PART I

INTRODUCTION: SOCIAL DILEMMAS AND TRUST

Chapter 1

Introduction

Elinor Ostrom and James Walker

Imagine the following decision situation involving two individuals. Individual 1 is endowed with $10.00. She can keep the entire sum or send some part of the $10 to individual 2, an anonymous counterpart. Individual 1 knows that any money sent to individual 2 will later be tripled in value. Furthermore, individual 2 knows that he will have the option of sending back to individual 1 some part of the tripled sum of money, a sum determined by individual 2.

Imagine that this game is to be played only once and that both individuals know that all decisions will remain completely anonymous. How might individual 1 go about making a decision in this situation? Based on motivations rooted only in pecuniary payoffs from the game, the noncooperative game-theoretic equilibrium is for individual 1 to send no money to individual 2. Individual 1 would reason, by *backward induction,* that individual 2 would have no incentive to send any money back. Thus, individual 1 would send no money to individual 2. On the other hand, there would be obvious gains to both players if individual 1 were to trust individual 2 to return an amount at least as great as the sum she sends him. The joint gains would be as much as three times the original endowment to individual 1.

What types of decisions will individuals make when they actually encounter this one-way game of trust? (see chapter 3, this volume). Joyce Berg, John Dickhaut, and Kevin McCabe (1995) conducted just such a study. Using a *one-shot* decision setting with double-blind experimental procedures to ensure complete anonymity, these researchers find that thirty of thirty-two subjects in the role of individual 1 sent money to their counterparts (in the role of individual 2)

($5.36, on average). Of the thirty subjects in the role of individual 2 who received funds, eighteen returned more than $1.00 ($4.66, on average), and eleven sent more than the original amount allocated by their partners. Interestingly, on average, those who sent at least $5.00 received an average return in excess of the amount they sent; those who sent less than $5.00 received a negative net-average return. In other words, those individuals in the first position who trusted their counterparts more were more likely than those who were less trusting to leave the game with more wealth than they had entered with.

In a follow-up study, Nancy Buchan, Rachel Croson, and Eric Johnson (1999) implement the experimental decision setting of Berg, Dickhaut, and McCabe (1995) in a cross-national design. They investigate the same game with 188 subjects from China, Japan, Korea, and the United States. They find no pure country effects in terms of the amount sent by subjects in the role of individual 1 (on average, 67 percent of the endowment) or in the amount returned (on average, 31 percent of the amount received). The researchers examine several other potential explanatory variables, in addition to country effects, including culture (based on survey data in which questions relate to an individual's attitude toward group versus individual outcomes), *social distance* (manipulated experimentally), and *communication* (manipulated experimentally). Culture, in the form of subjects' showing a greater orientation toward group outcomes, significantly and positively affects the amounts sent and returned.[1] The opportunity to communicate information about one's self and learn something about the other person with whom a subject is paired also had a positive effect on the amounts sent. Buchan, Croson, and Johnson find that "trusters prosper." "Subjects who sent above-average amounts to their partners took home greater wealth than did subjects who sent only average amounts or less" (Buchan, Croson, and Johnson 1999, 22).

In a third study, John Dickhaut and colleagues (1997) examine the impact of allowing individual 2 to build a reputation by adding a publicly announced second round to the basic structure of the first (Berg, Dickhaut, and McCabe) experiment, in which the individual continued to play with the same counterpart. According to standard *noncooperative game theory,* the existence of this second round should make no difference in the behavior of individuals in the first round. Dickhaut and colleagues wanted to ascertain whether individuals in the position of individual 2 would act even more trustworthy than those who had participated in the first study to assure those in the position of individual 1 that they could be trusted. In the first round, all twenty-three subjects in the role of individual 1 sent some money to their counterparts, and ten of the twenty-three sent the maximum

sum ($10.00). Those in the position of individual 2 showed higher levels of reciprocity in the first round than had been exhibited in the Berg, Dickhaut, and McCabe study. Twenty of them returned more than the original sum received, leading to a positive-sum outcome for all involved.

The second round followed a different pattern. Nineteen out of twenty of the individual 1s who had received positive returns in the first round made positive investments again in the second, and all three of the individual 1s who had received a negative return in the first round sent zero to their counterparts in the second round. The most significant difference in outcome was that only seven of the nineteen individual 1s received a positive net return in the second round. The reciprocity that had been exhibited in the first round was substantially reduced in the second and final round.

These three studies illustrate the intricate relation between trust and expectations of reciprocity.[2] In these experiments, a substantial proportion of individuals trusted that the person with whom they were paired—a stranger—would reciprocate trust by returning at least as much money as he or she had been sent by the first player. In all of these games, the predicted subgame-perfect equilibrium, assuming that utility is adequately captured by monetary payoffs, is for individual 2 to return zero. Because individual 2 is predicted not to reciprocate, the prediction is also that individual 1 will not invest anything, owing to the risk of losing everything that is sent. In all three studies, trust and trustworthy behavior exceeded predicted levels.

On the other hand, the degree of trust and trustworthiness varies substantially.[3] A variety of contextual factors associated with the structure of these experiments helps to explain the variance. Individual preferences toward payoffs, prior experience with this game structure, players' capacity to learn more about the personal characteristics of each other, and ability of players to build reputations all appear to affect the decisions of players in both positions. A second major finding is that the rates of trusting, reciprocity, and trustworthy behavior exhibit high variance across experimental conditions. Thus trusting and trustworthy behavior are not unchanging, universal attributes of all individuals but are rather the result of multiple contextual and individual attributes. Understanding trust and the conditions that are conducive to trust is a challenging task.

These three studies illustrate how experimental methods can help social scientists learn more about trusting and trustworthy behavior and contribute to the testing of diverse theories about the origin, sustenance, and outcomes of trust and trustworthiness. Furthermore, the structure of these experiments has a parallel structure to situations that Robert Trivers (1971) has posited as key for the expression of

reciprocal altruism. For an act to be an instance of reciprocal altruism, the following ingredients must be present:

- An action of the first actor is beneficial to the recipient but costly to the first actor.
- There is a time lag between giving and receiving.
- The actor who gives has an expectation of receiving an equal or larger benefit in the future.

In all three studies, the experimenter controls the pecuniary incentives associated with alternative decisions, the level and form of information subjects have about other subjects, and the extent to which reputations for trustworthiness can be built or destroyed. These studies are illustrative of the types of experimental environments, analyses, and outcomes that are reported in this volume. The studies reported here focus on multiple alternative decision settings that have been created to examine particular questions related to how pecuniary and nonpecuniary incentives and other structural variables affect trusting behavior, as well as trustworthiness.

Why is trust so important? These three studies also begin to provide a hint about the importance of trust in every society. In all of the gaming situations just described, an individual is faced with an "investment" decision that holds out the possibility of a substantial return if a counterpart proves trustworthy and returns more than the first individual invests in the relationship. Each individual 1, however, also faces the risk that his or her counterpart will take the money and run. A core aspect of most definitions of trust is the "intention to accept vulnerability based upon positive expectations of the intentions of the behavior of another" (Rousseau et al. 1998, 395).

Situations that have this core structure abound in all societies and at all times in history. Kenneth Arrow (1974) stresses the ubiquity of trust in almost every economic transaction. An empirical study of trust using a variety of cross-national indicators finds that trust is positively associated with homogeneity with respect to ethnicity and social and economic relationships and in situations in which diverse legal and social mechanisms for counteracting opportunism are well developed (Zak and Knack 1998). Furthermore, Arrow finds that higher rates of investment and growth are positively associated with higher levels of trust. The basic relationship between management and employees is one of mutual trust (see chapter 3, this volume) in which each player has to trust that contributions beyond the minimal will be forthcoming from the other (Fehr, Gächter, and Kirchsteiger 1996).

Elias Khalil argues that market-based societies, which are frequently characterized as encouraging self-interested, untrusting behavior, have a greater need for trust than kin-based or other forms of economic organization. "First, as economic exchange becomes less intermingled with kinship and more based on formal contractual relationships, the monitoring conducted by the kin members and the threat of ostracism almost vanish. Second, the modern judicial system, which replaces the threat system of ostracism and shunning, cannot practically monitor the extensive growth of contractual agreements—even the explicit ones" (Khalil 1994, 340). Simply thinking about the problem of buying a used car should remind every reader of the risks involved in the large number of diverse contractual arrangements that characterize life in a market economy (Buskens and Weesie 2000). Furthermore, our humanoid and primate ancestors faced many situations in which, to obtain long-term joint benefits, they had to decide to make themselves vulnerable based on positive expectations of the intentions of others (de Waal 1996; Boehm 1999). Part of the evolutionary heritage of the modern human is the capacity to recognize human faces and engage in a rough form of mental accounting so as to trust some individuals more than others in particular types of situations.

This book examines the foundations for trust and trustworthy behavior. The central theme that links all discussion relates to the gains from association that are achieved when individuals are able to develop trust and reciprocity. Whether they come in the form of market exchange or personal relationships, the gains from association depend on the willingness of individuals to take risks by placing their trust in others. Whether that trusting behavior is mutually beneficial and lasting depends on the trustworthiness of those in whom trust has been placed.

Concepts of Trust and Reciprocity

Chapter 2, by Elinor Ostrom, considers the abundant evidence of cooperation and reciprocity, focusing on the issue of how traditional noncooperative game theory might be enriched to become a behavioral-based theory. The author reviews the theoretical predictions of currently accepted *rational-choice theory* as it relates to *social dilemmas* and summarizes challenges to the reliance on a complete model of rationality. She focuses on how individuals achieve results that are "better than rational" by building conditions under which reciprocity, reputation, and trust help to overcome short-run, self-interested temptations. This parallels the arguments made in the experimental

chapters in part IV of this volume—in particular, the arguments made in chapter 10. Ostrom develops an initial framework for examining how contextual variables tend to affect reciprocity, reputation, and trust. She embeds this core in settings involving both small face-to-face groups and larger groups.

Russell Hardin concludes part I with his chapter, "Gaming Trust," by specifying the basic elements of a trust relationship between a truster and a trustee. First, the truster faces a situation in which he or she will be better off if the relationship is initiated and the trustee reciprocates. Second, the trustee may face some incentives to be trustworthy with regard to the specific relationship, such as internal norms or the prospect of extended relationships. Third, this incentive to be trustworthy can be overruled by a variety of other considerations, such as short-term gains from not being trustworthy. Hardin points to the cognitive structure of trust—the belief an individual has in the likely behavior of the other.

Hardin argues that to account for trust, one must first account for trustworthiness. Trustworthiness can be seen as primarily based either on incentives or on the normative attributes of the decision maker. Trustworthiness is incentive based if the trustee has reason to encapsulate the interests of the truster into his or her own preferences; an example is a situation in which the trustee values the continuance of the relationship. Trustworthiness is normatively based if the trustee feels morally obligated to fulfill the trust.

Building on the conceptual nature of associations dependent on trust, Hardin distinguishes between three models of trust that have been used in the literature: mutual trust, one-way trust, and thick relationships. Mutual trust is seen as an interaction that is part of a long sequence of association between the same parties, a relationship in which both parties play a role as truster and trustee. In one-way trust, the truster must trust the trustee in order to gain from their interaction. If the truster does not trust, the association ends, and potential gains are not realized. Once the truster demonstrates an act of trust, the trustee then is faced with a unilateral decision regarding how to respond to the trust. This game is one-way trust because only the truster faces the risk of trusting the trustee. The thick-relationship theory focuses on the idea that real relationships are experienced in complex settings based on diverse layers of interactions and that actions at one level may, in fact, depend on the action's ramifications at other levels. Actual motivations for trusting or being trustworthy are better understood when reputations in repeated relationships and third-party relationships are taken into account.

Biological Foundations of Trust and Reciprocity

A common theme found among many of the chapters in this volume is an evolutionary approach to understanding the bases for trust and trustworthiness in modern-day associations. The two chapters of part II lay the foundation for understanding this approach.

In chapter 4, Robert Kurzban presents a well-structured argument for the coevolution of cooperation and reciprocity. Biologists have struggled with the notion of altruism and cooperation, as the two behaviors appear to go against the principles of natural selection. If altruism is defined as helping another organism at a cost to oneself, then it seems logical that natural selection would eventually eliminate this behavior. The evidence of altruism as a stable evolutionary behavior is explained as a result of reciprocal behavior.

To explain the existence of altruism on the basis of reciprocity, Kurzban defines natural selection in terms of design or adaptations rather than behavior. Specifically, an organism can be thought of as being made up of subsystems, each designed to solve a particular problem and contribute to reproductive success. In this view, natural selection is a process that, over time, selects the best designs in solving a problem.[4]

Kurzban discusses two ways in which natural selection can lead to altruistic behavior: kin selection and reciprocal altruism. Kin selection theory (first articulated by William Hamilton [1964]) is the capacity of a gene to increase its rate of replication by facilitating the replication of exact copies of itself, even in different organisms. According to the kin selection theory, a gene would cause behavior in organisms that would benefit another (even at a cost to itself) if there is a high enough probability that the other organism contained that gene. Reciprocal altruism (first articulated by Robert Trivers [1971]) is used to explain altruistic behavior by unrelated organisms. The theory relies on one organism conferring a benefit on another, contingent on the second organism's having conferred a benefit on the first.

Certain conditions are necessary for both forms of altruism to occur. First, both rely on the contingent delivery of benefits. In kin selection, the contingency involves a cue to relatedness. In reciprocal altruism, the contingency is a particular history of interaction and conditions for the forms of such interactions. Second, the environment must be one in which there are benefits to be conferred. Third, organisms must have repeated interactions with one another. Fourth, organisms must have sufficient information-processing abilities to be able to distinguish among individuals and remember which ones

have or have not delivered benefits in the past. Finally, organisms must have sufficient information-processing sophistication and behavioral flexibility to interact with other organisms contingent on the history of interaction. Kurzban's discussion of the evolution of cooperation and reciprocity in human and nonhuman species focuses on how these conditions may have impacted the selection of specific kinds of behavior.

Kurzban's chapter is complemented by the detailed analysis found in chapter 5, by Frans de Waal, which stresses the deep evolutionary roots of trust and reciprocity. De Waal presents well-documented evidence to support the thesis that the exchange of social services among chimpanzees rests on cognitive abilities that allow current behavior to be contingent upon a history of interaction. Two types of behavior are formally supported: grooming increases the probability that the recipient will share with its donor at a later time, and sharing decreases the probability that the donor of the service will groom its recipient at a later time. In summary, evidence is presented for an exchange mechanism in which donations and receipts of services are stored in memory and exert partner-specific effects on subsequent behavior of donors and recipients.

Testing a model of reciprocal altruism, in which kin relationships are not a necessary condition, is a crucial step in establishing the evolutionary roots of observed reciprocity and trust. Although it is not possible to develop a quantitative test based on a large number of observations of very costly acts, de Waal does develop an ingenious experimental design for observing whether individual chimpanzees who had given something of value to other chimpanzees were more likely to receive something in return. De Waal concludes that we have evidence "for the entire set of features expected if reciprocity is cognition-based: partner specificity, selective protest, retaliation, turn taking, and the effect of one service on another."

The chapters in part III build on the biological foundations of trust. In particular, two modeling chapters are presented that summarize key arguments related to the linkages between trust, intention detection, and cooperation in complex societies.

The Links Between Evolution, Cognition, and Behavior

Kevin McCabe's chapter 6 can be viewed as a modeling "bridge" to the preceding chapters by Kurzban and de Waal. McCabe develops a theory of exchange based on the idea that one can conceptualize decision makers as using "mental modules" in their behavioral decisions in social exchange. In this theory of exchange, the expectation of re-

ciprocal behavior is fundamental to the implementation of exchange through trust and conciliation.

As McCabe notes, however, "Trust and conciliation are relatively easy to explain once reciprocity is established, but how does one explain reciprocity?" He explores three areas important to addressing this question. First, he examines evidence of the adaptation of the human mind to allow for complex commitments that are required for reciprocal actions. Second, he argues that commitment and reciprocal actions are dependent on adaptations that allow for delay of gratification: "cool" mental mechanisms may have evolved that overcome "hot" mental mechanisms that call for immediate gratification. The final area focuses on the notion of a cognitive accounting system utilized to implement delay of gratification and resulting in reciprocal decision making. A "goodwill" accounting system tracks trading partners and forms the foundation for decisions regarding whether to trust, based on expectations regarding reciprocity. Such an accounting system is seen to be dependent upon a form of "mind reading" whereby decision makers contemplate the possible behavior of those with whom they may interact.[5] McCabe ends his chapter with a summary of a thought-provoking set of brain-imaging experiments that supply evidence of how the human brain may encode goodwill accounting. These experiments suggest that the frontal lobe regions of the brain are activated in situations in which subjects are attempting to cooperate with other human subjects, in a manner that does not occur when these subjects are playing against computer-simulated decision makers.

In chapter 7, James Hanley, John Orbell, and Tomonori Morikawa develop an evolutionary framework and computer simulations to examine more closely the role of evolutionary processes in altering dispositions toward trust and cooperative behavior. In particular, they examine a simulated world in which decision makers play both cooperative and conflictual games and players control information related to the accuracy of intentions. A key presumption in this analysis is that "if cognitive evolution has made it easy—or at least cost-effective—to detect cheaters, then cooperators will be able to avoid exploitation, and cooperation is likely to flourish." In the simulation study, cooperative games are modeled as *prisoner's dilemma games,* and conflictual games take the form of *hawk-dove games.* In the latter, decision makers must decide either to fight or not to fight. In the games investigated, if both fight, the winner captures the resource and the loser incurs a cost. If both do not fight, they both receive a reservation payoff that equals the value incurred by not playing.

The simulations suggest several key findings that researchers interested in complex systems and evolutionary processes should find in-

teresting. Cooperation does not prosper in simulated worlds void of opportunities for fighting. The results suggest that in such worlds, over time, increases in mutual defection and corresponding decreases in mutual cooperation occur. Over time, the population of agents becomes relatively more adept at issuing persuasive lies than at penetrating the lies of others. The inclusion of the hawk-dove option yields a different dynamic. Over time, there is an increase in mutual cooperation and a decrease in mutual defection. Why? With the hawk-dove option, decision makers who believe they face protagonists who are not sufficiently cooperative have a choice beyond opting out of the prisoner's dilemma game. Furthermore, such agents are disproportionately inclined toward cooperation.

Each of the modeling chapters points to the complex relationship between agents' dispositions toward trust and cooperation and their resulting behavior. In part IV, the authors examine such behavior in laboratory-created decision environments, in which subjects face salient rewards for their decisions and those of their counterparts.

Experimental Evidence

Part IV of this volume contains several chapters that review previous experimental work and recent research by the authors. The overriding theme linking much of this work is stated clearly by McCabe and Vernon Smith: "We summarize and report data from a research program motivated by concepts of reciprocity and intentionality detection that we propose as supplements to the standard game-theoretic principles. The question is not whether game-theoretic principles find application in the data—they do—but rather whether they suffice as the only principles that govern behavior."

In chapter 8, Karen Cook and Robin Cooper provide a review of experimental studies in sociology, social psychology, and economics that focus on issues of trust. They begin by confronting the issue of attitudinal versus behavioral measures of trust and the difficult task of disentangling acts of cooperation from trust. They stress that sequential games that allow one player to move before the other, and games that allow players to exit, provide clear evidence regarding "trust" as contrasted to simple cooperation.

They then turn to a review of selected studies focusing on subjects' inherent motivations, the incentive structure and strategic nature of the game facing experimental subjects, and the social context in which decisions are made. In conclusion, Cook and Cooper argue for the need for a more complete theory of the basis of trust, as well as a more complete and unified experimental approach, that examines

trust more extensively in terms of the social context in which decision makers find themselves.

Parallel to the evolutionary approach discussed by Kurzban, in chapter 9 Catherine Eckel and Rick Wilson conjecture that humans are conditional cooperators and that the decision to cooperate is based on initial expectations about the likely behavior of others with whom an association is contemplated. They explore the idea that "humans share a capacity to read one another's intentions through a set of cues such as facial expressions, body language, and tone of voice." They report results from two experiments designed to examine how facial expressions may affect an individual's perception of the trustworthiness of a total stranger and the resulting behavior of this individual. Similar to arguments made by Kurzban and McCabe, Eckel and Wilson develop the notion of a "theory of mind"—the ability to read directly another's intentions. In their principal study, subjects in a one-shot trust game view the faces of their presumed counterparts. They move first, in what is in essence a decision to trust or not to trust their counterparts. The main result is that, at the margin, subjects draw inferences from the images they observe and act accordingly. Subjects choose the trusting move 62 percent of the time. The rate is significantly increased if subjects have been paired with counterparts who display smiling images. Trusting behavior is also increased in cases in which subjects have rated their counterparts high on a trustworthiness index, based purely on the inferences they have drawn from their counterparts' photographic images.

In their chapter 10, Kevin McCabe and Vernon Smith offer evidence from a series of one-shot and repeated games in which players are randomly matched or matched repeatedly with counterparts. Reciprocity is based on triadic knowledge—information gained through the shared-attention mechanism. Shared attention, as a mental module, operates on two kinds of information: *focal points,* such as symmetry and *Pareto dominance* of payoffs that are likely to be observed by the other player, and move information, which communicates intentions through both foregone opportunities and assessment of the payoffs that remain as admissible future opportunities.[6]

McCabe and Smith test reciprocity theory by performing experiments based on three variations of an *extensive-form* bargaining-tree structure in which players alternate moves. Each variation has a *subgame-perfect Nash equilibrium* and a symmetric cooperative outcome. Following player 1's decision either to take a sure payoff or to continue the game, player 2 makes the choice as to whether to head toward the subgame-perfect Nash equilibrium or to put the decision back into the hands of player 1. At this point, player 1 can choose the

cooperation outcome. In games 0 and 1, player 1 can deviate from the cooperative equilibrium, attempting to raise his or her payoffs and reduce player 2's payoffs. However, player 2 can punish this action at a cost to his or her own payoff. In game 2, however, player 1 can deviate from the cooperative equilibrium with no threat of punishment from player 2 (note that game 0 differs from game 1 in that the cooperative payoffs are equal to subgame-perfect Nash equilibrium payoffs in game 0, whereas in game 1 subgame-perfect Nash equilibrium payoffs are less than cooperative payoffs).

In summary, McCabe and Smith find support for several key findings: first, analysis identifies factors affecting choice other than those predicted by noncooperative game theory; second, no meaningful support is observed for the hypothesis that subjects attach a personal and positive utility to the payoff of their counterpart; third, reciprocity does not require a positive probability of repeat interaction with the same person; fourth, reciprocity increases with the probability of repeat interaction with the same person; fifth, players use the payoffs of counterparts in their strategic analysis. Concerning the last finding, players use the payoffs of their counterparts not to identify the subgame-perfect Nash equilibrium, however, but to identify and try to achieve the cooperative outcome through reciprocity. The possibility that cooperative play results from the threat of punishment by player 2 for player 1's deviations from the cooperative outcome cannot be ruled out. When the threat of punishment is removed, the cooperative outcome diminishes, and the defection rate increases. However, individuals in the position of player 2 rely largely on trust in attempting to achieve the cooperative outcome, with no decline in play toward the cooperative outcome.

William Harbaugh, Kate Krause, Steven Liday, and Lise Vesterlund, in chapter 11, examine how children in third, sixth, ninth, and twelfth grades play the game of trust as designed by Berg, Dickhaut, and McCabe (1995). They do find trusting behavior among children. Even eight-year-olds pass at least one token to the trustees in this game and thus do not behave consistently with the game-theoretic prediction. On the other hand, they find the children and teenagers in their study to be somewhat less trusting than the adults in the Berg, Dickhaut, and McCabe study and the extensive replications of this study with adult subjects. Consistent with a similar finding by T. K. Ahn, Elinor Ostrom, David Schmidt, and James Walker, in the following chapter, Harbaugh and colleagues find a negative relation between the passing of funds from a truster to a trustee and the response made by subjects to a general question used extensively to measure trust on national sample surveys. Trust appears to be a complex concept strongly affected by context. General survey responses

may not be a valid measure of the level of trusting behavior found in particular situations.

In chapter 12, Ahn, Ostrom, Schmidt, and Walker turn the focus to play in one-shot and repeated play prisoner's dilemma games. The authors concentrate on four attributes of strategic games as components in understanding cooperation, trust, and reciprocity: pecuniary benefits, player types, information about player types, and the linkages between players that occur in repeated game situations. Results from several experimental decision settings are presented. Cooperation is shown to be significantly related to previous gains from cooperation. The level of cooperation and trust is also shown to be minimal in situations in which reputations for trust cannot be established. In settings in which players have the opportunity to build associations with others of their type, gains from association are larger, less dependent on game parameters, and more dependent on the exact nature of repeated play.

Toshio Yamagishi, in chapter 13, confronts the common characterization of American society as relatively less trusting and Japanese society as relatively more so. Consistent with the findings reported by Buchan, Croson, and Johnson (1999) discussed earlier, Yamagishi challenges the presumption that national identities are reflective of the degree of general trust in a society. Yamagishi provides clear evidence that what has been labeled in Japan as "trust" might be better conceptualized as assurance of mutual cooperation in commitment relations. He argues that trustworthy behavior is assured in Japanese society by the nature of the incentives surrounding exchange partners. "This assurance of mutual cooperation in Japan, however, does not imply that the Japanese are generally trustful." Reporting on a series of experiments conducted by himself and his colleagues, Yamagishi shows that in the context of games requiring trust, but in the absence of mutual commitment relations, Japanese and American subjects behave in fundamentally similar manners. Yamagishi develops an institutional view of culture that is useful in understanding what seems to be contradictory evidence between survey instruments and experimental behavior. Building on the idea of the link between social uncertainty and cooperation, Yamagishi shows that one can predict alternative institutional arrangements in associations dependent on the extent of a lack of common information present in particular settings.

Conclusion

Our conclusions are similar to those expressed by Margaret Levi in chapter 14 of this volume. Whereas she started out a skeptic of the

usefulness of experiments, we started out, and continue to be, strong enthusiasts of the experimental method. We are glad that the experience of participating in the two workshops held at the Russell Sage Foundation and a careful reading of the final set of papers has transformed her views. Although the experimental method is certainly not the only way to study trust and trustworthiness, we are confident that a great deal has been learned by utilizing the experimental method to test an evolving set of theories relevant to the study of productive social and economic relations.

Another development that has grown out of experimental studies of trust and reciprocity is a better connection to the biological foundations of human behavior. The determinism of some forms of social Darwinism so alienated many social scientists that the important presumption that human development has been shaped by evolution has repeatedly been rejected. Many of the chapters in this volume presume that humans are also evolved creatures. As such, they have the capacity to learn cultural norms and institutional rules that affect the incentives they face and their resulting interactions. These are the factors that strongly affect how trust can develop in ongoing situations.

Although the original intent of this volume was to study trust using the experimental method, we are pleased that the effort broadened to include a wider discussion of conceptual and modeling issues that must be faced in studying trust. It is, of course, a truism of the social sciences that more research is needed. In this case, more research is indeed needed to sort out exactly which combinations of contextual variables are most conducive to high levels of trust and trustworthiness. In addition to the normative foundations of trustworthy behavior, knowledge of the "other," repeated interactions, and the strong possibility of future interactions are strong predictors of both trustworthy and trusting relationships. How these can be enhanced by various structural conditions will provide a rich set of questions for future work.

Notes

1. Joseph Henrich and colleagues (2001) briefly report on an experimental study conducted in fifteen small-scale cultural groups from multiple continents. Each group was asked to play *ultimatum, public-good,* and *dictator games.* The variance in the level of the amount of funds offered by the first player in ultimatum games is substantial and varies in a systematic way with the general cultural orientation of the groups.
2. Still further experiments with the basic trust game have been reported by Friedel Bolle (1998), René Fahr and Bernd Irlenbusch (2000), and Ernst

Fehr, Simon Gächter, and George Kirchsteiger (1996). Jörg Rieskamp and Gerd Gigerenzer (2001) report on an interesting effort to program the heuristics that individuals tend to use when faced with situations involving the structure of the basic trust game. Thomas Gautschi (2000) provides yet a further test of the findings in the trust game, with consistent results.

3. A modification of the trust game was run by Werner Güth, Peter Ockenfels, and Markus Wendel (1997). In this experiment, they had the players experience the game once and then bid for the roles of player 1 and player 2. Very low levels of trust were extended when the game was played by those who had bid for these positions (and low levels of reciprocity were extended by those who were trusted). Thus, as Karen Cook and Robin Cooper stress in chapter 8 of this volume, context makes a substantial difference in how trust is developed.
4. The position that natural selection picks out optimal designs is vigorously debated in contemporary evolutionary biology literature.
5. The experiments by Kathleen Valley, Joseph Moag, and Max Bazerman (1998) on the effect of face-to-face communication in two-person bargaining games are particularly relevant to the question of goodwill accounting.
6. Given the number of disciplines represented in this volume, some technical terms used will be unfamiliar to some readers. We hope the glossary helps all readers understand the terms used in this book.

References

Arrow, Kenneth. 1974. *The Limits of Organization.* New York: Norton Press.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10(1): 122–42.

Boehm, Christopher. 1999. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior.* Cambridge, Mass.: Harvard University Press.

Bolle, Friedel. 1998. "Reward Trust: An Experimental Study." *Theory and Decision* 45(1): 83–98.

Buchan, Nancy R., Rachel T. A. Croson, and Eric J. Johnson. 1999. "Getting to Know You: An International Experiment of the Influence of Culture, Communication, and Social Distance on Trust and Reciprocation." Working paper. Madison, Wisc.: University of Wisconsin, Department of Marketing.

Buskens, Vincent, and Jeroen Weesie. 2000. "An Experiment on the Effects of Embeddedness in Trust Situations: Buying a Used Car." *Rationality and Society* 12(2): 227–53.

de Waal, Frans. 1996. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals.* Cambridge, Mass.: Harvard University Press.

Dickhaut, John, John Hubbard, Kevin McCabe, and Vernon Smith. 1997. "Trust, Reciprocity, and Interpersonal History: Fool Me Once, Shame on You, Fool Me Twice, Shame on Me." Working paper. Tucson, Ariz.: University of Arizona, Economic Science Laboratory.

Fahr, René, and Bernd Irlenbusch. 2000. "Fairness As a Constraint on Trust in Reciprocity: Earned Property Rights in a Reciprocal Exchange Experiment." *Economic Letters* 66(3): 275–82.

Fehr, Ernst, Simon Gächter, and George Kirchsteiger. 1996. "Reciprocal Fairness and Noncompensating Wage Differentials." *Journal of Institutional and Theoretical Economics* 152(4): 608–40.

Gautschi, Thomas. 2000. "History Effects in Social Dilemma Situations." *Rationality and Society* 12(2): 131–62.

Güth, Werner, Peter Ockenfels, and Markus Wendel. 1997. "Cooperation Based on Trust: An Experimental Investigation." *Journal of Economic Psychology* 18(1): 15–43.

Hamilton, William D. 1964. "The Genetic Evolution of Social Behavior." *Journal of Theoretical Biology* 7(1): 1–52.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "Cooperation, Reciprocity, and Punishment in Fifteen Small-Scale Societies." *American Economic Review* 91(May): 73–78.

Khalil, Elias L. 1994. "Trust." In *The Elgar Companion to Institutional and Evolutionary Economics,* edited by Geoffrey M. Hodgson, Warren J. Samuels, and Marc R. Tool. Hants, England: Edward Elgar.

Rieskamp, Jörg, and Gerd Gigerenzer. 2001. "Simple Heuristics for Social Interactions: The Role of Trust, Reciprocity, and Fairness." Working paper. Berlin: Max Planck Institute for Human Development.

Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. 1998. "Not So Different After All: A Cross-Discipline View of Trust." *Academy of Management Review* 23(3): 393–404.

Trivers, Robert. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46(4): 35–57.

Valley, Kathleen, Joseph Moag, and Max H. Bazerman. 1998. "A Matter of Trust: Effects of Communication on the Efficiency and Distribution of Outcomes." *Journal of Economic Behavior and Organization* 34(2): 211–38.

Zak, Paul J., and Stephen Knack. 1998. "Trust and Growth." Working paper. Claremont, Calif.: Claremont Graduate University, Department of Economics.