE UNITS AND LEVELS OF SELECTION]). Additional mechanisms which have been identified include coercion, mutualism, by-product mutualism, and effects of local interactions.

The most commonly studied model of cooperation is the Prisoner's Dilemma, shown in Figure 22.1, originally developed by Merrill Flood and Melvin Dresher in 1950 while at the Rand Corporation for analyzing strategic conflict during the Cold War. The Prisoner's Dilemma encapsulates the strategic problem underlying the evolution of

cooperation produced when individual and collective interests conflict. In the Prisoner's Dilemma, achieving the collectively best outcome – the cooperative outcome – produces a suboptimal result from the point of view of the individual. In this model, each individual faces two courses of action, labeled "Cooperate" and "Defect." If both individuals cooperate, each has a fitness of R, the *reward*. If one individual cooperates and the other defects, the defector has the greatest possible fitness of T, the *temptation* for defecting, while the cooperator earns the lowest possible fitness of S, the *sucker's* payoff. If both individuals defect, each receives a fitness of P, the *punishment* for defecting, which is less than R. (The further condition that $(T+S)/2 < R$ is often imposed to insure that, in repeated interactions, cooperative behavior remains more beneficial than alternation of cooperate and defect.) With these particular fitness payoffs, it would seem that natural selection should favor Defect, since it maximizes one's own fitness independent of the behavior of others.

1. Kin Selection

After the modern synthesis, another solution to the problem of cooperation became available. The gene-centered view of evolution (see Dawkins, 1976) recognized that, since it is ultimately genes which are passed from parent to offspring, and individual organisms share portions of their genetic material with other members of the same species, natural selection may favor behaviors that successfully promote the propagation of an individual's genes even if that behavior reduces the number of viable offspring an organism has. [SEE THE UNITS AND LEVELS OF SELECTION]. This view was first given a precise formulation and analysis by Hamilton (1964), who introduced the concept of *inclusive* fitness, which can be thought of as the number of an individual's alleles present in the next generation rather than the actual number of viable offspring of an individual. More precisely, inclusive fitness is the relative representation, in the next generation, of an individual's genes in the overall gene pool. Kin selection is the process of selection which increases the inclusive fitness of the individual.

The theoretical result underlying kin selection is Hamilton's rule, which states that a gene possessed by an individual i increases in frequency whenever $\sum_{j=1}^{n} r_{ij}b_{ij} - c > 0$, where n is the number of individuals affected by the trait the gene encodes, r_{ij} denotes the degree of relatedness between individuals i and j, b_{ij} the benefit conferred by i to j, and c is the associated cost to i of bearing the trait. (The degree of relatedness of two individuals is a real number between 0 and 1 indicating the proportion of genes held in common between the two individuals.) According to Hamilton's rule, cooperative or altruistic acts can evolve provided that the cost/benefit ratio of the act is less than the degree of relatedness between the affected individuals. For example, evolution would favor one sibling sacrificing all of his fitness to help his brother (a degree of relatedness of 0.5) provided that the altruist's act increases his brother's fitness by at least twofold.

A common misinterpretation of Hamilton's rule is that it says organisms are expected to act altruistically toward relatives according to the degree that they are related. (This

mistake was made by Dawkins in the first edition of *The Selfish Gene*, corrected in the endnotes to the second edition.) Hamilton's rule states a condition under which altruistic or cooperative behavior toward relatives can evolve; it does not say that evolution is expected to produce an array of behaviors which distribute altruism accordingly across one's relatives.

Hamilton's work, and the idea of kin selection in general, has had great impact upon the field of evolutionary biology for two reasons. The first is that it seemed to provide a more parsimonious account of the evolution of cooperation than Darwin's preferred explanation of group selection. The second is that it provided a theoretical explanation for the haplo-diploid sex determination and eusociality of the social insects. Whereas in most animals sex differentiation occurs through the possession of a different set of sex chromosomes (a heterogametic and homogametic sex), among the social hymenopterans males develop from haploid (unfertilized) eggs and females from diploid (fertilized) eggs. This system of genetic determination of the sexes modifies the degrees of relatedness in such a way so as to strongly favor eusociality. Indeed, eusociality has independently evolved among the social insects no fewer than eleven times.

2. Reciprocity

While kin selection can account for the evolution of cooperation among genetic relatives, it cannot account for the evolution of cooperation among individuals who are not genetically related. Reciprocal altruism, first introduced in an influential paper by Trivers (1971), provides a mechanism through which altruistic or cooperative behavior can evolve even when the individuals who engage in altruistic behavior are not genetically related to one another. Reciprocal altruism is found in a variety of natural environments. Commonly cited examples of this phenomenon include mutual symbioses such as ants and ant-acacias, where the trees provide housing for the ants which, in turn, provide protection for the trees (Janzen, 1966); figs trees and fig-wasps, where the wasps are parasites on the fig flowers but provide the fig trees' method of pollination (Wiebes, 1976; Janzen, 1979); and cleaning symbioses, discussed at length in Trivers' original article. Reciprocal altruism is a robust phenomenon, having independently evolved many times (Trivers notes that it has arisen independently at least three times among shrimp alone).

In Trivers' original model, what promoted the flourishing of cooperative behavior in reciprocal interactions was a common threat from the environment which all faced; engaging in altruistic behavior served to reduce the environmental threat sufficiently so as to be worth each person's incurring the fitness cost imposed by altruistic action. For example, consider the act of saving someone from drowning. Suppose that the probability of dying from drowning is 50 percent if no one attempts a rescue, and that the probability of the rescuer drowning is 5 percent. In addition, assume that the drowning person always dies if his rescuer drowns and the drowning person is always saved if the rescuer does not drown (which is taken to mean that the rescue attempt was successful). If interactions between the drowning person and rescuer were never repeated, then there would be no reason for anyone to attempt to rescue a drowning person. However, if interactions are repeated, so that an individual who was saved from

drowning can reciprocate and come to the aid of his rescuer at a later point in time, it is in the interest of each to come to the aid of the other. If every person in the population has the same risk of drowning, people who come to the aid of the other will have, in effect, reduced the original 50 percent chance of dying to only a 10 percent chance. While reduction of risk posed by common threats provides a particularly striking example of the contexts in which reciprocal altruism can arise, the phenomenon is much more widespread, as the examples of mutual symbioses indicate.

Perhaps the most well-known (if somewhat overstated, see Binmore, 1998) example of the evolution of cooperation through reciprocity is the success of Tit-for-Tat in the repeated Prisoner's Dilemma (Axelrod, 1984). Axelrod conducted a computer tournament in which sixty strategies, solicited from many different individuals, were pitted against each other in a "round-robin" competition. Each strategy played five runs of the repeated Prisoner's Dilemma against every other strategy. Each run consisted of the Prisoner's Dilemma being repeated a certain number of times, where the number of repeats was fixed in advance, and common among all strategy pairings.

What Axelrod found, both in the original computer tournament and in a second, larger, tournament held later, was that a very simple strategy favoring cooperative behavior won both tournaments. The strategy, known as Tit-for-Tat, begins by cooperating and then simply mimics the previous play of its opponent in all rounds after the first. If its opponent always cooperates, then Tit-for-Tat will always cooperate. If its opponent defects in the nth stage of the game, Tit-for-Tat will reciprocate by defecting in the $n + 1$st stage of the game; if its opponent should then "apologize" for its nth stage defection with cooperative behavior in the $n + 1$st stage, Tit-for-Tat will accept the apology by cooperating in the $n + 2$nd stage. The simple feedback mechanism employed by Tit-for-Tat is, Axelrod found, remarkably successful at rewarding cooperative behavior and punishing defections in certain environments.

In addition, when Axelrod took the initial strategies and performed an "ecological analysis," modeling a dynamic environment in which more successful strategies became more prolific, Tit-for-Tat still won. This simulation proceeded as follows: initially, each of the submitted strategies was considered to be equally likely in the population. The results from the tournament were assembled into a large payoff matrix specifying how well each strategy did when paired against every other strategy. This matrix was then used to calculate the expected fitness of each strategy in the population, which in the first generation simply equaled the actual fitness earned by each strategy at the end of the original tournament. However, after the first generation, the frequency of each strategy in the population was adjusted according to how well it did at the end of the current generation. From this point on, the expected fitness of each strategy in the population need not necessarily agree with the fitness of each strategy in the original tournament. Even so, within two hundred generations Tit-for-Tat became the most frequently used strategy in the population.

Axelrod identified four beneficial properties of Tit-for-Tat that enabled it to be successful: (1) it was not envious, (2) it was not the first to defect, (3) it reciprocated both cooperation and defection, and (4) it was not too clever (Tit-for-Tat outperformed a strategy which modeled the actions of its opponent as a Markov process, then using Bayesian inference to select which move – Cooperate or Defect – was deemed most likely to maximize its payoff in the next round). He also claimed to provide necessary and

sufficient conditions for the collective stability of Tit-for-Tat, where "collectively stable" means that if everyone in the population follows it, no alternative strategy can invade (Axelrod, 1984, p.56). The precise result Axelrod proves is the following proposition:

> Proposition 2. Tit-for-Tat is collectively stable if and only if *w* is large enough. This critical value of *w* is a function of the four payoff parameters *T, R, P*, and *S*. (Axelrod, 1984, p.59)

The parameter *w* denotes the probability that both individuals will have another round of interaction in the future, and the critical value which makes Tit-for-Tat collectively stable is $\max\left\{\dfrac{T-R}{T-P}, \dfrac{T-R}{R-S}\right\}$.

Unfortunately, Tit-for-Tat's success in Axelrod's tournaments has led some to regard it as *the* solution to the Darwinian problem of cooperation, or as *the* optimal behavior to adopt in the repeated prisoner's dilemma. Tit-for-Tat is not optimal – indeed, it can be proven that in the indefinitely repeated Prisoner's Dilemma no optimal strategy exists. Axelrod himself noted that Tit-for-Tat would not have won the two computer tournaments if two other "natural" competitors had been submitted. One competitor which would have beat Tit-for-Tat is Win–stay, lose–shift (also known as "Pavlov"). Win–stay, lose–shift, like Tit-for-Tat, begins by cooperating on the first move, and then cooperates on future moves if and only if both players adopted the same strategy on the previous move. Suppose that the first individual follows the strategy Win–stay, lose–shift. If both cooperate, he will continue to cooperate on the next move as mutual cooperation is considered to be a "win" and the strategy recommends staying with a win. If both defect, he will switch to cooperating on the next move: mutual defection is considered to be a "loss," so he adopts the other alternative for the next move, which is cooperation. If the first individual defects and the second cooperates, the first individual will continue to defect on the next move, as defection against a cooperator is considered to be a "win." If the first individual cooperates and the second defects, he will switch to defection on the next move, as cooperating against a defector is a "loss," so he switches to the other alternative for the next move, which is in this case defection (Nowak & Sigmund, 1993).

Aside from the fact that Tit-for-Tat would have been beaten in the original tournament by only a marginally simpler strategy, which also does well on the four criteria identified by Axelrod, many other shortcomings of Axelrod's analysis have been identified (Binmore, 1998). Perhaps the most important one is that Tit-for-Tat is not actually immune to being invaded by competing strategies, contrary to Axelrod's claim that it is collectively stable. Lindren and Nordahl (1994) show how, in a model of the infinitely iterated Prisoner's Dilemma with noise and a strategy space which is not bounded in memory length (Tit-for-Tat only has a memory of 1), Tit-for-Tat can be invaded by a variety of other strategies.

Reciprocity promotes cooperation effectively by transforming the structure of the problem from the Prisoner's Dilemma into a different one. Consider what happens in the case where Tit-for-Tat plays against All Defect with the abovementioned payoffs and a probability of future interactions given by *w*. When Tit-for-Tat plays against Tit-for-Tat, it always cooperates, so the payoffs for the indefinitely iterated interaction are

	Cooperate	Defect
Cooperate	(3,3)	(0,5)
Defect	(5,0)	(1,1)

	Tit-for-Tat	All Defect
Tit-for-Tat	(9,9)	(2,7)
All Defect	(7,2)	(3,3)

Figure 22.2 Reciprocity changes the Prisoner's Dilemma into an Assurance Game. Payoffs listed for (row, column), and $w = \dfrac{2}{3}$

$$W(\text{TfT}|\text{TfT}) = R + Rw + Rw^2 + Rw^3 + \ldots$$
$$\sum_{i=0}^{\infty} Rw^i = \frac{R}{1-w}.$$

Likewise, the payoffs for the other three possible pairings of Tit-for-Tat and All Defect are as follows:

$$W(\text{TfT}|\text{AllD}) = S + Pw + Pw^2 + Pw^3 + \ldots$$
$$S + \sum_{i=1}^{\infty} Pw^i = S + \frac{Pw}{1-w}.$$
$$W(\text{AllD}|\text{AllD}) = P + Pw + Pw^2 + Pw^3 + \ldots$$
$$\sum_{i=0}^{\infty} Pw^i = \frac{P}{1-w}.$$
$$W(\text{AllD}|\text{TfT}) = T + Pw + Pw^2 + Pw^3 + \ldots$$
$$T + \sum_{i=1}^{\infty} Pw^i = T + \frac{Pw}{1-w}.$$

If the probability of future interactions is sufficiently high, the payoff matrix for choosing between reciprocating cooperative behavior and always defecting becomes that shown in Figure 22.2. Reciprocity can transform the Prisoner's Dilemma into an Assurance Game, or Stag Hunt (Skyrms, 2004).

3. Group Selection

Although the possibility that cooperative behavior might originate through selection acting on levels higher than the individual was first put forward by Darwin in *The Origin of Species*, group selection fell into disrepute when Williams (1966) argued that most alleged instances of group selection could be understood in individualist terms. In

recent years, though, Wilson (1980) and Wade (1978) have sought to rehabilitate theories of group selection, arguing for multilevel selection theory. Sober and Wilson (2000) show how group selection can support the emergence and persistence of cooperative behavior under certain conditions.

Whether group selection supports cooperation depends crucially on details of the selection process. For example, Maynard Smith's (1964) "haystack model" of group selection does not support the emergence of cooperation. In this model, field mice live in haystacks, where each haystack is initially populated by a single fertilized female. Each female gives birth in the haystack, which remains populated for several generations. At the end of the first generation, brothers and sisters from the original founding female mate with each other; at the end of the second generation, first cousins mate with first cousins, and so on. After a certain number of generations, all of the haystacks empty, mice mate with randomly chosen partners, and then each fertilized female goes on to found another colony in a new haystack, repeating the process described above. Maynard Smith showed that, under these conditions, cooperation tends to be driven to extinction.

Sober and Wilson's (1998, 2000) model of group selection modifies the process through which groups form. Unlike Maynard Smith's model, where each group (haystack) is initially occupied by a single pregnant female, in the Sober and Wilson model, groups periodically merge into a larger population and re-form by a partitioning of that population into smaller groups. This change, along with the fact that groups may include more than one cooperator at the time of formation, enables cooperation to emerge.

More precisely, suppose that cooperators incur a fitness cost of c and that individuals who receive the benefit of cooperation have their fitness increased by b. In addition, suppose each individual has a baseline fitness of X. If there are n individuals in the group, with p of them being cooperators, then the fitness of a cooperator is

$$W_C = X - c + \frac{b(np-1)}{n-1}$$

since each cooperator has his baseline fitness reduced by c and may possibly receive a benefit from any one of the $np - 1$ other altruists in the group. (The expression $\frac{b(np-1)}{n-1}$ denotes the expected benefit of each altruist in the group.) The fitness of a defector is simply

$$W_D = X + \frac{bnp}{n-1}$$

which exceeds the fitness of a cooperator for two reasons: first, the defector does not incur the fitness cost of cooperating; second, a defector is eligible to receive a benefit from any one of the np cooperators in the group, whereas a cooperator is eligible to receive a benefit from only $np - 1$ cooperators (it is assumed that cooperators cannot bestow benefits to themselves).

Now, suppose we have an initial population consisting of 200 individuals, in which exactly half of the population cooperate. Suppose further that the population divides into two groups of equal size, with the first group containing 20 percent cooperators and the second group contains 80 percent cooperators. The fitness of cooperators and defectors in the first group is then

$$W_C^1 = 10 - 1 + \frac{5(20-1)}{99} = 9.96$$
$$W_D^1 = 10 + \frac{5(20)}{99} = 11.01$$

and the fitness of cooperators and defectors in the second group is

$$W_C^2 = 10 - 1 + \frac{5(80-1)}{99} = 12.99$$
$$W_D^2 = 10 + \frac{5(80)}{99} = 14.04.$$

In both groups, cooperators have lower fitness than defectors, as one would expect given the basic structure of the Prisoner's Dilemma. After reproduction, group one increases in size from 100 to 1,080, with cooperators accounting for only 18.4 percent of the total, and group two increases in size from 100 to 1,320, with cooperators accounting for 78.7 percent of the total. In both groups, the frequency of cooperation has decreased.

However, considering the population as a whole, the total frequency of cooperation has increased. Initially we started with only 200 individuals and a frequency of cooperation of 50 percent. After the first generation, the total population size is 2,400 with the frequency of cooperation being $\frac{0.184 \cdot 1080 + 0.787 \cdot 1320}{2400} = 0.516$. The fact that the frequency of cooperation can decrease in each group individually while increasing in the overall population is an example of Simpson's paradox (see Simpson, 1951; Sober, 1984; and Cartwright, 1978).

4. Coercion

According to coercive theories of cooperation, individuals are coerced into cooperative or altruistic acts by dominant members of the population and face the threat of ejection if they do not comply. Although there is some evidence of coercion in cooperative societies of fish (Balshine-Earn et al., 1998), fairy wrens (Mulder & Langmore, 1993), and naked mole rats (Reeve, 1992), it seems that the majority of forms of cooperation are not coerced.

Closely related to coercive theories of cooperation are retributive theories (Boyd & Richardson, 1992). In this model, groups of size n are formed by random sampling from a large population. Within each group, individuals interact in two stages: the first being

a cooperative stage where individuals have a choice of either cooperating or defecting (as in the Prisoner's Dilemma), the second being a punishment stage where individuals can punish any member in the group. Boyd and Richardson find that, under certain conditions, retribution-based processes facilitate cooperation in larger groups than is possible with mere reciprocity-based processes. Retribution-based processes can also be a powerful selective and stabilizing force since "moralistic" behaviors, which punish individuals who do not comply with the required behavior, are capable of rendering *any* individually costly behavior evolutionarily stable.

5. Mutualism

For certain animals, the fitness of individual group members tends to increase with group size (Courchamp, Clutton-Brock, & Grenfell, 2000). Mutualist explanations of cooperative behavior point to correlations between group size/success and individual fitness, which thereby reduce the expected gain to individuals by defecting. Kokko, Johnstone, and Clutton-Brock (2001) identify several processes which lead to the creation of these correlations. For example, when greater group size/success leads to greater feeding success in adults, increased success in defending food supplies from competitors, greater efficiency in defending and providing for young, and so on, cooperative group behavior need not be eliminated by defection. While some of the evidence linking group size/success with individual fitness need not differentiate between mutualism and reciprocity, such as when unrelated group members contribute to the common good (Cockburn, 1998), cases where groups accept unrelated immigrants (Piper, Parker, & Rabenold, 1995) or kidnap individuals from other groups (Heinsohn, 1991) seem to favor mutualist accounts over reciprocal altruism.

6. Byproduct Mutualism

Byproduct mutualism occurs when the cooperative behavior benefiting the group coincides with the behavior that maximizes individual fitness. In these cases, the production of beneficial consequences for others through cooperative behavior might be entirely coincidental (Bednekoff, 1997). Note that byproduct mutualism therefore concerns instances of cooperation where the fitness payoffs do not conform to the basic structure of the Prisoner's Dilemma. Hence, there is some question as to whether the behavior deserves the label of "cooperative" in the first place.

Brown (1983) introduced byproduct mutualism by noting that "in many cases of mutualism, CC > DC will be found to prevail rather than DC > CC as required by the prisoner's dilemma." Contrary to the DC > CC > DD > CD ordering of payoffs for the Prisoner's Dilemma, a more likely ordering for species where cooperative activities are more profitable in groups than alone would be "CC > CD > DC = DC" (Brown, 1983, p. 30). Figure 22.3 illustrates the payoff matrix for cooperative behavior generated in the context of byproduct mutualism. The structure of the payoff matrix is that of a coordination game, where the choice to Cooperate dominates Defect.

	Cooperate	Defect
Cooperate	(x,x)	(y,w)
Defect	(w,y)	(z,z)

Figure 22.3 The payoff matrix for cooperative behavior generated through byproduct mutualism. Payoffs listed for (row, column), where values indicate relative changes in individual fitness, and $x > y > w \geq z$

Although it is easy to see why natural selection would favor "cooperative" behavior in these instances, part of the interest in byproduct mutualism derives from the fact that, in the study of the evolution of cooperation, it is difficult to determine the payoffs for the acts of Cooperate and Defect. When uncertainty exists as to what the payoffs are, it is an open question as to which payoff matrix best describes the interactive problem. Some experiments with bluejays (Clements & Stephens, 1995) suggest that the observed cooperative behavior is better explained as a result of byproduct mutualism than alternative mechanisms.

7. Local Interactions

Large, panmictic populations that reproduce asexually do not favor the formation of cooperative behavior. One well-known model of this is the *replicator dynamics* by Taylor and Jonker (1978). Suppose we have a large population, where each agent has a certain phenotype σ. For simplicity, assume that there are only finitely many phenotypes $\sigma_1, \ldots, \sigma_m$. Let n_i denote the total number of agents in the population with the phenotype σ_i, with the total size of the population given by $N = \sum_{i=1}^{m} n_i$. For large, panmictic populations, all of the relevant information about the population is contained in the state vector $\vec{s} = \langle s_1, \ldots, s_m \rangle$, where $s_i = \dfrac{n_i}{N}$ for all i. If the growth rate of the ith phenotype approximately equals the fitness of that phenotype in the population, one can show that the rate of change of the ith phenotype is given by

$$\frac{ds_i}{dt} = s_i (W(i|\vec{s}) - W(\vec{s}|\vec{s}))$$

where $W(i|\vec{s})$ denotes the mean fitness of i in the population and $W(\vec{s}|\vec{s})$ denotes the mean fitness of the population at large. This is continuous replicator dynamics, which assumes that the increase or decrease of the phenotype frequencies occurs without well-defined generational breaks; that is, it assumes there is not a well-defined notion of "next generation" applying to the population (such as biological reproduction in humans).

425

In a population where p individuals Cooperate and $1 - p$ Defect, the expected fitness of Cooperate and Defect are, respectively,

$$W(C|\vec{s}) = p \cdot W(C|C) + (1-p) \cdot W(C|D)$$

and

$$W(D|\vec{s}) = p \cdot W(D|C) + (1-p) \cdot W(D|D).$$

Since $T > R$ and $P > S$, the expected utility of defecting is greater than the expected reward of cooperating, so it follows that $W(D|\vec{s}) > W(\vec{s}|\vec{s}) > W(C|\vec{s})$. From this, it follows that,

$$\frac{ds_D}{dt} = (1-p)(W(D|\vec{s}) - W(\vec{s}|\vec{s})) > 0$$

and

$$\frac{ds_C}{dt} = p(W(C|\vec{s}) - W(\vec{s}|\vec{s})) < 0.$$

Over time, the proportion of the population not defecting will eventually be driven to extinction.

However, if spatial location constrains interaction between individuals, cooperation may emerge. Nowak and May (1992, 1993) show that the spatialized Prisoner's Dilemma favors the evolution of cooperation provided that the fitness payoffs for cooperation lie in a certain range and that there are a certain number of cooperators initially present. In their model, organisms are positioned at fixed locations on a square lattice and interact with their eight nearest neighbors. (In the original paper, all locations on the lattice are occupied and the lattice is considered to wrap at the edges. Although the former assumption is important for their results, the latter is not.) All individuals interact simultaneously and receive a total fitness payoff equaling the sum of all eight interactions. After interacting, behaviors are replicated according to the following rule: if an organism's fitness is lower than the fitness of at least one of his neighbors, that organism will be replaced in the next generation by an offspring from his neighbor who has the highest fitness. (If several neighbors are tied for having the highest fitness, then the neighbor whose offspring replaces the unfit individual is chosen at random.) If an organism's fitness is higher than the fitness of all of his neighbors, that organism's offspring will occupy the same site in the lattice for the next generation.

There are three possible outcomes: cooperation and defection may coexist in stable oscillating patterns, defection may drive cooperation to extinction, or cooperation and defection may coexist in chaotic patterns of mutual territorial invasion. Figures 22.4, 22.5, and 22.6 illustrate each of these possibilities in turn. In Figure 22.4, the case of stable coexistence, the fitness values are $T = 1.1$, $R = 1$, $P = 0$ and $S = -0.1$. In figure 22.5,
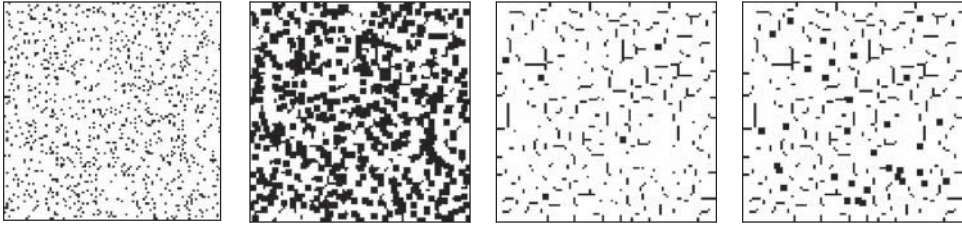
Figure 22.4 The spatial prisoner's dilemma illustrating the evolution of stable cooperative regions. $T = 1.1$, $R = 1$, $P = 0$ and $S = -0.1$
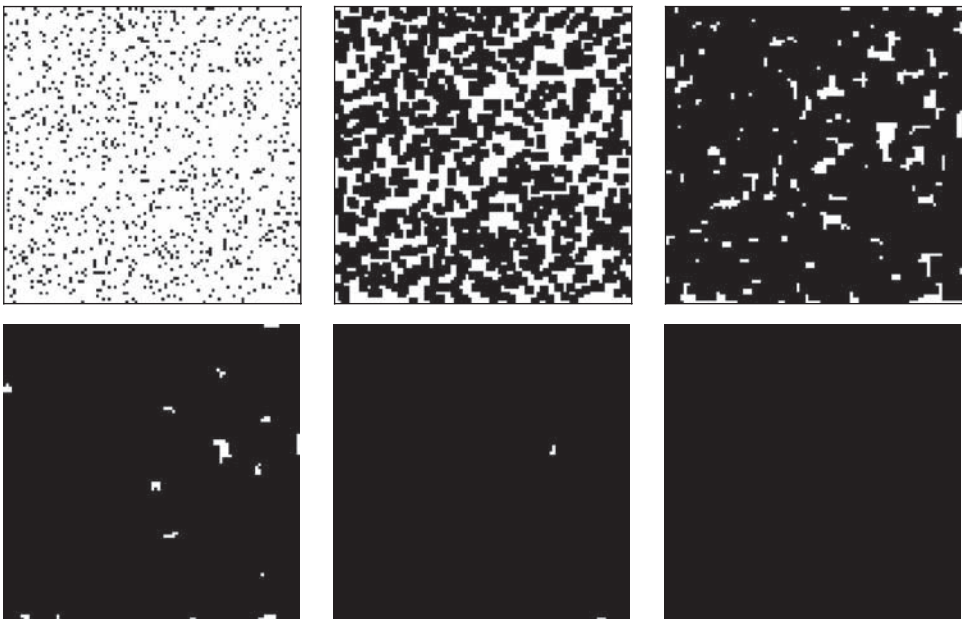

Figure 22.5 The spatial prisoner's dilemma illustrating the evolution of stable cooperative regions. $T = 2.7$, $R = 1$, $P = 0$, $S = -0.1$

with fitness values of $T = 2.7$, $R = 1$, $P = 0$, $S = -0.1$, defectors come to dominate within a relatively short period of time. (Note, though, that these particular fitness values violate the requirement that $(T + S)/2 < R$.) Of particular interest is Figure 22.6, which uses payoff values of $T = 1.6$, $R = 1$, $P = 0$, $S = -0.1$. In this case, the mix of cooperators and defectors in the population fluctuates chaotically. Cooperative regions can be invaded by regions of defectors, and vice versa, without ever settling into a stable evolutionary state.
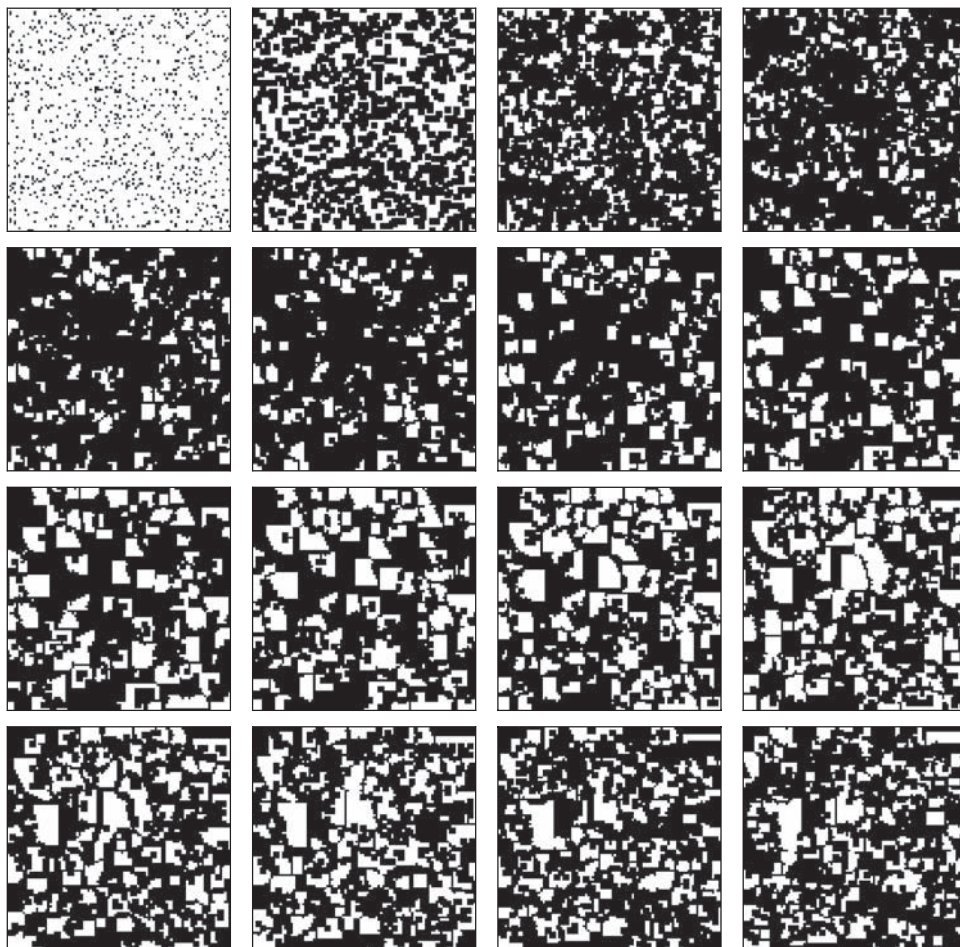
427

Figure 22.6   The Spatial Prisoner's Dilemma, $T = 1.6$, $R = 1$, $P = 0$, $S = -0.1$

References

Axelrod, R. (1982). *The evolution of cooperation*. New York: Basic Books.

Balshine-Earn, S., Neat, F. C., Reid, H., & Taborsky, M. (1998). Paying to stay or paying to breed? Field evidence for direct benefits of helping behavior in a cooperatively breeding fish. *Behavioral Ecology*, 9, 432–8.

Bednekoff, P. A. (1997). Mutualism among safe, selfish sentinels: A dynamic game. *The American Naturalist*, 150, 373–92.

Binmore, K. (1998). Review of "The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration". *Journal of Artificial Societies and Social Simulation*, 1(1). http://jasss.soc.surrey.ac.uk/1/1/review1.html.

Boyd, R., & Richardson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizeable groups. *Ethology and Sociobiology*, 13, 171–95.

Brown, J. L. (1983). Cooperation: a biologist's dilemma. In J. S. Rosenblatt (Ed.). *Advances in the study of behavior* (pp. 1–37). New York: Academic Press.

Cartwright, N. (1978). Causal laws and effective strategies. *Noûs*, 13, 419–37.

Clements, R., & Stephens, D. C. (1995). Testing models of non-kin cooperation-mutualism and the prisoner's dilemma. *Animal Behaviour*, 50(2), 527–35.

Cockburn, A. (1998). Evolution of helping behavior in cooperatively breeding birds. *Annual Review of Ecology and Systematics*, 29, 141–77.

Courchamp, F., Clutton-Brock, T., & Grenfell, B. (2000). Multipack dynamics and the Allee effect in African wild dogs *Lycaon pictus. Animal Conservation*, 3, 277–86.

Darwin, C. (1985 [1865]). *The origin of species*. London: Penguin.

Dawkins, R. (1976). *The selfish gene*. London: Oxford University Press. Second edition, December 1989.

Hamilton, W. D. (1964). The genetical evolution of social behavior. I and II. *Journal of Theoretical Biology*, 7, 1–52.

Heinsohn, R. G. (1991). Kidnapping and reciprocity in cooperatively breeding white-winged choughs. *Animal Behavior*, 41, 1097–100.

Huxley, T. H. (1888). The struggle for existence in human society. In *Collected Essays* (Vol. 9, pp. 195–236). London: Macmillan.

Janzen, D. H. (1966). Coevolution of mutualism between ants and acacias in Central America. *Evolution*, 20, 249–75.

Janzen, D. H. (1979). How to be a fig. *Annual Review of Ecology and Systematics*, 10, 13–51.

Kokko, H., Johnstone, R. A., & Clutton-Brock, T. H. (2001). The evolution of cooperative breeding through group augmentation. *Proceedings of the Royal Society of London, B*, 268, 187–96.

Kropotkin, P. (1902). *Mutual aid*. London: William Heinemann.

Lindgren, K., & Nordahl, M. (1994). Evolutionary dynamics of spatial games. *Physica*, D, 75, 292–309.

Maynard Smith, J. (1964). Group selection and kin selection. *Nature*, 200, 1145–7.

Mulder, R. A., & Langmore, N. E. (1993). Dominant males punish helpers for temporary defection in superb fairy-wrens. *Animal Behavior*, 45, 830–3.

Nowak, M., & May, R. (1992). Evolutionary games and spatial chaos. *Nature*, 359, 826–9.

Nowak, M., & May, R. (1993). The spatial dilemmas of evolution. *International Journal of Bifurcation and Chaos*, 3, 35–78.

Nowak, M., & Sigmund, K. (1993). A strategy of win–stay, lose–shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364, 56–8.

Piper, W. H., Parker, P., & Rabenold, K. N. (1995). Facultative dispersal by juvenile males in the cooperative stripe-backed wren. *Behavioral Ecology*, 6, 337–42.

Reeve, H. K. (1992). Queen activation of lazy workers in colonies of the eusocial naked mole rat. *Nature*, 358, 147–9.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society B*, 13, 238–41.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. London: Cambridge University Press.

Sober, E. (1984). *The nature of selection: evolutionary theory in philosophical focus*. Cambridge, MA: MIT Press.

Sober, E., & Wilson, D. S. (1998). *Unto others: the evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Sober, E., & Wilson, D. S. (2000). Summary of "Unto Others the Evolution and Psychology of Unselfish Behavior". In L. D. Katz (Ed.). *Evolutionary origins of morality: cross disciplinary perspectives* (pp.185–206). Bowling Green: Imprint Academic.

Summers-Smith, J. D. (1963). *The house sparrow*. London: Collins.

Taylor, P. D., & Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40, 145–56.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Journal of Biology*, 46, 189–226.

Wade, M. J. (1978). A critical review of the models of group selection. *Quarterly Review of Biology*, 53, 101–14.

Wiebes, J. T. (1976). A short history of fig wasp research. *Gardens' Bulletin*, Singapore 29, 207–32.

Wilkinson, G. (1984). Reciprocal food sharing in vampire bats. *Nature*, 308, 181–4.

Williams, G. C. (1966). *Adaptation and natural selection: a critique of some current evolutionary thought*. Princeton: Princeton University Press.

Wilson, D. S. (1980). *The natural selection of populations and communities*. Menlo Park, CA: Benjamin Cummings.

Wilson, E. O. (1975). *Sociobiology: the new synthesis*. Cambridge, MA: Harvard University Press.