Articles

The Further Evolution of Cooperation

ROBERT AXELROD AND DOUGLAS DION

Axelrod's model of the evolution of cooperation was based on the iterated Prisoner's Dilemma. Empirical work following this approach has helped establish the prevalence of cooperation based on reciprocity. Theoretical work has led to a deeper understanding of the role of other factors in the evolution of cooperation: the number of players, the range of possible choices, variation in the payoff structure, noise, the shadow of the future, population dynamics, and population structure.

COOPERATION IS A TOPIC OF CONTINUING INTEREST FOR the social and biological sciences. A theory of cooperation based upon reciprocity (*1–3*) has engendered a wide literature concerning the evolution of cooperation. In this article, we survey this literature in order to determine what new insights have been gained.

Although we shall concentrate on theoretical work, it is important to note that scholars have been active in pursuing empirical applications of the theory. Huth (*4*) found that military conflict during the last century was most successfully deterred when a challenge was met with reciprocity. Cooperation based on reciprocity has been supported for vampire bats (*5, 6*), vervet monkeys (*7*), and sessile invertebrates (*8*). Experimental simulations of defection have been presented to stickleback fish (*9*) and tree swallows (*10*); the findings are consistent with reciprocity. On the other hand, Nol (*11*) rejected reciprocity as an explanation of cooperation between the sexes in chick-rearing by the American oyster catcher, arguing that data on foraging trips and energy use support a theory based on the efficient allocation of energy by the birds. Other investigators have pointed to the difficulties in determining whether observed cooperation is due to a tit-for-tat-like process (*12*). Despite these difficulties, cooperation based upon reciprocity has received substantial empirical support.

In addition, advice has been offered for problems of breach of contracts (*13*), child custody (*14*), superpower negotiations (*15*), and international trade (*16, 17*).

We begin our analysis of recent theoretical work on the evolution of cooperation by reviewing Axelrod's original formulation and theoretical results (*2*). Is there any reason for an individual to cooperate when noncooperative behavior is rewarded? This question underlies the problem of cooperation in the Prisoner's Dilemma. In the Prisoner's Dilemma, each player has two choices: cooperate or defect (Fig. 1). The choice may or may not be made rationally. In either case, if the game is only played once, then each player gets a higher payoff from defecting than from cooperating, regardless of what the other player does. However, if both players defect, they both do worse than had both cooperated.

If the game is played repeatedly (the iterated Prisoner's Dilemma, or IPD), there is greater room for cooperation. To see which strategies would be effective in exploiting this opportunity for cooperation, game theorists were invited to submit programs for a round-robin IPD computer tournament having the following properties:

1) The interactions were between pairs of players.

2) Each player had two available choices on each move: cooperate or defect. Choices were made simultaneously.

3) The payoffs (Fig. 1) were fixed before play and announced to all players.

4) At each move in the game, each player had access to the history of the game up to that move—in short, there was no noise in the transmission of strategy choices between players.

Rankings of the strategies were determined by the number of points achieved overall. The winner of the first tournament was TIT FOR TAT (TFT), a program that uses cooperation on the first move of the game and then plays whatever the other player chose on the previous move. Results of this tournament were publicized and a second tournament received 62 entries. Unlike the first tournament, in which the number of iterations was known beforehand, the second tournament had the following additional characteristic.

5) There was a fixed probability (called the "shadow of the future") of the game ending on the next move.

The winner of this second tournament was again TFT. TFT's success was due to its being nice (not the first to defect), provokable (responding to the other player's defection with a defection), forgiving (punishing and then cooperating after a defection), and clear (easy for other players to understand) (*2*).

But success in the computer tournaments did not prove that TFT would perform well as an evolutionary strategy. After all, TFT's success might have been due to its performance with other strategies that would not themselves survive for very long. To check this possibility, an ecological simulation was conducted. The initial population consisted of the entries to the second round of the computer tournament with the following characteristic:

6) The population dynamics of the ecological simulation were determined by setting the change in frequency of each strategy in

Player A \ Player B	C Cooperation	D Defection
C Cooperation	$R=3$ Reward for mutual cooperation	$S=0$ Sucker's payoff
D Defection	$T=5$ Temptation to defect	$P=1$ Punishment for mutual defection

Fig. 1. The Prisoner's Dilemma game. The payoff to player A is shown with illustrative numerical values. The game is defined by $T > R > P > S$ and $R > (S + T)/2$.

R. Axelrod is a professor of political science and public policy at the Institute of Public Policy Studies, University of Michigan, Ann Arbor, MI 48109. D. Dion is a graduate student in the Department of Political Science, University of Michigan.

any given round to be proportional to its relative success in the previous round.

In this ecological simulation, TFT quickly became the most common strategy.

Once cooperation based upon reciprocity is established, no player can do any better than to cooperate as well, provided the chance for future interaction, w, is high enough. For example, if everyone is using TFT then no one mutant can do better if $w \geq \max [T - R)/(R - S), (T - R)/(T - P)]$ (*2*, pp. 207–208). In game theory terms, this means that everyone using TFT is a Nash equilibrium when the shadow of the future is sufficiently high (*18*).

Getting cooperation started required something more because a single cooperative player can never do better than the population average of a group of noncooperating players. So the implications of clustering of strategies were explored by supposing the following condition holds (at least initially).

7) Populations are structured in the sense that strategies have more-than-random interaction with strategies of their own type.

Axelrod (*2*, pp. 63–69) showed that even a small cluster of players who used cooperation based upon reciprocity could establish themselves in a population of noncooperative players, and even take over such a population. Moreover, once established, the reciprocating cooperative players would be immune from re-invasion by a similar cluster of noncooperating players. Thus, the evolution of cooperation contains a ratchet.

In the remainder of this article, we review the work that has been done on altering or relaxing each of the seven assumptions listed above.

Interactions

In the original formulation, the interactions are between pairs of players (*2*, p. 30). For certain applications, however, interactions involve more than two players. Consequently, the corresponding game is the *n*-player PD (NPD), in which players make a choice (cooperate or defect) which they play with all other players.

Theorists have shown that increasing the number of individuals in the NPD makes cooperation more difficult (*19–22*). Taylor (*19*), in a nonevolutionary framework, proved that for cooperation to be part of an equilibrium in an NPD when some players are playing ALL D (the strategy which defects on each move), it is necessary that either the shadow of the future is long or the number of cooperators is large. In an evolutionary setting, Joshi (*20*) has found that if individuals play a "hard" TFT (meaning that they will cooperate until one player defects), and the number of individuals playing a hard TFT passes a certain threshold, then hard TFT can dominate a population of ALL D players. But this threshold rises as the number of individuals in the society increases.

The effectiveness of TFT in the NPD, however, depends upon the type of NPD that is being played. Pettit (*23*) has distinguished "free-riding" from "foul-dealing" NPDs. In a free-riding NPD, a lone defection places the remainder of the population worse off than total cooperation but still better off than complete defection. In a foul-dealing NPD, the defection of one individual places some subset of the population in a position worse than universal defection. Pettit argues that only in the foul-dealing iterated NPD is the retaliation implied by TFT rational, since in the free-ridng iterated NPD each player would rather cooperate with the remainder of the population than lose the benefits of cooperation by punishing the lone defector.

Pettit's distinction raises two questions. First, is it true that the coalition behavior Pettit suggests would actually be effective in a variegated environment? An answer is provided by Fader and Hauser (*24*) who ran a computer tournament involving a three-firm oligopoly game. They found that effective strategies tended to take advantage of the opportunity for cooperating with other players, including cooperating with the remaining player if the third player defects (charged the competitive price). Winners in the Fader and Hauser tournament also tended to be magnanimous, offering potential partners a coalition price that was slightly above the optimal coalition price (in this way, they avoided being misperceived as competitive). This work is paralleled in the study of alliances among primates (*25*). Alliances potentially can form when a pair of organisms is engaged in a game with each other as well as a game against a common enemy. Aoki (*26*) has formally demonstrated that if the gains to altruism (both in terms of direct benefits and inclusive fitness) exceed the coats to altruism, then a genetic trait for forming reciprocal alliances will increase in the population.

The second question: if there is an incentive not to punish defectors, can some change in the game structure reintroduce the incentive to retaliate? Axelrod's (*27*) analysis of the metanorms game suggests there is. In the metanorms game, individuals still have the option to punish a defector. However, individuals also have the opportunity to punish others who have failed to punish defectors. An evolutionary simulation of the metanorms game shows that the population quickly becomes vengeful in their punishment of defections and not so bold in the attempting of defections. This suggests the establishment of a norm of cooperation. This analysis of the evolution of norms may give us a solution to the difficulty posed by Pettit. Although individual punishment of defection may not be individually optimal in the free-riding iterated NPD, in the context of a metanorm it may be optimal for individuals to punish defectors.

Choices

One limitation of the IPD is the assumption of two possible choices (cooperate and defect) chosen simultaneously by the two players. In many applications there is an option to leave the game (exit) or to force others to leave the game (ostracism). One might suppose that when exit was possible, the cooperators would leave first to avoid exploitation by the defectors. But in an experiment allowing human subjects to exit from a nine-player PD, this did not happen (*28*). The reason was the cooperators' greater concern for the welfare of the group that led them to cooperate in the first place. Hirshleifer and Rasmusen (*29*) have used the presence of ostracism to derive the stability of cooperation in the finitely repeated PD in all but the last round. The cooperation-inducing effects of ostracism have been empirically supported in a case study by Barner-Barry (*30*), who found that ostracism of a school-yard bully led to the bully's eventual attempts at cooperation. The assumption of simultaneous choice has been relaxed by Kondo (*31*), who has shown that the stable equilibria of the nonsimultaneous move game are either ALL D or a form of conditional cooperation.

Payoffs

How robust are the original solutions to changes in the payoff matrix? Lipman (*32*) has generalized the Nash equilibrium results to iterated Chicken, a game that models crisis bargaining. Dacey and Pendegraft (*33*) ran a computer tournament in which programs played both the IPD and Chicken. They found that TFT does well in the provokable worlds of the simulation, but that PERMANENT RETALIATION (cooperate until the other player defects and then defect for the remainder of the game) excelled in the Chicken games.

In the IPD, the shadow of the future and reciprocity help support cooperation. This is a special case of the folk theorem, a theorem that applies to all iterated games. The folk theorem states that any individually rational outcome (*34*) of any one-shot game can be supported as a Nash equilibrium of the infinitely repeated game, assuming the future is important enough (*35*). In the strategies constructed in the proof of this (and similar theorems) conditional cooperation is used as the basis for deterring defecton. Such theorems also show the variety of outcomes (besides mutual cooperation) that can be supported in equilibrium in the IPD.

Noise

Faulty transmission of strategy choices (noise) severely undercuts the effectiveness of reciprocating strategies. Molander (*36*) has shown that in the presence of any amount of noise, two TFT players will in the long run average the payoffs of two interacting RANDOM players, each of whom cooperates and defects with equal probability. In fact, a risk-averse player would prefer a strategy of any weighted average reciprocator (reciprocates based upon a longer history of the game than TFT) to TFT, as Bendor (*37*) has shown.

If few environments were characterized by noise, then TFT's vulnerability would not be very important, but this is not the case. Various species lack the capacity for individual recognition, which implies that they will be unable to determine the history of interaction with any given individual. Species with individual recognition may still only imperfectly recognize others. Noise can even be a serious problem between people or between nations (*15, 38*) and can occur in any game in which monitoring of the other player is difficult.

Two tournaments have been conducted to see how TFT would do in a noisy variegated environment. The first is reported in Axelrod (*2*, p. 183), who noted that rerunning the first round of the computer tournament with a 1 percent probability of misperceiving the other player's choice still found TFT to be the best decision rule. The second tournament, by Donninger (*39*), used a 10 percent chance of misperception. In this tournament, TFT finished sixth out of 21. In variegated populations, then, TFT may still perform fairly well, even with noise.

How can TFT be modified to more easily cope with environments of noise? The solutions depend on the type of noise. One type of noise can be labeled "misimplementation." In this case, the player making the mistake knows that a mistake has been made and that the other player cannot distinguish the mistake from an intentional cooperation or defection. A variety of authors have suggested some unconditional cooperation (increasing the probability of playing C after misimplementing a prior C choice) on the part of the individual making the error as a response to small probabilities of misimplementation (*40*).

A second type of noise is misperception, where a player actually does make one choice but the other player believes that a different choice was made. In the case of misperception, if there is no information that the players can use to distinguish an intentional from an unintentional defection (or cooperative move), then there is no (game-theoretic) difference between the noiseless IPD and the IPD with misperception.

A variant of misperception is "noisy channels," where neither player knows that an error has been committed but they both know that such errors have some probability of occurring. In environments with sufficiently small amounts of this type of noise, risk-averse, self-interested players will both converge on the same level of generosity (defined as the probability that one's strategy plays C after the other player has played D) (*36*). Mueller (*41*) finds that the best strategy to maintain already established cooperation is a more or less restrained version of TFT, with the amount of restraint depending on the necessary safety margin against an eventual reinvasion of noncooperators.

Finally, another sort of noise, in which players observe each other's choices but a stochastic variable affects the payoffs, has been considered for the case of a "favor granting" game by Calvert (*42*). In equilibrium each player grants all favors asked to the other player. Intuitively, each player takes into account the stochastic element of the game, and therefore is not provoked when low payoffs are realized. If strategy choices are unobservable then cooperative outcomes may still be obtained, even in the NPD, by the use of trigger strategies (*22, 43*). A trigger strategy specifies defection for t moves whenever the player's payoff falls below a certain level. Such strategies are restrained in their retaliation, and explain temporary failures of cooperation as elements of Nash equilibrium behavior.

We have dealt so far with responses that players can make to noise within the context of the IPD, but it is important to note that changing the structure of the game can give players additional ways of dealing with noise. Players can choose strategies which are less ambiguous, as Dixit (*16*) has suggested for U.S. trade policy. There can be an explicit convention about what constitutes defection and cooperation (*44*). And players can unilaterally take steps that minimize the probability of error. For example, vampire bats tend to have very poor individual recognition. However, grooming behavior, with its attendant close contact, may provide the clues needed for a system of reciprocal food sharing to overcome the effects of noise from inadequate individual recognition (*6*).

The lessons of the literature on noise in the IPD suggest that for sufficiently small amounts of noise, unilateral generosity is the best response. However, for larger amounts of noise, there is a trade-off: unnecessary conflict can be avoided by generosity, but generosity invites exploitation.

The Shadow of the Future

The original paradigm assumed an indefinite ending to the game, with the probability of the game ending with the current move equal to w. This probability is called the shadow of the future and is equivalent to a discount rate. As mentioned earlier, if $w \geq \max[(T - R)/(R - S), (T - R)/(T - P)]$, there exists a Nash equilibrium in which all players use TFT. In this section, we shall see how various changes in the paradigm affect the way in which the shadow of the future supports cooperation.

If cooperation depends on a long enough shadow of the future, then it should not be surprising that variation in w can affect observed patterns of cooperation. In research on the implications of a nonconstant discount parameter, behavior-dependent contexts of play and finite repetition have been studied.

In work on behavior-dependent contexts of play, it is assumed that choices affect the probability of future interaction. Eshel and Weinshall (*45*) consider an iterated game with a stochastic payoff matrix. In some moves of this game, the players are in a PD, while in other moves cooperation is the dominant choice. If the probability of surviving the current move depends on the cumulative payoffs, an egoist may cooperate on the PD iterations of the game to guarantee the presence of a partner for reaping the benefits of mutual assistance in other iterations. In work reviewed below, Feldman and Thomas (*46*) investigate the case where the probability of future interactions depends on the choices of the players.

In work on finite repetition, the possibility of cooperation in the finitely repeated prisoner's dilemma, or FRPD, has been analyzed. Luce and Raiffa (*47*) proved by backward induction that cooperation was not rational in the FRPD (cooperation does not pay in the last round, and hence cannot pay in the next to last round, and so forth). In recognition of this problem, the second computer tournament used a probabilistic treatment of the possibility that the two players will meet again (*2*, p. 42).

Efforts to establish conditions for cooperation in the FRPD fall into three categories. The first employs the theory of automata to study players with limited computational ability. These theoretical results have been mixed, with cooperation being obtained in some models (*48*) but not in others (*49*). A second approach uses various combinations of player beliefs about payoffs and strategies. In one model, each player attaches a certain probability to the proposition that the other player will use TFT even if it is irrational to do so (*50*). A final approach uses learning theory to account for the tendency of college students engaged in a series of FRPDs to start defecting more and more moves before the end of the game (*51*). The results of these three lines of inquiry are particularly relevant because the rationality and information of people as well as animals is incomplete.

In a similar attempt at realism, finite population versions of the Nash equilibrium result have been presented (*21, 52, 53*). Aoki (*53*) has noted that if the size of the group numbers several tens, as is the case with primate groups or many human societies, then the infinite population theorem provides a good approximation to the finite population case. Pollock (*54*), however, notes that if the population consists of a single dyad, then for any discount rate TFT is not stable in the presence of an ALL D mutant. This point emphasizes the value to a TFT player of the presence of other reciprocating players.

The theoretical work on the importance of the shadow of the future has received empirical support from social and biological scientists. Moore (*55*) has found that the continued interactions among individuals in the Lebanese banking community laid the groundwork for the stability of that sector despite the great instability of Lebanese affairs. In a more rarefied setting, Murnighan and Roth (*56*) found that the probability of repeated play had a significant impact on the number of cooperative choices made in an IPD by undergraduates. In avian studies, Davies and Houston (*57*) suggest that the failure of two male dunnocks to practice cooperative polyandry reflects the low chance of their surviving the season and meeting again.

Population Dynamics

Readers familiar with the literature on game-theoretic models of evolution will note the difference between an evolutionarily stable strategy (ESS) (*58*) and a strategy that is in Nash equilibrium with itself. Let $V(X|Y)$ be the payoff to an X player when playing with a Y player. A strategy X is said to be an ESS if for all Y either (i) $V(X|X) > V(Y|X)$, or (ii) $V(X|X) = V(Y|X)$ and $V(X|Y) > V(Y|Y)$. By contrast, the Nash condition is that a strategy is stable if there is no Y such that $V(Y|X) > V(X|X)$. All strategies that are ESS are stable in the Nash sense but the converse need not hold.

The implications of this distinction can be best understood in terms of stability versus neutral stability. A strategy that is evolutionarily stable can completely overcome invasion by mutants. This is stability. A strategy that is stable in the Nash sense may be only neutrally stable for a unilateral deviation from the equilibrium (*59*). This allows mutants to drift into the population.

As it turns out, if one allows invasion by multiple mutants, there is no single strategy that is evolutionarily stable (ESS) in the IPD. This fact was discovered, apparently independently, by Pudaite (*60*) and Boyd and Lorberbaum (*61*). Pudaite showed that if there is some chance for future interaction, then for each strategy X there is always some set of strategies Z such that X is not an ESS against Z. Boyd and Lorberbaum showed that no pure strategy whose behavior is determined solely by the history of the game is an ESS if the future is important enough (that is, if $w > \min[(T - R)/(T - P), (P - S)/(R - S)]$).

These negative results place new value on the intensive analysis of sets of strategies, because it is only by understanding the set of possible competitors that we can understand the particular evolutionary path of cooperation. Two procedures seem particularly well adapted for such analysis.

The first procedure applies the mathematics of dynamical systems to the interactions of sets of representative strategies in order to determine the full characteristics of possible dynamics. We already noted the work of Feldman and Thomas (*46*) in which behavior-dependent discount parameters were used. One of their results points to, not only the existence, but also the local stability of a full polymorphic equilibrium (composed of TFT, ALL D, and a strategy that defects on all but the first move) given certain conditions on the discount rates. This result is of particular importance in determining whether observed genotypic variation is merely transitory or in fact stable. Other population dynamic treatments are presented in Blad (*62*), Cave (*21*), and Pudaite (*60*).

A limitation of these dynamic treatments, as well as the ecological tournament, is their inability to develop new strategies. One method of overcoming this limitation is to use a genetic approach to develop new strategies for playing the IPD. A good method of implementing this on a computer is Holland's genetic algorithm (*63*). A strategy can be represented as a "chromosome" describing what to do in each different context that the strategy can distinguish. The genetic algorithm simulates the evolutionary process by mutating the chromosomes and by allowing sexual reproduction to recombine features from two different strategies. It then subjects the resulting offspring to competition with other programs in the population. Axelrod (*64*) used the genetic algorithm and a chromosome of 70 genes to study the evolution of strategies in a fixed environment composed of eight representative strategies from his second computer tournament. From a random start, strategies similar to TFT often evolved within a few dozen generations. Strategies sometimes evolved that outperformed TFT in this environment. These well-adapted strategies were not nice: they defected initially to discriminate between the representatives they were facing so that they could exploit those that were exploitable and cooperate with the others.

These two methods of overcoming the limitations of an ecological simulation approach are complementary. The differential equations approach allows a relatively complete analysis of the interactions among a small set of selected strategies or strategy types. The genetic algorithm approach provides a way of exploring a potentially huge strategy space, making possible the discovery of new strategies not previously specified.

Population Structure

As noted earlier, the initiation of cooperation in a population of noncooperators was possible if the cooperative strategies invaded in clusters (*2*, p. 63–68). Can cooperation evolve without the presence of population structure?

The theoretical results on ESS supply one answer. Because no strategy is ESS if the future is important enough, cooperation may be started without population structure if the "correct" combination of strategies is present (*60, 61*). This also undoes the "ratchet effect" discovered by Axelrod. Boyd and Lorberbaum (*61*) suggest the following example. Suppose that a population consists of TFT, SUSPICIOUS TFT (or STFT, exactly like TFT but defects on the initial move) and TIT FOR TWO TATS (or TF2T, which defects only if the other player defected on the two preceding moves). STFT and TF2T will end up cooperating after the first move, STFT and TFT will each continue cooperating while the other defects (and vice versa), and TFT and TF2T will cooperate on every move. In such a situation, TF2T will be able to invade both TFT and STFT, even without clustering. This example would not be so startling (since a cooperative strategy is merely invading another cooperative strategy), except that TF2T is easily invaded by a strategy which initially defects and then alternates cooperation and defection (*65*).

Other attempts at minimizing the need for clustering in order to initiate cooperation have relied on certain informational requirements. Examples are positive assortative mating and meeting (*66*). Intuitively, cooperative strategies "pick" fellow cooperators and shun noncooperators, an outcome that is quite similar to ostracism.

Informational requirements may not be necessary, however, if the payoffs reflect certain behavioral characteristics of the game. For instance, Peck and Feldman (*67*) prove that cooperation can get started in a population of defectors if the payoffs to cooperative acts depend on the frequency with which those acts are performed. And Feldman and Thomas (*46*) have found that in the case where the probability of another interaction depends on the choices of the players, TFT can increase when rare, although this does not guarantee that TFT will increase to fixation. In particular, in a population of mostly ALL D, where w is the probability that a player continues to interact after choosing cooperation and u is the probability that a player continued to interact after defection, a rare TFT trait can increase if $S(1 - u) > P(1 - w)$.

Work on initial viability of cooperative strategies has mixed implications for the evolution of cooperation. On the plus side, initial viability of cooperation may not require the assumption of population structure. On the minus side, however, such theoretical results imply that cooperative strategies need not be stable against invading ensembles of strategies.

Conclusion

Research has shown that many of Axelrod's findings (*2*) can be generalized to settings that are quite different from the original two-player iterated Prisoner's Dilemma game. Our review has identified the following extensions. Increasing the number of players who simultaneously interact tends to make cooperation more difficult. The results on the importance of reciprocity and the shadow of the future apply not only to the iterated Prisoner's Dilemma, but also to Chicken and more generally to any game in which individually rational choices lead to suboptimal decisions. Altering the set of strategies to include nonsimultaneous play and options for exit and ostracism does not disturb Axelrod's conclusions. Several solutions to the problem of noise and misperception have been suggested; the prescriptions entail increased generosity for small amounts of noise. However, for larger amounts of noise, increased generosity may invite exploitation. The empirical relevance of the shadow of the future in inducing cooperation has been demonstrated, and theoretical researchers have been pushing beyond the IPD formulation to derive cooperation for the finitely repeated Prisoner's Dilemma. The sophistication of evolutionary models of cooperation has been advanced by the introduction of full population dynamics and by the use of the genetic algorithm for the discovery of entirely new strategies. Work in theoretical biology has demonstrated that no strategy is evolutionarily stable in the iterated Prisoner's Dilemma if the shadow of the future is long enough and multiple mutants occur. This also implies that it is possible for cooperation to begin without a structured population.

Further exploration of the evolution of cooperation can profit from the greater specification of the pattern of interactions and the process by which those interactions themselves evolve. Research along such lines will not only help unite theoretical and empirical work even further than has so far been possible, but may also deepen our understanding of the initiation, maintenance, and further evolution of cooperation.

REFERENCES AND NOTES

1. R. Axelrod, *J. Conflict Resolut.* **24**, 389 (1980); *ibid.*, p. 379; *Am. Polit. Sci. Rev.* **75**, 306 (1981).
2. ______, *The Evolution of Cooperation* (Basic Books, New York, 1984).
3. ______ and W. Hamilton, *Science* **211**, 1390 (1981).
4. P. Huth, *Extended Deterrence and the Prevention of War* (Yale Univ. Press, New Haven, CT, in press).
5. G. Wilkinson, *Nature* **308**, 181 (1984).
6. ______, *Behav. Ecol. Sociobiol.* **17**, 111 (1985).
7. R. M. Seyfarth and D. L. Cheney, *Nature* **308**, 541 (1984).
8. L. W. Buss, *Science* **213**, 1012 (1981).
9. M. Milinski, *Nature* **325**, 433 (1987).
10. M. P. Lombardo, *Science* **227**, 1363 (1985).
11. E. Nol, *Behavior* **95**, 232 (1985).
12. H. R. Pulliam, G. H. Pyke, T. Caraco, *J. Theor. Biol.* **95**, 89 (1982); R. L. Mumme and A. De Queiroz, *Behavior* **95**, 290 (1985).
13. S. B. Katz, *Southern Calif. Law Rev.* **60**, 509 (1987).
14. A. Schepard, *Texas Law Rev.* **64**, 687 (1985).
15. S. Brown, *Negotiation J.* **2**, 371 (1986).
16. A. Dixit, in *U.S. Trade Policies in a Changing World*, R. M. Stern, Ed. (MIT Press, Cambridge, MA, 1987), p. 245.
17. J. L. Goldstein and S. D. Krasner, *Am. Econ. Rev.* **74**, 282 (1984).
18. Axelrod (*2*) termed a strategy that is in Nash equilibrium with itself is called a collectively stable strategy.
19. M. Taylor, *Anarchy and Cooperation* (Wiley, New York, 1976).
20. N. V. Joshi, *J. Genet.* **66**, 69 (1987).
21. J. A. K. Cave "A formal analysis of the evolution of cooperation" (Rand Paper Series P-7031, Santa Monica, CA, 1984).
22. J. Bendor and D. Mookherjee, *Am. Polit. Sci. Rev.* **81**, 129 (1987).
23. P. Pettit, *J. Philos.* **83**, 361 (1986).
24. P. Fader and J. R. Hauser, *J. Conflict Resolut.* **32**, 553 (1988).
25. A. H. Harcourt, in *Comparative Socioecology of Mammals and Humans*, V. Standen and R. A. Foley, Eds. (Blackwell, Oxford, in press).
26. K. Aoki, in *Population Genetics and Molecular Evolution*, T. Ohta and K. Aoki, Eds. (Springer-Verlag, Berlin, 1985), p. 429.
27. R. Axelrod, *Am. Polit. Sci. Rev.* **80**, 1095 (1986).
28. J. M. Orbell, P. Schwartz-Shea, R. T. Simmons, *ibid.* **78**, 147 (1984).
29. D. Hirshleifer and E. Rasmusen, *Cooperation in a Repeated Prisoners' Dilemma with Ostracism* (Graduate School of Management, UCLA, 1988).
30. C. Barner-Barry, *Ethol. Sociobiol.* **7**, 281 (1986).
31. T. Kondo, *Int. Relat. Stud.* **5**, 3 (1986).
32. B. J. Lipman, *Public Choice* **51**, 315 (1986).
33. R. Dacey and N. Pendegraft, *Int. Interactions*, in press.
34. An outcome is individually rational if it provides each player with at least that payoff below which the player cannot be forced by the other players.
35. For a review of work since 1976, see D. Fudenberg and E. Maskin, *Econometrica* **54**, 533 (1986).
36. P. Molander, *J. Conflict Resolut.* **29**, 611 (1985).
37. J. Bendor, *Am. J. Polit. Sci.* **31**, 531 (1987).
38. J. Gowa, *Int. Org.* **40**, 167 (1986); R. Axelrod and R. Keohane, *World Polit.* **38**, 226 (1985).
39. C. Donninger, in *Paradoxical Effects of Social Behavior*, A. Diekmann and P. Mitter, Eds. (Physica-Verlag, Heidelberg, 1986), p. 123.
40. R. Sugden, *The Economics of Rights, Co-operation and Welfare* (Basil Blackwell, New York, 1986); P. S. Segerstrom, *J. Econ. Theor.*, in press.
41. U. Mueller, *J. Conflict Resolut.* **31**, 692 (1987).
42. R. Calvert, *Reciprocity Among Self-Interested Actors: Asymmetry, Uncertainty and Bargaining* (Carnegie-Mellon University and Washington University, Pittsburgh, 1985).
43. E. J. Green and R. H. Porter, *Econometrica* **52**, 87 (1984).
44. R. Keohane, *After Hegemony* (Princeton Univ. Press, Princeton, 1984).
45. I. Eshel and D. Weinshall, *Theor. Popul. Biol.*, in press.
46. M. W. Feldman and E. R. C. Thomas, *J. Theor. Biol.* **128**, 297 (1987).
47. R. D. Luce and H. Raiffa, *Games and Decisions* (Wiley, New York, 1957).
48. A. Neyman, *Econ. Lett.* **19**, 227 (1985); N. Megiddo and A. Wigderson, "On play by means of computing machines" (IBM Research Report RJ 4984, Computer Science, Yorktown Heights, NY, 1986).
49. A. Rubinstein, *J. Econ. Theor.* **39**, 83 (1986).
50. D. M. Kreps, P. Milgrom, J. Roberts, R. Wilson, *ibid.* **27**, 245 (1982).
51. R. Selten and R. Stoecker, *J. Econ. Behav. Organ.* **7**, 47 (1986).
52. I. McLean, *Public Choice* (Basil Blackwell, New York, 1987).
53. K. Aoki, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4065 (1983).
54. G. Pollock, *Am. J. Phys. Anthropol.*, in press.
55. C. H. Moore, *Am. Polit. Sci. Rev.* **81**, 201 (1987).
56. J. K. Murninghan and A. Roth, *J. Conflict Resolut.* **27**, 279 (1983).
57. N. B. Davies and A. I. Houston, *J. Anim. Ecol.* **55**, 139 (1986).
58. J. Maynard Smith and G. R. Price, *Nature* **246**, 15 (1973).
59. W. G. S. Hines, *Theor. Popul. Biol.* **31**, 195 (1987).
60. P. Pudaite, *On the Initial Viability of Cooperation* (Merriam Laboratory for Analytic

Political Research, University of Illinois, Urbana–Champaign, 1985).
61. R. Boyd and J. P. Lorberbaum, *Nature* **327**, 58 (1987).
62. M. C. Blad, *Int. J. Game Theory* **15**, 83 (1986).
63. J. Holland, *Adaptation in Natural and Artificial Systems* (Univ. of Michigan Press, Ann Arbor, 1975).
64. R. Axelrod, in *Genetic Algorithms and Simulated Annealing*, L. Davis, Ed. (Pitman, London, 1987), p. 32.
65. R. M. May, *Nature* **327**, 15 (1987).
66. L. L. Cavalli-Sforza and M. W. Feldman, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2017 (1983); I. Eshel and L. L. Cavalli-Sforza, *ibid.* **79**, 1331 (1982).
67. J. R. Peck and M. W. Feldman, *Am. Nat.* **127**, 209 (1986).
68. We thank the anonymous reviewers for helpful suggestions. Supported by the MacArthur Foundation through a grant to the University of Michigan and by the NSF grant SES-8808459.

Experimental Constraints on Theories of High–Transition Temperature Superconductors

W. A. LITTLE

Recent experiments have revealed several key features of the unique nature of the new, high–transition temperature cuprate superconductors. These results provide an easily understandable, physical picture of the structure and behavior of the charge carriers in these materials, and point to the mechanism responsible for their existence. These experiments are now placing strong constraints on possible theoretical models of the phenomenon.

THE RECENT DISCOVERIES OF SUPERCONDUCTIVITY IN $La_{2-x}Sr_xCuO_4$ (*1*) with a transition temperature, T_c, above 30 K, and of superconductivity in $YBa_2Cu_3O_7$ (*2*) above 90 K, have come as a shock to most physicists familiar with superconductivity. These cuprate ceramics superconduct at temperatures many times higher than any other materials. Superconductivity had been discovered in 1911 by Kammerlingh Onnes in mercury at 4.2 K, but prior to 1985 the highest transition temperature observed among the thousands of alloys prepared was only 23.2 K. The physics community has now been presented with an existence proof of high-temperature superconductivity. In response, a flood of theories have come forth, which range from modest additions to the Bardeen, Cooper, and Schrieffer (BCS) theory (*3*) which had explained so successfully conventional superconductors, to theories suggesting the existence of a totally new type of metal. It is not yet clear, even to the experts, how valid or applicable many of these are. It is even more difficult for the nonexpert to appreciate the subtleties of this complex phenomenon.

While the theorists were rethinking the problems of superconductivity, the experimentalists have been busy. A huge amount of work has been done during the past 18 months studying the electronic, magnetic, thermal, structural, and optical properties of these materials. Included in this study are a few basic experiments, which have established some strong constraints on the possible theoretical explanation of the phenomenon. My purpose is to present these and to discuss their implications from a general point of view for physicists, chemists, materials scientists, and molecular biologists. I will begin by considering the nature of the charge carriers and then show how the study of these reveals something of the interactions responsible for the superconductivity.

The Charge Carriers

For a material to exhibit the essential features of superconductivity—persistent currents, perfect diamagnetism, or quantum interference behavior—a finite fraction of all the charge carriers must be in the same quantum state (*4*). The reason is that this single state must make a significant contribution to the total free energy of the system, which, in turn, requires that it have a macroscopic occupation.

The charge carriers in a normal metal are electrons, which obey the Pauli exclusion principle. One and only one such particle can be in any one state at a time. So the charge carriers in the superconducting state cannot be single electrons but must be composite particles, of an even number of electrons. These then are bosons, and obey Bose-Einstein statistics, which allows an arbitrary number of particles to be in the same state. The wave function of such a composite particle is a linear combination of products of single-particle states. Many such composite particles can be in the same state because, although each is described by the same linear combination, the same terms in the linear combination in different particles are never occupied by electrons at the same time. Thus the exclusion principle is obeyed by the individual electrons but the composite particle as a whole behaves as a boson.

In conventional superconductors the bosons are pairs of electrons—the "Cooper Pairs" of the BCS theory. This was established by the beautiful flux quantization experiments of Deaver and Fairbank (*5*) and Doll and Nabauer (*6*) in 1962. In these experiments it was shown that the magnetic flux trapped in a hollow superconducting cylinder was an integral multiple of a fundamental unit of flux, $hc/2e$. Here h is Planck's constant, c the velocity of light, and e the charge of the electron. The presence of the factor of 2 in the denominator shows that the carriers are pairs. Similar and related experiments on a large number of conventional superconductors show, without exception, that the charge carriers in these also are pairs. Are the charge carriers in the cuprates pairs, as in the BCS theory, or quartets or more complex structures?

A clean and elegant answer to this question was given by a flux quantization experiment done by Gough *et al.* (*7*) in Birmingham. The values of the flux trapped in a superconducting ring of

The author is a professor in the Physics Department, Stanford University, Stanford, CA 94305.