The Power of Reciprocity

FAIRNESS, RECIPROCITY, AND STAKES IN VARIANTS OF THE DICTATOR GAME

ANDREAS DIEKMANN
Swiss Federal Institute of Technology

In two experiments, the existence and extent of altruistic reciprocity is explored in the context of a simple experimental game, "the sequential dictator." Findings show that altruistic reciprocity is frequent and robust, and the reciprocity norm does not erode if stakes are raised. Implications of the findings for social theory and further empirical research are discussed.

Keywords: *reciprocity; fairness norms; altruism; dictator game; behavioral game theory*

The topic of reciprocity has long played an important role in anthropology, ethnology, and sociological thinking, for instance, in the classic writings of Georg Simmel (1950), Bronislaw Malinowski (1926), Marcel Mauss (1950/1990), or Alvin Gouldner (1960). More than four decades ago, Gouldner clarified the concept and its dimensions and assumed the existence of a universal norm of reciprocity in a well-known article. Recently, new interest in the issue of reciprocity has grown in sociology and political science in the context of the vivid debate about "social capital." Although there is a long tradition in sociology of research on reciprocity, a systematic theory leading to empirically testable predictions is still lacking. In economics and game theory, on the other hand, reciprocity was incorporated into rigorous models and has given rise to a more coherent theoretical perspective. However, the standard economic approach has the weakness that it cannot account for altruistic reciprocity or compliance to reciprocity norms in unrepeated interactions. Yet field studies and experimental studies alike support the existence of a norm of reciprocity for a wide array of social activities and even among strangers. Recent developments in experimental game theory try to account for these observations contradicting the standard approach.

Building on models from "behavioral game theory" (Camerer 1997), I will explore some of the core hypotheses of reciprocity theory using experimental data. In the sec-

AUTHOR'S NOTE: I would like to thank Elisabeth Coutts, Axel Franzen, Gabriela Koenig, Rachel Matthey, Anatol Rapoport, Carmen Tanner, and Manuela Vieth for their helpful comments. I am also grateful to Regula Bieri who performed the experimental work with the assistance of Andrea Hungerbuehler. Experimental data are available for replication on the author's homepage: http://www.socio.ethz.ch/de/diekmann/. The data used for this study are available at www.yale.edu/unsy/jcr/jcrdata.htm/.

JOURNAL OF CONFLICT RESOLUTION, Vol. 48 No. 4, August 2004 487-505
DOI: 10.1177/0022002704265948
© 2004 Sage Publications

ond section, I sketch some new developments in reciprocity theory. The third section describes the method and the data, and the fourth section reports on the empirical results. In the final section, I discuss implications of the results for reciprocity theory and further empirical research.

THEORIES OF RECIPROCITY

SOCIAL CAPITAL

Whereas Gouldner (1960, 172) was concerned with the functionalist argument that reciprocity promotes the stability of a social system, social capital theory shifted the attention to the effects of reciprocity and trust on cooperation and economic prosperity. *Reciprocity* and *social capital* are close associates. I use the term *social capital theory* with some reservations. At least until now, there has been no measurable "rate of return" on social capital as there is for physical and human capital. Unlike theories of physical capital and human capital, social capital theory is not a deductive set of propositions but a heuristic framework of more or less precise hypotheses. The notion of social capital was introduced 100 years ago and reinvented or readapted by Bourdieu (1983), Coleman (1988), and others (see Putnam 2001 for a short history). Until now, a great variety of definitions of social capital has been suggested (for a synopsis of definitions, see Freitag 2001). Some of them explicitly incorporate the norm of reciprocity. For example, Putnam (2001, 21) defines the concept as follows: "social capital—that is, social networks and the associated norms of reciprocity." In Fukuyama's (2000) norm-oriented approach, reciprocity is a basic element. "The norm that constitutes social capital can range from a norm of reciprocity between friends, all the way to complex and elaborately articulated doctrines like Christianity or Confucianism" (p. 3). Both adherents to social capital theory believe that the constituent norms of social capital enhance cooperation. Of course, cooperation does not always have a positive impact on the welfare of the whole society as the example of the mafia demonstrates. However, if norms of reciprocity transcend family relations and are applied to members of closed groups and outsiders alike, these norms will reduce transaction costs that facilitate social exchange and market activities. By this and other effects of reciprocity and trust, social capital is expected to be positively associated with economic growth. More elaborate theoretical arguments for the prosperity-fostering effect of social capital and some empirical evidence are provided by Knack and Keefer (1997) and Whiteley (2000).

In discussing the effects of social capital, it is usually argued that norms of reciprocity, trust, social networks, and other elements, which constitute social capital in the respective definition, have an effect on certain variables, such as economic growth, the crime rate, political participation, or other aggregated social statistics. Hence, the relation between social capital and some other variable is explained in terms of the underlying components of social capital. I assume that this is the reason for the theoretical deficiency of the social capital framework. Social capital is a conglomerate of several components whereby the bundles of components vary from definition to definition of

respective authors. This deficiency may be eliminated if one disentangles the components and formulates hypotheses on the level of the constituent elements. Hence, I expect greater theoretical progress in social capital research if one focuses more precisely on the basic elements. Obviously, reciprocity is a central element.

ASPECTS OF RECIPROCITY

One should distinguish among several aspects of the concept of reciprocity. In his "preliminary statement," Gouldner (1960) already made important distinctions. A first distinction is between conditional and unconditional generosity. For example, the moral of "turning the other cheek" in the New Testament refers to the latter aspect (p. 164). If someone voluntarily shares a cake with somebody else, he or she adopts an approach of unconditional fairness. Reciprocity, on the other hand, is conditional fairness. Moreover, reciprocity is not a categorical—not an "all or none matter" (p. 164)—but a dimensional variable. Alter may reciprocate ego's action to a greater or lesser degree. These are two features of "reciprocity as a pattern of mutually contingent exchange" (p. 161). Besides the pattern of exchange, there is a norm of reciprocity that "evokes obligations toward others on the basis of their past behavior" (p. 170). As a first aspect, the norm can refer to heteromorphic or homomorphic reciprocity. In the former case, a good or service is paid back by a different good or service of equal value ("tit-for-tat"). In the latter case, a good or service is reciprocated by exactly the same good or service ("tat-for-tat") (p. 172). Second, the norm of reciprocity does not apply only to benevolent actions. Reciprocity may be positive or negative (p. 172). "An eye for an eye" in the Old Testament clearly is negative (and homomorphic) reciprocity. Negative sanctioning of norm violators, even if sanctioning is costly, is often stimulated by a norm of reciprocity. Hence, negative reciprocity, too, is an important mechanism for promoting cooperation (Fehr and Gächter 2002). Gouldner (1960, 173) further indicates that reciprocal behavior might be driven by egoistic motives—"if you want to be helped by others you must help them." This leads to another important distinction between altruistic and egoistic reciprocity.

RECIPROCITY IN REPEATED SITUATIONS

Since the publication of Gouldner's (1960) article, research in biology (Trivers 1971) and game theory (e.g., Friedman 1977; Taylor 1976; Axelrod 1984) has been able to demonstrate that purely egoistic and rational actors may mutually adopt conditional cooperative strategies such as reciprocal behavior in their own interest. Trivers (1971) had already developed a model on the "evolution of reciprocal altruism" in accordance with Darwin's theory of evolution. For example, his model explains the cooperation of nonkin symbiotic organisms by repeated interactions.[1] Similarly, in game theory the important structural condition is the repetition of the exchange situation (or social dilemma) where actors in each round face the decision to cooperate or to

1. Trivers's (1971) well-known example is the "cleaning symbioses" of a cleaning fish and a host fish. "The host's altruism is to be explained as benefiting him because of the advantage of being able quickly and repeatedly to return to the same cleaner" (p. 43).

try to exploit their coplayer. More technically, it was proven that in a situation of infinitely repeated social dilemmas, conditional cooperative strategies under certain circumstances combine to a (pareto-optimal) Nash equilibrium (e.g., Fudenberg and Tirole 1991, 150). The latter is the term for a situation where none of the actors have incentive to change their course of action as long as the other actors stick to their choice of strategy. (For example, as with driving on the left, if all other actors employ this strategy, no single actor has an incentive to deviate from the strategy of driving on the left. Hence, the strategy "driving on the left" is a Nash equilibrium, obviously not a unique one.) In Axelrod's (1984) well-known computer tournaments, the strategy of tit-for-tat, sent in by Anatol Rapoport, won the first prize twice and the "ecological tournament" in addition.[2] In Axelrod's terms, if "the shadow of future" (measured by a discount parameter) is large enough, reciprocal cooperation is in the interest of purely egoistic and rational actors. In such a situation, "reciprocity as a pattern of mutually contingent exchange" (Gouldner 1960, 161) may evolve among self-interested actors without the existence of a norm of reciprocity. Also, it can be seen that Gouldner's distinction of a manifest "pattern of reciprocity" from the concept of a "norm of reciprocity" is conceptually fruitful. However, even if a norm of reciprocity is not a necessary precondition to establishing reciprocal cooperation, such a norm may evolve from a pattern of reciprocity and also may greatly facilitate the stability of reciprocal exchange.

INDIRECT RECIPROCITY

In the Axelrod tournament, participants played the repeated prisoner's dilemma in dyads, not knowing in advance when the game would end. Reciprocal cooperation is, however, more fragile if larger sets of actors rather than dyads play the repeated game simultaneously. Although in principle there are (many) Nash equilibria of conditional cooperative strategies in an *n*-person social dilemma, the equilibrium is difficult to attain and may decay rapidly (Boyd and Richerson 1988). A new solution to the problem of cooperation in larger groups was presented by Nowak and Sigmund (1998). Here, an actor is assigned to a coplayer for a single round, and his or her decision is observed by other group members with a certain probability. For example, an actor has the choice to give or not to give to a person in need (Wedekind 1998). Nowak and Sigmund assume that actors cumulate an "image score" for each cooperative decision. The higher the image score—a reputation for cooperation—the more other actors are inclined to cooperate with the reputed actor, even if they had no relation in the past. Under this condition of "indirect reciprocity," Nowak and Sigmund proved that cooperation will emerge if certain requirements regarding information about the image score are met. Wedekind and Milinski (2000) conducted a sophisticated experiment confirming empirically that indirect reciprocity promotes cooperation via image scoring. Interestingly, Putnam (2001), probably unaware of the theory of indirect reciprocity, emphasized the importance of a "norm of generalized reciprocity," meaning that "I'll do this for you without expecting anything specific back from you, in the confi-

2. For the evolution of tit-for-tat in an experimental context, see Rapoport and Chammah (1965).

dent expectation that someone else will do something for me down the road." He added that "generalized reciprocity is more efficient than a distrustful society, for the same reason that money is more efficient than barter" (p. 21). Putnam's "generalized reciprocity" very much resembles the notion of indirect reciprocity. Nevertheless, besides conceptual similarities, the strength of the theory of indirect reciprocity is that it offers a deeper explanation for the evolution of cooperation in larger groups. Again, norms of generalized reciprocity may support the evolution of cooperation in large groups. However, according to the theory of indirect reciprocity, an internalized norm is not a necessary condition to promote cooperation.

In sum, theoretical progress in game theory provides explanations for the fact that, under certain conditions, selfish actors will choose strategies of reciprocity, which in turn lead to cooperation. This may be true in repeated dyadic games, and it may be true in larger groups via the mechanism of indirect reciprocity.

RECIPROCITY IN NONREPEATED INTERACTIONS

Yet there are many situations where both positive and negative reciprocal behavior is observed that cannot be explained in terms of strategic and far-sighted self-interest. For example, a person acts favorably toward a stranger, and the stranger reciprocates the favor although it is unlikely that they will ever meet again. Also, the likelihood that friends, colleagues, and relevant others will hear about the act of charity is zero. A visitor abroad honors the excellent service in a restaurant with a large tip although he knows that he will never return to the place. Apart from observations of everyday life, experiments provide evidence for altruistic reciprocity. In a carefully arranged experiment by Regan (1971), subjects who received a favor were strongly inclined to return the favor, while positive feelings toward the donor were of much less importance. More recently, Fehr, Fischbacher, and Tougarova (2002) provided support for Akerlof and Yellen's (1988) "fair wage hypothesis" with an experiment with a "gift-exchange market." Subjects in the role of workers reciprocated "fair" wages by choosing more effort, and subjects in the role of employers granted higher wages above market equilibrium. Even for high stakes, the pattern of cooperation did not differ from the behavior in the low-stake condition. Altruistic reciprocity is a key element of the "fair wage/effort hypothesis" based on work from sociological exchange theory and psychological equity theory (Akerlof and Yellen 1988).

Gouldner (1960, 176) has already pointed out that "the employer may pay his workers not merely because he has contracted to do so; he may also feel that the workman has earned his wages." Or the employer may be aware that higher wages induce workers to expend more effort—like Henry Ford, who doubled wages on January 5, 1914, with the result of a productivity increase from 40% to 70% and an increase in profits of 20% (Borjas 1996, 427). There is also much evidence of altruistic negative reciprocity. People complain about norm violations and unfair treatment or exercise revenge even if there are considerable costs associated with sanctioning. Experiments with the ultimatum game (Güth, Schmittberger, and Schwarze 1982; Thaler 1992) and experiments on sanctioning behavior (Fehr and Gächter 2002) consistently demonstrate that victims treated unfairly sacrifice material payoffs to punish offenders.

ALTRUISTIC RECIPROCITY AND BEHAVIORAL GAME THEORY

Standard game theory and neoclassical economics cannot account for altruistic reciprocity, that is, a responder who is fully aware that the sequence of interactions will not continue but nevertheless returns a favor or employs a negative sanction that reduces his own material payoff. Thus, reciprocity is observed in "one-shot" games. Gouldner (1960) and others explain altruistic reciprocity through an internalized obligation: the norm of reciprocity. Although Axelrod's (1984) "shadow of the future" explains reciprocity in repeated interactions, Gouldner (1960, 174) coined the concept of a "shadow of indebtedness," which clearly refers to a moral obligation, to the altruistic compliance with a norm of reciprocity. Since then, a new branch in experimental game theory, called "behavioral game theory" (Camerer 1997), has sought to develop more rigorous and systematic explanations to bridge the gap between theory and empirical observations. Two approaches are of relevance here: first, the models of inequality aversion proposed by Bolton and Ockenfels (2000) and Fehr and Schmidt (1999), and second, the "fairness equilibrium" model introduced by Rabin (1993). The basic idea of the former models is the assumption of a utility function, which is dependent on material payoffs as well as on some measure of distance to a fair reference payoff. In Bolton and Ockenfels' "theory of equity, reciprocity, and competition" (ERC), the reference payoff is an equal share of the total material payoff. Actors who gain more or less than their fair "share of the cake" are punished, that is, their utility will be reduced. Although ERC theory is formalized more generally by a set of axioms, the following special utility function (or "motivation function," as it is called by the authors) demonstrates the argument:

$$u_i = a_i y_i - b_i(1/2 - \sigma_i)^2. \tag{1}$$

This additive-separable utility function for $n = 2$ players was proposed as an example by Bolton and Ockenfels (2000). Here, subscript $i = 1, 2$ denotes the player, u_i is player i's utility, y_i is the material (for example, monetary) payoff, σ_i is player i's share ($\sigma_i = y_i/(y_1 + y_2)$), and a_i and b_i are individual-specific parameters.

It can be seen that actors have to balance material payoff and inequality aversion. In the two-player interaction, the reference point is the equal share of 1/2. Deviations of the actual share σ_i from the equal share are punished. Parameter a_i is the weight for selfishness, whereas b_i is a measure of the strength of player i's motive of inequality aversion ($a_i \geq 0$, $b_i > 0$).

Fehr and Schmidt (1999) suggest an alternative utility function. In the "theory of fairness, competition, and cooperation" (FCC), player's utility is diminished by the weighted sum of distances to richer actors and the weighted sum of distances to poorer actors. The weight for the former distance ("disadvantageous inequality") is assumed to be larger than the weight for "advantageous inequality."

Both theories incorporate material payoffs and the loss of unfair appropriation of money or goods as arguments in the utility function. The theory then deals with the utilities as in the standard game theory; that is, it predicts that the actors' choice is a subgame-perfect Nash-equilibrium strategy. Although predictions of the two variants

of theories of inequality aversion differ for certain situations (and, therefore, it would be possible to discriminate between the theories with empirical data), both theories account for a large body of empirical observations. For example, the theories predict that a certain proportion of actors chooses cooperation in a single-shot prisoner's dilemma, contribute to a public good, reciprocate in a gift-exchange market, or reject offers in an ultimatum game.

The theories of inequality aversion do not assume a norm of reciprocity. However, under certain conditions, the theories imply altruistic positive and negative reciprocal behavior. For example, in my experiment, I use a simple variant of the dictator game, which I call "sequential dictator" (see below). In contrast to standard game theory, the theories of inequality aversion predict that in a sequential dictator game, subjects may exhibit reciprocal behavior to a certain degree.

EROSION OF THE NORM OF RECIPROCITY IF STAKES ARE HIGH

Imagine the situation of altruistic reciprocity. Actors are in conflict to win a material gain or to bear a loss and follow the norm of reciprocity. It is expected that a large proportion of actors will follow the norm if the material loss is low. However, will people comply to a norm of reciprocity if stakes are high? Rabin's (1993) model of a "fairness equilibrium" explicitly assumes that stakes matter. The logic of the Rabin model is somewhat different from ERC theory and FCC theory. Rabin assumes that *Ego* forms a belief about the intentions of his coplayer. Without going in technical details, it is sufficient to say that these beliefs determine a fairness payoff that is added or subtracted from the material payoff. The fairness payoff is defined such that it is positive if *Ego* believes *Alter* will act kindly, and it is negative if he or she expects an unkind decision. In the former case, the (positive) fairness payoff is enhanced if *Ego* behaves kindly, and in the latter case, the (negative) value of the fairness payoff is reduced by punishing *Alter*'s (believed) unkind behavior. Hence, Rabin's formula takes into account that positive or negative reciprocity will increase *Ego*'s fairness payoff. As in the ERC model and the FCC model, there is a trade-off between the material payoff and the fairness payoff. In Rabin's model, because the fairness payoff is bounded, it follows that the higher the material payoff, the more it will dominate the fairness payoff. In the limit for increasingly high stakes, only material payoffs matter, and actors will choose Nash-equilibrium strategies in material payoffs. With increasing payoffs, the actor's behavior converges to the "ideal type" of a *homo economicus*. For example, for a nonrepeated prisoner's dilemma game, there will be two fairness equilibria if stakes are low, one in cooperative and one in noncooperative ("defective") strategies. If stakes are high, there will be only one fairness equilibrium, which is identical to the Nash equilibrium of defective strategies. Thus, Rabin's theory predicts that actors will increasingly opt for noncooperation if stakes increase. Thus, Rabin restricts self-sacrificing reciprocity to the region of low-cost decisions, whereas *homo economicus* rules in the sphere where decisions lead to high gain or loss.

In the following, I propose a simple design to test for altruistic reciprocity effects. In a first experiment, the degree of reciprocity is ascertained. I ask, also, whether empirically observed reciprocity effects match the predictions of behavioral game the-

ory. Main emphasis is given to a second experiment in which the hypothesis of stake effects is explored. In this experiment, Gouldner's (1960) assumption of a universal norm of reciprocity is confronted with the hypothesis that the norm of reciprocity will erode if stakes are high. In an additional replication of a field experiment, the reciprocity hypothesis is applied to a real life example.

EXPERIMENTAL DESIGN AND METHOD

"SEQUENTIAL DICTATOR"

In many sequential decision situations, an actor's choice is driven by the motive to reciprocate a favorable or unfavorable action of his or her coplayer. For example, in an ultimatum game, the proposer suggests a certain division of "the cake," and the responder has the veto power to accept or to reject the division. If he or she rejects, both actors earn nothing. So the responder is able to exercise negative reciprocity if he or she believes the proposer's decision to be unfair. Rejection is self-sacrificing reciprocity because the responder earns less compared to the acceptance of an unfair offer. The motive of positive altruistic reciprocity is involved in a sequential prisoner's dilemma or in a trust game (Dasgupta 1988). In the latter game, the trustor has the option to transfer a certain amount or good to the trustee, who in turn faces the choice to cooperate or to exploit the trustor. Cooperation by the trustee contains elements of altruistic positive reciprocity because he reciprocates the trustor's cooperation, thereby gaining lower material payoffs than by behaving opportunistically. The problem is that reciprocity is not the only reason for cooperation. Honoring trust; attaining an "efficient," that is, a pareto-optimal payoff; or receiving the maximal collective payoff may be rival motives in the decision process. Although it is likely that positive altruistic reciprocity is involved, one cannot equate cooperation with reciprocity in a trust game. Thus, to explore effects of reciprocity, I suggest a simple alternative design. I choose a variant of the dictator game that I call "sequential dictator." In a first round, the dictator (actor 1) decides how to share a good or an amount z of money ("the cake") with his coplayer (actor 2). In the second round, both players change roles, such that the new dictator (actor 2) has to divide a (new) cake z with the former dictator (actor 1). In the simple version of sequential dictator, which I will use here, the cake in round 2 is of the same size z as in round 1, and the game ends after two rounds.[3] Player 1's behavior is programmed by the experimenter, whereas real subjects are always in the position of player 2. By this design, responses of subjects to various divisions of z can be investigated in a controlled experiment. Sequential dictator is a very simple and easy to understand decision situation. It is, so to say, a minimal social situation to study reciprocity.

3. I would like to thank James Walker for the suggestion to vary the size of the cake or to multiply the sacrifice to player 2 by $m > 2$. In the latter case, the game would be similar to a trust game (Dasgupta 1988), and players would have to deal with an efficiency problem. However, for the present study, I preferred to stick to a minimal social situation to study reciprocity.

Denote by $x_1$the proportion of the cake player 1 is willing to sacrifice to player 2. The offer of player 2 to player 1 is denoted by x_2. The result is that player 1 will earn $y_1 = z(1 - x_1) + zx_2$, and player 2 gains $y_2 = zx_1 + z(1 - x_2)$.

Both variants of behavioral game theory, the ERC model and the FCC model, predict that player 2 will never offer more than he or she received in round 1. By paying back more than was received, player 2 will reduce his or her material payoff, and in addition, player 2 has a utility loss by increasing inequality. Independent of parameters in the utility or motivation function for $x_2 > x_1$, player 2 can always increase his or her utility by reducing x_2. On the other hand, in the region $x_1 \leq x_2$, player 2 has to balance between a material loss and a gain by decreasing inequality. This can be inferred from equation 1 and follows in general from the ERC model. The result of the trade-off depends on the individual and unobserved parameters (the weights for selfishness and inequality aversion a_i, b_i). Hence, the model allows for reciprocal behavior up to the point of full reciprocation.[4]

DESIGN AND METHOD

Participants in experiments 1 and 2 were university students, whereas the field experiment was based on a subsample of a general population survey. The university registration office provided addresses of students with various backgrounds, and students were asked via postal mail to participate in an experiment. Appointments were made with those who agreed to participate. Participants were gathered in one classroom in groups of five to eight. The experimenter then distributed written general instructions on the course of the experiment.

A crucial point was maintenance of anonymity. Participants were assured that their identity would not be disclosed to their coplayers. In addition, participants' decisions were not disclosed to the experimenter. To guarantee double anonymity, the participants were asked to select and write down a six-digit number. This number had to be written on the questionnaire(s) that was to be placed in a voting box later. After finishing the experiment, each participant had to pick up his or her payoff in the secretary's office in an envelope with his or her specific code number.

After the general instructions, the experimenter distributed a questionnaire explaining the decision situation. The instructions in the questionnaire emphasized that the participant was assigned to a coplayer by lot and that he or she would interact with the coplayer only once. For the purpose of anonymity, the coplayer was in a different room, and his or her coplayer's decision was transmitted in a letter. In fact, participants were always in the position of player 2, the coplayer was programmed by the experimenter, and the lot determined the random assignment to the experimental condition. The questionnaire described the decision situation and gave information about the amount of money the subject could earn. The size of the cake (z) was 10 tokens in each of the two rounds. Then the participant received a letter from the coplayer with a written note of his or her share (x_1). In response, the participant was asked to write his or her

4. Dependent on values of the parameters, the fairness, competition, and cooperation (FCC) model predicts a share of 0% or 50% for the dictator game. A nonlinear modification of the utility function eliminates this weakness (Fehr and Schmidt 1999).

decision, that is, the number of tokens (x_2) shared with the coplayer, in a letter and to put the letter in an envelope. A messenger delivered the sealed envelope to the coplayer. On the questionnaire, the participant was asked to give his or her coplayer's decision, his or her personal decision, the monetary payoff resulting from these decisions (to control for understanding), a verbal description of the decision, and a description of his or her own and the coplayer's expectations. Then a second questionnaire with psychological items on fairness and sociodemographic questions, as well as a second decision situation on a dilemma game, was administered. Subjects received 10 Swiss francs (sFr.), roughly U.S. $6.50 in 1999-the year the first two experiments were conducted-for participation and additional monetary rewards, depending on experimental condition and decision behavior.

EXPERIMENT 1

In experiment 1, the level of generosity x_1 of (programmed) player 1 was varied. I investigated reciprocity effects by assigning subjects randomly to three experimental conditions: x_1 = .20, .50, and .60, respectively. As mentioned before, the size of the cake in each round was 10 tokens. A token was worth sFr. 2. (~U.S.$1.30). Depending on experimental condition and participant's decision, possible earnings for this task ranged from sFr. 4 to sFr. 32. For example, in the condition x_1 = .60, the participant received 6 tokens from player 1. If the decision was made to divide the cake evenly (x_2 = .5), he or she kept another 5 tokens, which amounted to a total of sFr. 22 for this task. The number of participants in experiment 1 was 69. According to the strict rationality hypothesis, the participant's offer should be 0 in all three experimental conditions. In contrast, according to the hypothesis of altruistic reciprocity, the amount shared with player 1 is expected to be positive, and it should increase with the proportion of the cake that player 1 sacrificed to player 2.

EXPERIMENT 2

With experiment 2, I explored the hypothesis of stake effects. Primarily, I compared two groups, one with low payoffs and one with high payoffs. In both groups, x_1 is .5, that is, the programmed player divided the 10 tokens evenly in round 1. Participants were in the position of player 2 of a sequential dictator game. In addition, in this experiment, there were two other experimental conditions, an "ordinary" dictator game and a game called "simultaneous dictator." In the ordinary dictator game, participants had to divide 10 tokens, and the game ended thereafter. In simultaneous dictator, there are two players, A and B, both in the role of a dictator. A shared the cake with B, and B shared the cake with A simultaneously. A knew in advance that B would share the cake with him or her and vice versa. This game was designed to test an implication from the Rabin (1993) model, but this is not of main importance here.

As in experiment 1, in all conditions except for high stakes, 10 tokens were worth sFr. 20. In the high-stakes condition, payoffs were multiplied by a factor of 5.

Under this condition, participants received five tokens worth sFr. 50 from programmed player 1. Thus, a selfish *homo economicus* would have earned sFr. 150 (~U.S.$100) in this decision situation.

Participants were assigned to experimental conditions randomly, although not with equal probabilities. As mentioned before, the main emphasis was on the comparison of the low-stake group with the high-stake group. To increase the power of the test (to reduce the β-error, the probability of a participant's being assigned to the low-stake group was doubled. The number of participants in experiment 2 was 105.

By using this design, the robustness of altruistic reciprocity or reciprocity norms is explored. According to the hypothesis of the erosion of reciprocity norms, the proportion shared with player 1 is expected to diminish if stakes increase.

EXPERIMENT 3: FIELD EXPERIMENT

In a field experiment, the hypothesis concerning a reciprocity effect of a donation on the response probability of a mailed questionnaire was explored. Compared to an experiment conducted in the laboratory, a field experiment sheds more light on the external validity of the reciprocity effect. The study replicates earlier experiments reported in a meta-analysis by Church (1993). In the second wave of a panel study with a subsample of the Swiss Labor Market Survey, 600 participants were randomly assigned to three experimental conditions.[5] The subsample consisted of labor force participants between the ages of 18 and 60, living in the German-speaking part of Switzerland. In the control group, participants received a letter with a questionnaire. The second group received the questionnaire with a letter promising that respondents would receive a phone card worth sFr. 10 when the questionnaire was returned. The third group received the questionnaire and the letter with the phone card enclosed as thanks for the effort and participation.

A *homo economicus* in the latter group who perceives the effort of answering and returning the questionnaire as a cost would keep the phone card and not return the questionnaire. Of course, he or she also would not return the questionnaire in the control group. If the phone card outweighs the cost and he or she believes the promise, he or she will return the questionnaire in the second group. Hence, according to strict rationality theory, the response rate should be highest in the group with a promised donation. However, according to the hypothesis of reciprocity norms, subjects feel the obligation to reciprocate if they receive the donation in advance. In contrast to strict rationality theory, the reciprocity hypothesis predicts that the highest response rate will be in the "enclosed" condition.

5. The field experiment was conducted with Ben Jann. See Diekmann and Jann (2001) for a more detailed account.

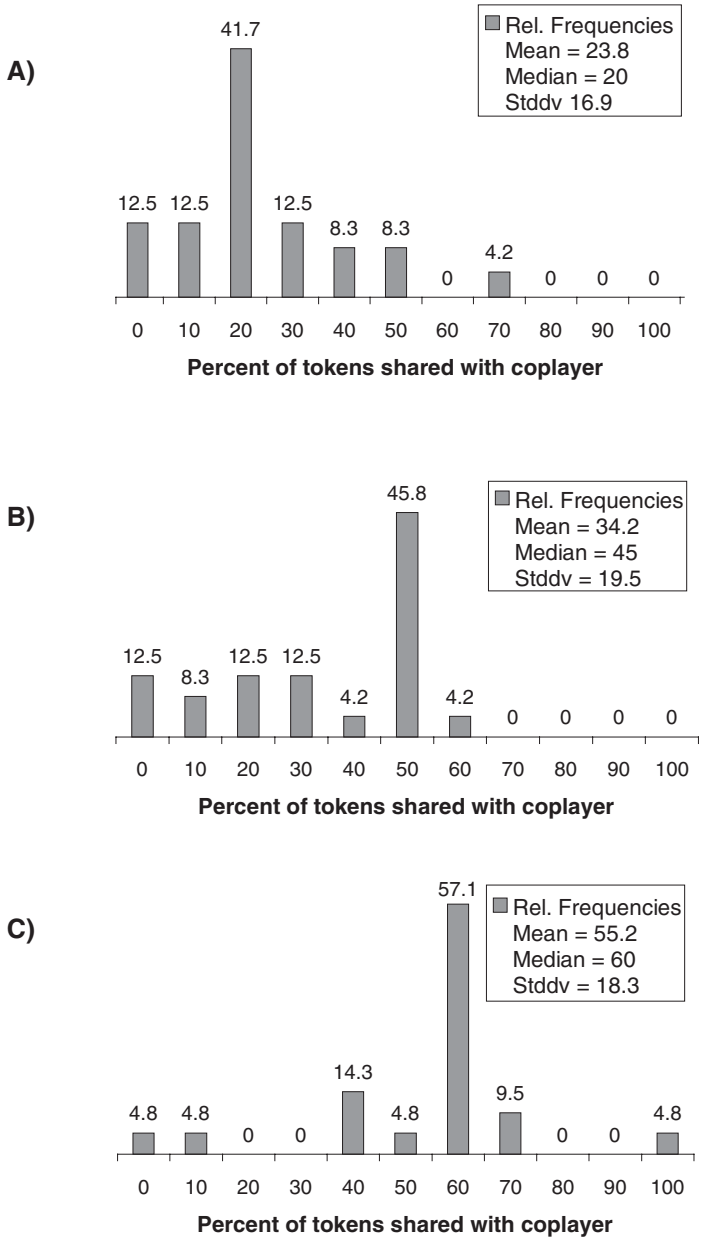


Figure 1: Distribution of Tokens Shared by Actor 2 under Various Conditions of Generosity of Actor 1

NOTE: (A) $n = 24$, player 1 shared 20% in the first round; (B) $n = 24$, player 1 shared 50% in the first round; (C) $n = 21$, player 1 shared 60% in the first round. Analysis of variance: $F(2, 66) = 17.03$; $p = 0.000$.

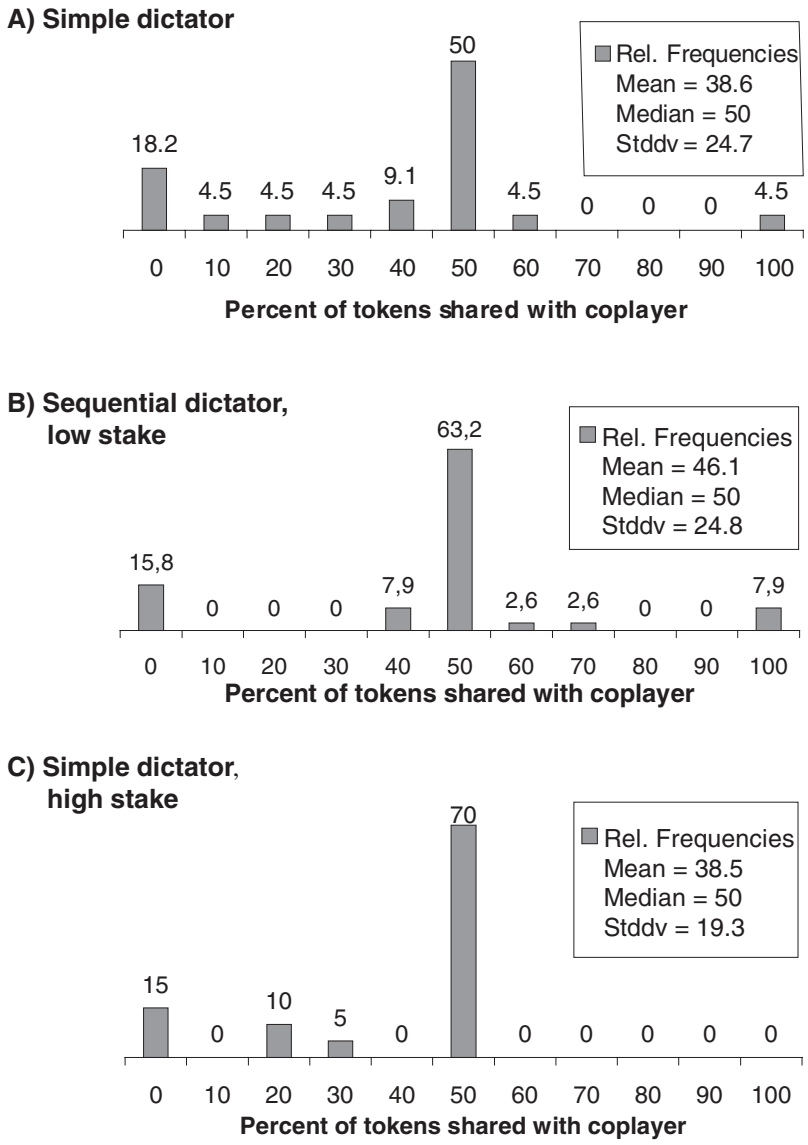
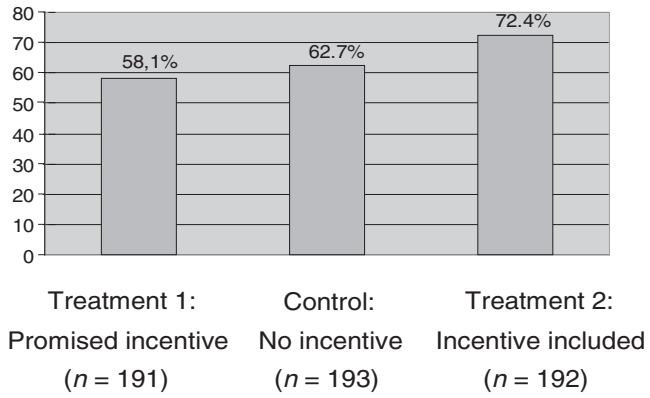


Figure 2: Percent of Tokens Shared in the Low-Stake and High-Stake Conditions
NOTE: Analysis of variance: $F(2, 77) = 1.01$; $p = 0.37$. (A) $n = 22$; (B) $n = 38$, player 1 shared 50% in the first round; (C) $n = 20$, player 1 shared 50% in the first round.



Treatment 2 versus treatment 1: $p = .003$.

Figure 3: Experiment with Mailed Questionnaire

RESULTS

Figure 1 displays the distributions of participants' share (x_2) in the sequential dictator game under various conditions of the coplayer's generosity. It can be seen that the mode of the x_2-distribution corresponds to x_1. If player 1 shares 20%, the relative majority of participants (42%) paid 20% back in round 2. If player 1 shares the cake evenly, 46% of participants sacrificed half of the cake as well. And if player 1 behaves overly fairly by sharing 60% in round 1, the majority of players (57%) reciprocated a 60% share for the coplayer. Also, the arithmetic means of x_2 differ substantially. These are 24, 34, and 55 in the 20%, 50%, and 60% condition, respectively. An analysis of variance yields a highly significant value of $F = 17.03$ ($df_{\text{between}} = 2$, $df_{\text{within}} = 66$, $p = .000$). The ERC model predicts behavior quite well, although there are cases that do not conform to the model. Two-thirds of participants complied with the theoretical expectation ($x_2 \leq x_1$) in the 20% condition, 96% comply in the 50% condition, and 86% confirm the prediction in the 60% condition. Only a small fraction (7 out of 69 participants) behaves in a strictly selfish manner. In sum, about 90% of participants do not exhibit a behavior in compliance with the strict rationality theory. Experiment 1 clearly demonstrates a strong motivational force of altruistic reciprocity. Participants will reciprocate, even if this behavior is not in their short- or long-term material interest.

As in experiment 1, findings from experiment 2 do not confirm the strict rationality hypothesis. Eighty-six participants out of 105 in the four experimental conditions share more than zero tokens with their coplayers, and 97 out of 105 share 50% or less as predicted by the ERC model. Figure 2 depicts the distribution of generosity for the

simple dictator game, the low-stake condition of sequential dictator, and the high-stake condition of the sequential dictator game.[6] Again, in both conditions of sequential dictator, a majority of participants reciprocate the fair play of the coplayer. Yet the main purpose of experiment 2 is the exploration of stake effects. By the modified or weak rationality hypothesis, a rise in stakes should lead to an increase in the proportion of players employing the strategy of a *homo economicus*. On the contrary, higher stakes do not reduce the level of generosity. In the low-stake condition as well as in the high-stake condition, the fraction of purely selfish actors is 15% to 16%. Under both conditions, the majority of subjects shares evenly. The relative frequency of the modal behavior is even slightly increased in the high-stake group. Seventy percent of participants share sFr. 100 evenly, despite the opportunity to gain the whole amount in full anonymity. The corresponding figure is 63% in the low-stake condition. However, there might be a tendency that high stakes reduce overly fair behavior. In the low-stake condition, 5 out of 38 subjects offer more than 50%. This figure is 0 in the high-stake group. Possibly, a small fraction of players will be less serious with their decisions if stakes are relatively low. This is the reason that the mean (but not the median) of x_2 is larger in the low-stake group compared to the high-stake group, although the difference is not significant for $\alpha = .05$ ($t = 1.19$, $df = 56$, $p = .241$). Also, in accordance with findings surveyed by Camerer and Hogarth (1999), there is a tendency that raising the incentives reduces the variance of x_2. The difference, however, is not significant. To summarize, there is clear evidence that altruistic reciprocity for high stakes is as strong as for low stakes.

Now I will turn to the field experiment. Six hundred questionnaires were mailed to participants in the first wave of the Swiss Labor Market Survey. Twenty-four addresses were invalid, leaving a sample of 576 cases for analysis. Of course, based on the level of the response rate, the sample is selective. About 2 years before the field experiment, all participants had answered a telephone interview and a mailed questionnaire in the first part of the survey. Thus, the response rate was probably biased upwards, compared with the rate in the general population. However, I am interested only in the differences between experimental conditions. Response rates in the time interval of 42 days before a reminder was sent, are displayed in Figure 3. (After receiving the reminder, response rates increased by another 12 to 16 percentage points.) Again, the results of the experiment do not support the strict rationality hypothesis. Response rates are lowest in the group with a promised donation, and they are highest in the group with the donation enclosed. The percentage difference of 14 percentage points is significant for $\alpha = .05$ ($p = .003$). The result of the replication is in accordance with other studies comparing response rates for enclosed versus promised donations (for a meta-analysis, see Church 1993). These findings clearly confirm the hypothesis of the behavioral effect of a reciprocity norm.

6. For the simultaneous dictator game, the distribution is 0 (24), 10 (4), 20 (4), 30 (8), 40 (4), 50 (52), 100 (4), with the (%) relative frequencies in parentheses. Number of subjects is 25, mode is 50, and arithmetic mean is 35.2. Analysis of variance for all four experimental conditions yields no significant differences between arithmetic means for $\alpha = .05$ ($F = 1.18$, $df_{\text{between}} = 3$, $dfithin = 101$, $p = .32$).

DISCUSSION

The experiments show that a norm of reciprocity shapes behavior and that altruistic reciprocity remains robust if stakes are high. Also, results of a controlled experiment outside the laboratory demonstrate the importance of a reciprocity norm in daily life.

Robustness of altruistic reciprocity if stakes are high justifies the notion of the "power of reciprocity." Of course, there is the objection that the threshold for a decay of reciprocity may be above the payoffs of experiment 2. If one assumes that thresholds vary individually, a rise in stakes should concern at least a certain proportion of the sample, thereby reducing the relative frequency of full reciprocation. Yet the expected reduction in fairness was not observed in the experiment. An example of conditional fairness for extremely high stakes is Frederick G. Banting, who, in collaboration with his young assistant Charles H. Best, discovered insulin. Banting received the Nobel Prize but shared the prize with Best.[7] It may be that the average proportion of the share would be much smaller if one replicated the sequential dictator game with the monetary payoff of the Nobel prize. Of course, this will never be known because there is certainly no sponsor for an experiment like that. For now, it is sufficient to know that the norm of reciprocity has the power to outweigh a considerable amount of monetary payoffs.

There are other experiments with high stakes, some conducted in low-income countries. Evidence is mixed concerning the effect of an increase in incentives on the frequency of altruistic or "non-Nash" cooperative behavior (Camerer and Hogarth 1999). Fehr, Fischbacher, and Tougarova (2002) report findings from a "gift-exchange" experiment with students in Moscow. Participants' payoffs were raised to the level of an average monthly income. Nonetheless, Fehr, Fischbacher, and Tougarova observed the same pattern of a high degree of reciprocity as in the low-stake condition. On the other hand, there are two experiments with the ordinary dictator game that yield the result that stakes increase selfish behavior (Forsythe et al. 1991; Sefton 1992; see also the survey by Camerer and Hogarth 1999). I conjecture that unconditional fairness, as exhibited in an ordinary dictator game, is less robust with regard to the increase in incentives than conditional fairness. The obligation of reciprocal behavior seems to be more powerful than unconditional fairness, a conjecture that should be proved in further research. Physiological research matches the observation of social science experiments. A powerful behavioral principle may have physiological roots in the human brain. Rilling et al. (2002) measured the activity in certain parts of the brain of subjects playing a repeated prisoner's dilemma game by scanning them with magnetic resonance imaging (MRI). They believe that neural networks seem to reward altruistic reciprocity (see also Cory 1999; Hoffman, McCabe, and Smith 1998).

7. Best's son reports on Banting's telegram, sent after having received the Nobel Prize in 1923: "At any meeting or dinner, please read the following stop I ascribe to Best equal share in the Discovery stop hurt that he is not so acknowledged by Nobel trustees stop will share with him stop" (Best 1996). Banting received the prize together with John R. Mcleod, who also voluntarily shared the prize with James B. Collip, another member of the "team." There was much controversy on the priority of the discovery. For a detailed account of the history of insulin research, see Bliss (1982).

In future research, physiologists will possibly discover more about the "neural basis of social cooperation" (Rilling et al. 2002). However, the task of sociologists is to study social and cultural variations of reciprocal behavior and their implications on society. So far there are many important implications for social theory and a great variety of applications of the principle of reciprocity to diverse sociological areas. The "fair wage hypothesis" of "efficiency wage theory" proposed by Akerlof and Yellen (1988) was already mentioned. Although in classical economics, marginal productivity determines the wage level, there is also the reverse causal relation. Higher wages are reciprocated by a rise in workers' efforts. Hence, higher wages increase productivity. By this rationale, employers are willing to pay wages above the market clearing level with the unintended side effect of raising the unemployment rate. Moreover, negative reciprocity of workers inhibits employers from reducing wages in economic recessions, an explanation for the phenomenon of wage rigidity (Fehr and Gächter 2000). Also, customers exert altruistic negative reciprocity when refusing unfair prices. Thaler (1992) demonstrates the argument using the example of shopkeepers who refrain from increasing the price of snow shovels after a snowstorm because they are afraid of a consumer strike. Marketing sellers often send test products or donations, expecting potential customers to place an order (Cialdini 1993). Fraud in Internet auctions is relatively rare because of the establishment of a reputation system. Such a reputation system, that is, the evaluation of sellers by buyers (and vice versa) would not work well without an obligation of positive or negative reciprocity. And why do children care for their elderly parents? This is not a repeated game but altruistic reciprocity. Intergenerational relations are very much shaped by the norm of reciprocity.

For social theory, the reciprocity hypothesis is a key element in explaining the production of collective goods, the emergence of social cooperation, and the existence of social norms (Fehr and Gächter 2000). Social norms promoting cooperation would erode if not stabilized by sanctions leading to a sanctioning dilemma or "second-order free-rider problem" (Heckathorn 1989). Altruistic reciprocity contributes to the solution of the sanctioning dilemma because individuals are willing to employ sanctions even if they are costly (Fehr and Gächter 2000, 2002; Diekmann and Voss 2003).

The "shadow of indebtedness," as Gouldner (1960) coins the obligation of reciprocity, is a most powerful principle in shaping social interactions and furthering social integration. Sociologists are well advised to continue the tradition of Simmel (1950), Malinowski (1926), Mauss (1950/1990), Blau (1964), Homans (1958, 1961), and Gouldner (1960); to join the transdisciplinary debate on reciprocity; and "to bring reciprocity back" into sociological thinking.

REFERENCES

Akerlof, George A., and Janet L. Yellen. 1988. Fairness and unemployment. *American Economic Review* 78:44-49.

Axelrod, Robert. 1984. *The evolution of cooperation*. New York: Basic Books.

Best, Henry B. M. 1996. Speech to the Academy of Medicine. April 24. Toronto, Canada, The Vaughan Estate. http://www.discoveryofinsulin.com/Best.htm.

Blau, Peter M. 1964. *Exchange and power in social life*. New York: Wiley.

Bliss, Michael. 1982. *The discovery of insulin*. Chicago: University of Chicago Press.

Bolton, Gary E., and Axel Ockenfels. 2000. ERC: A theory of equity, reciprocity, and competition. *American Economic Review* 90:166-93.

Bourdieu, Pierre. 1983. Forms of social capital. In *Handbook of theory and research for the sociology of education*, ed. J. G. Richardson, 241-58. New York: Greenwood.

Borjas, George J. 1996. *Labor economics*. New York: McGraw-Hill.

Boyd, Robert, and Peter J. Richerson. 1988. The evolution of reciprocity in sizeable groups. *Journal of Theoretical Biology* 132:337-56.

Camerer, Colin F. 1997. Progress in behavioral game theory. *Journal of Economic Perspectives* 11:167-88.

Camerer, Colin F., and Robin M. Hogarth. 1999. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty* 19:7-42.

Church, A. H. 1993. Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly* 57:62-79.

Cialdini, Robert B. 1993. *Influence. Science and practice*. New York: HarperCollins.

Coleman, James S. 1988. Social capital in the creation of human capital. *American Journal of Sociology* 94 (Suppl.): 95-120.

Cory, Gerald A. 1999. *The reciprocal modular brain in economics and politics*. New York: Kluwer/Plenum.

Dasgupta, P. 1988. Trust as a commodity. In *Trust: Making and breaking cooperative relations*, ed. D. Gambetta, 49-72. New York: Blackwell.

Diekmann, Andreas, and Ben Jann. 2001. Anreizformen und Ausschöpfungsquoten bei postalischen Befragungen. Eine Prüfung der Reziprozitätshypothese [Incentives and response rates by mailed questionnaires: A test of the reciprocity hypothesis]. *ZUMA-Nachrichten* [Newsletter of the Center for Survey Research and Methodology], no. 48: 18-27. Mannheim, Germany: ZUMA.

Diekmann, Andreas, and Thomas Voss. 2003. Social norms and reciprocity. Discussion paper, University of Leipzig, Germany.

Fehr, Ernst, Urs Fischbacher, and Elena Tougarova. 2002. Do high stakes and competition undermine fairness? Evidence from Russia. Working paper no. 120, Institute for Empirical Research in Economics, University of Zurich, Switzerland.

Fehr, Ernst, and Simon Gächter. 2000. Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14:159-81.

———. 2002. Altruistic punishment in humans. *Nature* 415:137-40.

Fehr, Ernst, and Klaus M. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114:817-68.

Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton. 1994. Fairness in simple bargaining experiments. *Games and Economic Behavior* 6:347-69.

Freitag, Markus. 2001. Das soziale Kapital der Schweiz. Vergleichende Einschaetzungen zu Aspekten des Vertrauens und der sozialen Einbindung [Social capital in Switzerland: Comparative evaluation of aspects of trust and social embedding]. *Schweizerische Zeitschrift für Politische Wissenschaft* 7:87-117.

Friedman, James W. 1977. *Oligopoly and the theory of games.* Amsterdam: North-Holland.

Fudenberg, Drew, and Jean Tirole. 1991. *Game theory.* Cambridge, MA: MIT Press.

Fukuyama, Francis. 2000. Social capital and civil society. Working paper 00/74. Washington, DC: International Monetary Fund.

Gouldner, Alvin W. 1960. The norm of reciprocity: A preliminary statement. *American Sociological Review* 25:161-78.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3:367-88.

Heckathorn, Douglas D. 1989. Collective action and the second-order free-rider problem. *Rationality and Society* 1:78-100.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith. 1998. Behavioral foundations of reciprocity: Experimental economics and evolutionary psychology. *Economic Inquiry* 36:335-52.

Homans, George C. 1958. Human behavior as exchange. *American Journal of Sociology* 63:597-606.
———. 1961. *Social behavior: Its elementary forms.* New York: Harcourt, Brace and World.
Knack, Stephen, and Philip Keefer. 1997. Does social capital have an economic payoff? A cross-country investigation. *Quarterly Journal of Economics* 112:1251-88.
Malinowski, Bronislaw. 1926. *Crime and custom in savage society.* London: Routledge and Kegan.
Mauss. Marcel. 1950/1990. *The gift: The form and reason for exchange in archaic society.* London: Routledge.
Nowak, Martin A., and Karl Sigmund. 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393:573-77.
Putnam, Robert D. 2001. *Bowling alone.* New York: Simon & Schuster.
Rabin, Matthew. 1993. Incorporating fairness into game theory and economics. *American Economic Review* 83:1281-302.
Rapoport, Anatol, and A. Chammah. 1965. *Prisoner's dilemma.* Ann Arbor: University of Michigan Press.
Regan, Dennis T. 1971. Effects of favor and liking on compliance. *Journal of Experimental Social Psychology* 7:627-39.
Rilling, James K., David A. Gutman, Torsten R. Zeh, Giuseppe Pagnoni, Gregory S. Berns, and Clinton D. Kilts. 2002. A neural basis for social cooperation. *Neuron* 35:395-405.
Sefton, Martin. 1992. Incentives in simple bargaining games. *Journal of Economic Psychology* 13:263-76.
Simmel, Georg. 1950. *The sociology of Georg Simmel.* Ed. and trans., K. H. Wolff. Glencoe, IL: Free Press.
Taylor, Michael. 1976. *Anarchy and cooperation.* New York: Wiley.
Thaler, Richard H. 1992. *The winner's curse: Paradoxes and anomalies of economic life.* Princeton, NJ: Princeton University Press.
Trivers, Robert L. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35-57.
Wedekind, Claus. 1998. Give and ye shall be recognized. *Science* 280:2070-71.
Wedekind, Claus, and M. Milinski. 2000. Cooperation through image scoring in humans. *Science* 288:850-52.
Whiteley, Paul F. 2000. Economic growth and social capital. *Political Studies* 48:443-66.