Towards Realistic Models for Evolution of Cooperation

Lik Mui (lmui@lcs.mit.edu)

Abstract

Throughout evolution, human have exhibited cooperative behaviors which do not directly contribute to their individual survival. Activities such as warfare, caring for the young and sick, altruistic acts to the community, *etc*. contribute to the public good. This raises the question for why people regularly engage in these cooperative activities – many of which are costly to the individual. This paper summarizes the strengths and weaknesses of major approaches that have been taken to answer this question. Future directions for this future are suggested.

Outline

1. Introduction 2
2. Existing Approaches to Study Evolution of Cooperation 4
3. Extending Existing Models 9
4. Conclusion 10
Appendix: Prisoner's Dilemma 11
References 13

1. Introduction

Darwin's theory of evolution argues for the "survival of the fittest" but does not specify the unit for fitness measure. By "fittest", one could mean the individual, the population, the gene, or the organization around a group of agents. Each interpretation calls for very different models for how natural selection works.

Rational agency theory underpins most of modern economics (Kreps, 1990) and interprets "fittest" to refer to the individual. This individualistic view treats social organization as a byproduct of self-interest. Phenomena such as altruism and cooperation are dismissed as no more than disguises with selfish motivations (Kreps and Wilson, 1982).[1] Such interpretations are in contradiction with sociological and biological observations of human and animals (Trivers, 1971; Ostrom, 1998). Acts of giving and receiving permeate human life and the lives of many animals. However, the individualistic view of evolutionary theory predicts that reciprocally altruistic cooperation should be rare – natural selection should favor selfish behaviors.

The beginning half of the past century saw the proponents of interpreting "fittest" as that of the species or population (Haldane, 1932; Eshel, 1972). This is known as the "group selection" theory. Although contested by latest evolutionary theorists, group selection has for a time been able to account for why individual agents cooperate.

One dominant group of theorists interprets "fittest" to refer to the gene (Hamilton, 1964; Dawkins, 1979). Cooperation and related behaviors are explained in terms of fitness contribution to the individual gene. Individuals' aim in life is to maximize the chances that their set of genes (which also exists in others around them) is passed on to the next generation – even at the expense of the individual's fitness.

Yet another group of theorists considers "fittest" to refer to the social organization surrounding individual agents (Simon, 1969, 1990). This interpretation considers cooperation and sacrificial acts as enhancing the fitness of groups of agents within social structures, which in turn increases the individual fitness of its members.

This paper summarizes existing approaches to the study of evolution of cooperation. Strengths and weaknesses of these approaches are discussed. Suggestions on how this field can move forward are provided.

[1] Like Simon (1990), altruism is taken in this paper to mean behaviors which on average increases the reproductive fitness of others at the expense of the fitness of the altruist. Fitness refers to the expected number of progeny. Cooperation often refers to altruistic acts where the cooperator puts future self-interest on the line for the recipient of these acts.

	Group Selection	Kinship Theory	Direct Reciprocation	Indirect Reciprocation	Social Learning
Supporting	Eshel, 1972; Wilson, *et al.*, 1994;	Hamilton, 1963, 1964; Dawkins, 1976; Riolo, *et al.*, 2001;	Trivers, 1971; Axelrod, *et al.*, 1981; Axelrod, 1984; Roberts, *et al.*, 1998;	Alexander, 1987; Pollock, *et al.*, 1992; Nowak, *et al.*, 1998; Lotem, *et al.*, 1999; Wedekind, *et al.*, 2000;	Simon, 1969, 1991; Boyd, *et al.*, 1982;
Unsupporting	Fisher, 1958; Hamilton, 1963; Axelrod, *et al.*, 1981;	Murray, *et al*. 1984; Taylor, 1992; Wilson, 1992; Queller, 1994; West, *et al*., 2001; Axelrod, *et al.*, 1981;	Alexander, 1987; Boyd, *et al.*, 1987; Nowak, *et al.*, 1998;	Boyd, *et al.*, 1989; Riolo, *et al.*, 2001;	N/A

Table 1. A number of significant works have been put into the four major categories of theories on the evolution of cooperation

2. Existing Approaches to Study Evolution of Cooperation

This paper examines five major approaches to the study of evolution of cooperation over the past century. Significant publications for each approach are briefly summarized. Table 1 highlights the representative publications for these approaches.

2.1 Group Selection

Darwin's theory of evolution argues for the "survival of the fittest" but does not specify the unit of fitness. If the unit is on the individual level, cooperation among individuals would not be consistent with Darwin's theory. A group of evolutionary theorists (Eshel, 1972; Wilson and Sober, 1994) therefore postulated the unit of natural selection is on the group level: species, community, population, *etc*.

Many argue that group selection theory is misguided (Axelrod and Hamilton, 1981). Difficulties with this theory arise from the strife often observed within species, communities, and populations. Nevertheless, to explain for the manifest existence of cooperation and related behavior such as altruism, several alternative theories have been proposed.

2.2 Kinship Theory

Kinship theory is based on the commonly observed cooperative behaviors such as altruism exhibited by parents toward their children, nepotism in human societies, *etc*. Such behaviors toward one's kin not only not increase individual fitness of the donor while benefiting the fitness of others, they often incur costs – thereby decreasing personal fitness.

Hamilton's rule of relatedness provides the foundation of much of the work on kinship theory. This rule states that altruism (or less aggression) is favored when the following inequality holds:

$$rb - c > 0$$

where r is the genetic relatedness of two interacting agents, b is the fitness benefit to the beneficiary, and c is the fitness cost to the altruist. This rule suggests that agents should show more altruism and less aggression toward closer kin. Hamilton (1964) uses Wright's Coefficient of Relatedness for r.[2]

Hamilton (1964) suggests that it is not the individual fitness that is evolutionarily selected but it is that of the "inclusive fitness" of a gene. *Inclusive fitness* measures the ability for a gene to reproduce itself in the offspring. This differs from the classical Darwinian notion of "individual fitness" in that the central actor of evolution is the hereditary unit and not the individual. Dawkins (1979) popularized this "selfish gene" thesis in his 1979 book. By using inclusive

[2] Wright's Coefficient of Relatedness r=1 if the agent is compared to itself; r=0.5 for same parents siblings. r=0.25 for a grandparent and a grandchild; and so on. r can be thought of as the proportion of genes shared between two individuals.

fitness for analysis, Hamilton and others are able to account for cooperative behaviors as means to maximize the inclusive fitness of individuals' genes.

Problems with Kinship Theory

A number of studies have contested Hamilton's thesis (Taylor, 1992; Wilson, 1992; Queller, 1994). The main difficulty pointed out is the interplay between relatedness and competition among relatives. Factors that tend to increase the average relatedness of interacting agents, such as limited dispersal, also tend to increase the amount of competition among relatives (Murray, *et al.*, 1984). West, *et al.*, (2001) have empirically examined fig wasp in nature and reported that with limited dispersal, the increased competition between relatives can negate the effect of increased relatedness in favoring altruism. Different fixes to Hamilton's rule have been proposed. West, *et al.,* (2001) modifies the rule replacing b as follows:

$$b = B - a\,(B - c)$$

where c is the fitness cost as before and B is the benefit that would be given to the beneficiary in an interaction if competition does not exist. The parameter $a \in [0, 1]$ measures the extent to which neighbors compete (small a confers little competition while large a confers the opposite). $a = 0$ equals the original Hamilton' rule.

Kinship theory also has difficulty in explaining cooperation where relatedness is low or absent. Axelrod and Hamilton (1981) pointed out examples from mutualistic symbioses such as that between fungus and alga that compose a lichen; the fig wasps and fig trees where wasps serve as the tree's sole means of pollination and seed set. Cooperation in such symbioses can sometimes turn into antagonism (Caullery, 1952). Kinship theory cannot explain this dynamics of cooperation at all. The theory of reciprocation explains symbioses in a much more satisfactory manner.

2.3 Reciprocation Theory

Trivers (1971) postulated "reciprocal altruism" as an explanation for how individuals are willing to sacrifice personal gain for the good of another. Specifically, phenomena such as friendship, moralistic aggression, gratitude, sympathy, trustworthiness, *etc*. can be explained in terms of deferring immediate personal gain toward a future reciprocation.

Much of the work on reciprocation theory is carried out using a game theoretic framework based on the Prisoner's Dilemma (PD) game and the Tit-for-Tat (TFT) strategy. The appendix briefly reviews this game and its relevance to evolutionary simulations.

Direct Reciprocation

The strategy TFT has been postulated by Axelrod and Hamilton (1981) to be an evolutionary stable strategy for the iterated PD game, provided with a sufficiently large probability of repeated future encounters between agents. They label this probability "shadow of the future"

(w). In other words, cooperation between agents can evolve using the reciprocating strategy of TFT.

TFT is a direct reciprocating strategy for agents in the PD game of reciprocating what the other agent did in the last round. In the context of an iterated PD game, the threshold w can be calculated analytically. Consider a game with n TFT agents against n ALLD agents (*c.f.*, Appendix on agent strategies) and a given w and the payoff matrix shown in the Appendix:

Payoff for TFT against TFT: $R\,(1 + w + w^2 + \ldots) = R/(1-w)$
Payoff for TFT against ALLD: $T + P(w + w^2 + \ldots) = T + wP/(1-w)$

For TFT agents to defend themselves against ALLD agents, the following inequality must hold:

$$R/(1-w) > T + wP/(1-w) \qquad or \qquad w > (T-R)/(T-P)$$

Experimental findings of TFT behaviors have been found for tree swallows (Lombardo, 1985) and sticklebacks (Milinski, 1987) in nature.

Other theoreticians have disapproved of TFT's significance as a workable strategy toward cooperation. Specifically, Boyd and Lorberbaum (1987) have shown that there does not exist a single strategy (including TFT) which can be resistant to all kinds of invading strategies for the PD game. The counter proof they use to show that TFT is not an evolutionarily stable strategy is by introducing two others: the Tit-for-two-tat (TFTT) and Suspicious-tit-for-tat (STFT) strategies. A TFTT agent does not retaliate until there have been two successive defections; a STFT agent starts off defecting on the first encounter but thereafter plays TFT.[3]

Indirect Reciprocation

It is a common practice in human societies where the donor of a good deed might not necessarily be rewarded by the recipient directly but by other individuals who might be recipients of other good deeds by others. Alexander (1987) is the first to term this phenomenon "indirect reciprocity."

Boyd and Richerson (1989) are the first to develop a mathematical model of cooperation based on indirect reciprocity. In their model, donors are rewarded for their deeds by the last person in a ring of n indirectly-reciprocating individuals. Their analysis of the indirect reciprocity indicates that the conditions necessary for the evolution of indirect reciprocity "… become restrictive as group size increases." Their analysis however does not take into account social structures other than ringed interactions. Human agents and their interactions tend to form embedded graphs which are not rings (Granovetta, 1985).

Nowak and Sigmund (1998) developed an indirect reciprocation model based on image (reputation) scoring. In their non-overlapping evolutionary model, each agent has a genetic

[3] If a group of agents use all TFT, STFT, and TFTT strategies, TFTT has the edge over the other two. Therefore, TFT is not evolutionarily stable.

strategy and a non-heritable image score.[4] They showed that cooperation can be established under "global" image scoring if the number of interactions per generation is sufficiently large. "Global" refers to the nature of an agent's image being visible to all other agents. Following the results by Pollock and Dugatkin (1989), Nowak and Sigmund (1998) further studied the effect of randomly selected observers on the evolution of cooperation. Every interaction is no longer globally known but is observed by a limited number of observers -- thereby removing the global image constraint. They conclude that cooperation may have evolved through indirect reciprocity with or without global knowledge about agents' image scores. Wedekind and Milinski (2000) have experimentally verified Nowak and Sigmund's result that image scoring does play a role in human cooperation.

Extensions to Nowak and Sigmund's work have been done in two areas. Lotem, *et al.* (1999) have included persistent non-cooperators ("phenotypic defectors") to model after the sick, young or handicapped who may be unable to cooperate even if they are genetically predisposed to do so. Their simulations show that phenotypic defectors paradoxically allow persistent discriminating cooperation under a much wider range of conditions than found by Nowak and Sigmund.

Riolo, *et al.* (2001) have infused kinship theory into Nowak and Sigmund's model with "tag-based" reciprocity – cooperation based on inheritable and identifiable tags on agents. Their simulations indicate that cooperation can evolve even when reciprocity is absent.

2.4 Social Learning

Boyd and Richerson (1982) introduce a model for cooperation to evolve based on "conformist transmission," or *cultural transmission.* This mechanism refers to the preferential selection of the behaviors individuals encounter most frequently. In other words, individuals learn the most dominant behaviors in their embedded social network. Although cooperation can evolve using this mechanism, conformers and non-conformers are not distinguished. Along this line, Simon (1990) proposes an alternate social learning approach.

Simon (1990) argues that in human societies, cooperation exists where kinship and reciprocation are absent. Specifically, he uses the case of altruism often documented in organizational studies where real people who are satisfied with their positions adopt the organization's interest as their own interest.

Simon (1990) considers a population of two types of individuals in his model: A and S, in proportions p and $1 - p$, respectively. Individuals of type A are altruistic (who always cooperate) while those of type S are selfish. Each A exhibits behaviors that contributes b offspring to members of the population (including himself). The cost to A is c fewer children for A. The average number of offspring, F_A and F_S, of each A and S in the absence of social learning is:

[4] The next section will explain this model in details, along with simulation results not reported by Nowak and Sigmund (1998).

$$F_A = X - c + bp$$
$$F_S = X + bp$$

where X is the number of offspring in the absence of altruistic behavior. To clarify, any individuals can be recipients to the np altruists, and selfish S individuals have no altruism costs. Without social learning, since $c > 0$, altruists always have fewer children than selfish ones. In evolutionary game theory terms, selfish individuals are more fit and will therefore have the evolutionarily stable strategy over altruistic ones.

To model social learning (learning from others in the society), Simon introduces the docility parameter (d), which refers to the willingness of an individual to " … accept well the instruction society provides them." The content of what is learned will not be fully screened for its contribution to personal fitness. Docility models the level of bounded rationality and computation limitation of individuals. The higher the docility, the more bounded rational and the more computationally limited an individual is. F_A and F_S can be modified as follows:

$$F_A = X + d - c + b(c)\, p$$
$$F_S = X + b(c)\, p$$

b is now a function of c because the amount of altruism exacted from A depend on the society's definition of proper behavior. The condition for cooperation (and altruists) to dominate is:

$$d - c > 0$$

3. Extending Existing Models

Research attempting to explain how cooperation arises in society has been at least a century old. Relative to many other research problems, this is a mature field with the wisdom of many very intelligent and dedicated researchers. Nonetheless, existing models fall far short of fully explaining how and why cooperation evolves in natural population for the following reasons:

3.1 Unifying Perspectives

Firstly, existing models and theories appear to have conflicting assumptions and implications, as illustrated in the two rows of supporting and non-supporting publications in Table 1. Each of the five major approaches discussed in this paper seems to explain certain aspects of cooperation and how it evolves but lacking in other aspects. Clearly, cooperation does evolve. Therefore, a unified theory of how and why cooperation evolves still eludes the research community.

3.2 Towards Realistic Models

Secondly, most models are still based on the prisoner's dilemma game which is a gross over-simplification of actual agent-to-agent interactions. Important elements comprising actual animal interactions are missing in existing models. Glaringly amiss are the following:

- Nearly all existing models are based on asexual, non-overlapping generations of agents. Obviously, these are not realistic assumptions.
- Simultaneous play for every interaction is assumed in all but a few models (e.g., Abell and Reyniers, 2000)
- Interactions are restricted to dyadic. Clearly, animals are often involved in interactions involving more than one other animal. Studies into how other types of interactions are much needed.
- Agent behavior is assumed to be unmistakable. As pointed out by May (1987), human behavior is far from mistake-free. Future studies of how cooperation evolves should take the stochasticity of behavior seriously.
- Actions are discrete in existing models. Clearly cooperation is not just "cooperate" and "defect" alone but the levels of cooperation are continuous. Perhaps this continuity is determined by the amount trust between agents.
- Animals are not born with specific and unyielding strategies. Depending on the situation, animals might switch between different types of strategies in cooperating with others. Strategy modification and defection should be studied.
- Although works by Simon (1990, 1991), Cohen, et al., (2001), and others have brought social structure and learning into the study of cooperation, much work is still needed to under how and to what extent various types of structure affects cooperation. Given the importance placed on cooperation across economic entities, this should be of enormous interest to the business community.

It is hard to believe given the century old nature of the problem of cooperation, there are so many glaringly unrealistic aspects in current models. This lack of progress points to the difficulty and mysteriousness of this natural phenomenon.

4. Conclusion

Throughout evolution, human have exhibited cooperative behaviors which do not directly contribute to their individual survival. Activities such as warfare, caring for the young and sick, altruistic acts to the community, *etc*. contribute to the public good. This raises the question for why people regularly engage in these cooperative activities – many of which are costly to the individual. This paper has summarized the major approaches that have been taken to answer this question. After over a century of research, answer to this question remains open. This field continues to provide a fruitful area for evolutionary research.

Appendix: Prisoner's Dilemma

Nearly all quantitative modeling in the study of evolution of cooperation uses game theory (Axelrod, 2000). The most common game studied for evolution is the prisoner's dilemma game. This session briefly describes this game.

One Shot PD Games

For the PD game, the action space for each agent is:

Action = { *cooperate*, *defect* }

The payoff matrix for the prisoners' dilemma game is:

		agent 2	
		C	D
agent 1	C	R, R	S, T
	D	T, S	P, P

Figure 1. *Payoff matrix for a one-shot prisoners' dilemma game, where C = cooperate, D = defect.*

With the following two constraints, the "dilemma" can be easily seen (Axelrod, 1984):

$$T > R > P > S$$
$$2R > T + S$$

These assignments satisfy these constraints: T = 5, R = 3, P = 1, S = 0. Given agent 2 is choosing "cooperate," agent 1 can maximize its payoff by choosing "defect"; given agent 2 is choosing "defect," agent 1 can maximize its payoff by choosing "defect." Therefore, the *Nash Equilibrium* [5] of this one shot PD game is: Defect-Defect (the lower right quadrant of the payoff matrix).

By the rational agency theory of economics (Kreps, 1990), the utility maximization point is therefore *Defect-Defect*. However, both agents can do better by cooperating,[6] herein lies the dilemma.

Iterated PD Games

If agents are indeed selfish utility-maximizing, economics tells us that in one shot PD games, they cannot do better than what is indicated by the Nash Equilibrium solution of defection. However, if the PD game is played multiple times between two agents, cooperation has some

[5] "Nash Equilibrium" refers to a utility-maximizing stability point where either agent cannot be better off by perturbing their actions at the equilibrium (Kreps, 1990).

[6] *Cooperate-Cooperate* is termed *pareto-superior* to the Nash Equilibrium point *Defect-Defect* in the one-shot PD game.

chance. The details will be deferred to the discussion for direct reciprocity later in the paper. In the game-theoretic literature for describing payoffs for prisoners' dilemma (PD) game, four descriptions are used for the outcomes of the game:

- *Temptation (T)*: when an agent defects (D) while its opponent cooperates (C)
- *Reward (R)*: when both an agent and its opponent cooperate (C) with each other
- *Punishment (P)*: when both an agent and its opponent defect (D) against each other
- *Sucker (S)*: when an agent cooperates (C) while its opponent defects (D)

Evolutionary PD Games

To simulate the evolutionary process, birth and death of agents are introduced for multiple generations of iterated PD games among groups of agents. The most common scenario is the so called "non-overlapping" generations where all agents of one generation dies before the next generation is created (Axelrod and Hamilton 1984; Nowak and Sigmund, 1998). A generation is defined by a set of PD encounters among the agents. Participants in an encounter are chosen randomly from the total population. Often the goal of the evolutionary simulation is to assess the fitness of specific agent strategies or attributes.

In many studies, the goal is to understand the strategic implications for the survival of agents. To carryout these studies, at the end of a generation (where a certain number of dyadic encounters between agents have occurred), the total fitness of all agents with the same strategy is calculated. The number of progeny in the next generation with a given strategy is proportional to the total fitness associated with that strategy in the previous generation. The total population size is often fixed, so any increase in the number of one type of agent is balanced by a decrease in the numbers of other types of agents.

Commonly studies agent strategies are:

- *Cooperate (ALLC)*: always cooperates.
- *Defect (ALLD)*: always defects.
- *Tit-for-tat (TFT)*: initially cooperates, and thereafter does what the other agent did in the last round.
- *Suspicious Tit-for-tat (STFT)*: initially defects, and thereafter does what the other agent did in the last round.
- *Grim-trigger*: initially cooperates, but once defected by another agent, this agent always defects against that other agent.

Axelrod (1984) has documented empirical evidences for the robustness of TFT compared to other strategies under competitive settings of many types of strategies.

References

Abell, P. & Reyniers, D. Generalized reciprocity and reputation in the theory of cooperation: a framework. *Analyse and Kritik* **22**, 3-18 (2000).

Alexander, R. D. *The Biology f Moral Systems* (Aldine de Gruyer, New York, 1987).

Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390 (1981).

Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).

Axelrod, R. On six advances in cooperation theory. *Analyse & Kritik* **22**, 130-151 (2000).

Boyd, R & Lorberbaum, J. No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature* **327**, 58-59 (1987).

Boyd, R. & Richerson, P. J. The evolution of indirect reciprocity. *Social Networks* **11**, 213-236 (1989).

Boyd, R. & Richerson, P. J. Cultural transmission and the evolution of cooperative behavior. *Human Ecology* **10**, 325-351 (1982).

Caullery, M. *Parasitism and Symbiosis* (Sidgwick and Jackson, London, 1952).

Cohen, M. D., Riolo, R. L., Axelrod, R. The role of social structure in the maintenance of cooperative regimes. *Rationality and Society* **13**, 5-32 (2001).

Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, Oxford, 1976).

Kreps, D. *A Course in Microeconomics* (Princeton Univ. Press, Princeton, NJ, 1990).

Kreps, D. M. & Wilson, R. Reputation and imperfect information. J. Econ. Theory **27**, 253-279 (1982).

Eshel, I. On the neighbourhood effect and evolution of altruistic traits. *Theor. Popul. Bio.* **3**, 258-277 (1972).

Fisher, R. A. The genetic theory of natural section. 2nd ed. (Dover, New York, 1958).

Granovetter, M. Economic Action and Social Structure: the Problem of Embeddedness. *Am. J. of Soc.* **91**, 481-510 (1985).

Haldane, J. B. S. The causes of evolution. Longmans Geen & Co, London (1932).

Hamilton, W. D. The evolution of altruistic behavior. *Am. Nat.* **97**, 354-356 (1963).

Hamilton, W. D. The genetic evolution of social behaviour, I & II. *J. Theor. Biol.* **7**, 1-52 (1964).

Keller, L. & Reeve, H. K. Familiarity breeds cooperation. *Nature* **394**, 121-122 (1998).

Lombardo, M. P. Mutual restraint in tree swallows – a test of the tit for tat model of reciprocity. *Science* **227** (4692), 1363-1365 (1985).

Lotem, A., Fishman, M. A., Stone, L. Evolution of cooperation between individuals. *Nature* **400**, 226-227 (1999).

Milinski, M. Tit-for-tat in sticklebacks and the evolution of cooperation. *Nature* **325**, 433-435 (1987).

Murray, M. G. & Gerrard, R. J. Conflict in the neighbourhood: models where close relatives are in direct comopetition. *J. Theor. Biol.* 111, 237-246 (1984).

Nowak, M. & Sigmund, K. Evolution of Indirect Reciprocity by Image Scoring. *Nature* **393**, 573-577 (1998).

Nowak, M. & Sigmund, K. The dynamics of indirect reciprocity. *J. Theor. Biol.* **194**, 561-574 (1998).

Ostrom, E. A Behavioral Approach to the Rational-Choice Theory of Collective Action. *Am. Polit. Sci. Rev* **92**(1), 1-22 (1998).

Pollock, G. B. & Dugatkin, L. A. Reciprocity and the evolutionofreutation. *J. Theor. Biol.* **159**, 25-37 (1992).

Queller, D. C. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**, 70-73 (1994).

Riolo, R. L., Cohen, M. D. & Axelrod, R. Evolution of cooperation without reciprocity. *Nature* **414**, 441-443 (2001).

Roberts, G. & Sherratt, T. N. Development of cooperative relationships through increasing investment. *Nature* **394**, 175-179 (1998).

Simon, H. *The Science of the Artificial.* (Cambridge, MA: MIT Press, 1969).

Simon, H. A mechanism for social selection and successful altruism. *Science* **250**, 1665-1668 (1990).

Simon, H. Organizations and markets. *J. Econ. Perspectives* **5**, 859-871 (1991).

Taylor, P. D. Inclusive fitness in a homogeneous environment. *Proc. R. Soc. London. B*, **249**, 299-302 (1992).

Trivers, R. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35-57 (1971).

Wedekind, C., Milinski, M. Cooperation through image scoring in humans. *Science*, **288**, 850-852 (2000).

West, S. A., Murray, M. G., Machado, C. A., Griffin, A. S. & Herre, E. A. Testing Hamilton's rule with competition between relatives. *Nature*, **409**, 510-513 (2001).

Wilson, D. S., Pollock, G. B. & Dugatkin, L. A. Can altruism evolve in purely viscous populations. *Evol. Ecol.*, **6**, 331-341 (1992).

Wilson, D. S. & Sober, E. Reintroducing group selection to the human behavioural sciences. *Behav. Brain Sci.* **17**, 585-654 (1994).