TEXT SUPPLEMENT (Ch 8): Reciprocal Altruism

The theory of inclusive fitness, described in the text, is only one possible explanation for the evolution of altruism. Another completely different explanation for altruism is based on a concept called reciprocal altruism, which is the idea that if you have behaved kindly towards an individual in the past, that individual will be inclined to perform altruistic acts that benefit you. It is not at all necessary that the altruist expects anything in return; it is only important that altruists do generally benefit in the long run from reciprocal acts of kindness.

Researchers seeking to understand the evolution of reciprocal altruism have turned to mathematical models based on game theory. As with other mathematical models, game theory models have the form of an elaborate hypothesis—that the participants, viewed as players in a game, behave in the manner that the model predicts. In the mathematical theory of games, the underlying hypothesis is that each player makes certain choices that maximize their chances of achieving the highest possible score. The guidelines by which the choices are made are called a strategy, and the strategy that maximizes the chances of winning (or of getting the highest score) is considered a winning strategy. For example, in one of the most well known of these models, called prisoner's dilemma, two captured individuals find out that if they both defect to the police (confessing and also implicating one another), they will both receive jail sentences. If neither confesses, the prisoners hope that they might be released for lack of evidence. However, they are told, if only one of them defects, then the defector will receive a medal, while the other will be shot. The dilemma is that since neither knows what the other will do, the only way to avoid the worst outcome is to defect.

In the biological applications of game theory, different individuals (or their genotypes) are compared to players, and the biological concept of fitness is used to keep score. Thus any strategy that increases an individual's fitness will result in their leaving more offspring, and the strategy will thus spread more widely throughout the population.

Robert Axelrod and W.D. Hamilton used mathematical game theory to determine what would happen in a population in which many individuals were playing prisoner's dilemma repeatedly against one another. They tested a variety of gaming strategies, allowing the more successful individuals to occasionally 'reproduce' others like themselves and the least successful ones to die. Over time, certain strategies became more common, while others diminished and died out. In these models, any strategy that can successfully spread (increase in frequency) when it is initially rare is known as an evolutionarily stable strategy (ESS). Such strategies are stable because they increase fitness and thus resist the spread of alternative strategies.

By testing a variety of strategies, Axelrod and Hamilton were able to show that a strategy called TIT FOR TAT is the most successful strategy among those tested, that is, it increased fitness to the greatest extent. The TIT FOR TAT strategy is to act kindly toward any other individual who acted kindly toward you on your last encounter, or towards newcomers with no previous track record of behavior, but to act nasty toward anyone who was nasty to you the last time. If other members of the population practice the same TIT FOR TAT strategy, then individuals using this strategy will become reciprocal altruists towards one another, and they will all benefit from the altruism. The benefits of reciprocal altruism ensure that its frequency will increase in the population, even if it is initially rare. In other words, TIT FOR TAT is an evolutionarily stable strategy in this model.

In the 1990s, several researchers discovered that another strategy, which some have nicknamed PAVLOV, can increase fitness even more than TIT FOR TAT. A simple description of the PAVLOV strategy is, "If you were successful last time, repeat whatever you did; if not, switch your behavior." As a strategy, PAVLOV resembles TIT FOR TAT in encouraging reciprocal altruism, but it differs from TIT FOR TAT in more frequently taking advantage of other individuals, called 'suckers,' who are willingly victimized or exploited. The superiority of the PAVLOV strategy over TIT FOR TAT thus depends on the presence of 'suckers' in the population. Observations of natural populations have confirmed that PAVLOV is in fact a widely used strategy.

Reciprocal altruism begets deception: if I can take advantage of you, but make you think that I have acted kindly, then I receive *both* the direct advantage of my action and also the future benefit of your acting to return the imagined kindness. Among mammals especially, reciprocal altruism selects both for deception and for the ability to detect the deception or true motives of others. Often, it is the higher-ranked (more dominant) individuals that make a career out of deceiving others, while lower-ranked (more subordinate) individuals, including many females, become skilled at discerning the true motives of others so that they are not deceived. Both deception and the ability to detect deceptive behavior among others are favored by natural selection in the models made by game theorists.