Journal of Economic Behavior & Organization
Vol. 55 (2004) 93–104

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

Book reviews

Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research
Elinor Ostrom, James Walker (Eds.); Russell Sage Foundation, New York, NY, 2003, xiii and 409 pages, Index, US$ 39.95

Trust and Reciprocity is the sixth book in the Russell Sage Foundation's series on trust. This collection of essays differs from the other books in the series in two fundamental ways. First, every chapter of this book draws heavily on the results of behavioral experiments to explain how trust (and reciprocity) develop and a few of the chapters present important new experimental results. A second theme, and the second way that this book differs from others, is that the research described in this volume is truly interdisciplinary and therefore provides valuable perspectives not learned in the graduate economics curriculum. While many of the contributors are economists (Eckel, Harbaugh, Krause, McCabe, Schmidt, Smith, Vesterlund, and Walker), just as many are political scientists (Ahn, Hanley, Hardin, Levi, Morikawa, Orbell, Ostrom, and Wilson), and important contributions come from psychologists (Kurzban and Yamagishi) and sociologists (Cook and Cooper). For that matter, one of the most interesting chapters is written by an animal behaviorist (de Waal).

This book is an invaluable review of behavioral research conducted on trust and reciprocity. It is clear that this book will soon find its way onto the shelves of the most active behavioral researchers, as well as, into the backpacks of many students who are interested in prosocial behavior. Following (Arrow, 1974), the book identifies the problem of trust as one of the fundamental issues in the social sciences. Trust seems paradoxical to economists who know the behavioral literature. Time and again, experimental participants achieve Pareto superior outcomes while researchers scratch their heads and ask why homo economicus would trust others enough to take actions that lead to better outcomes for all when he must make himself vulnerable to exploitation in the process? And, moreover, why is he usually not disappointed in the trustworthiness of his counterparts? Granted, identifying the fact that behavior does not match the predictions of standard game theoretic models is old news. Thankfully, the contributors to this volume spend little time rehashing this point. Instead, their work is focused on moving past the straw man built on asocial preferences to examine the determinants of this prosocial behavior.

The book is organized into five broad sections. Part 1 introduces social dilemmas and the issues important to trust and reciprocity research. Part 2 provides the evolutionary rationale for trusting behavior. Part 3 examines the cognitive factors that provide foundations for the behavior we see in the lab. Part 4 presents a few new experiments and provides a review of the existing experimental literature on trust. Concluding thoughts are expressed in Part 5.

Part 1 begins with the introductory essay (Chapter 1) by Elinor Ostrom and James Walker. Here the authors hit the mark by stating that, now that the amazement with the fact that the asocial preferences model does not predict has worn off, instead of focusing entirely on the average behavior in experiments, we should shift attention to explaining the heterogeneity of behavior. At the individual level, a few people never trust or cooperate while another few seem completely altruistic. However, the majority of participants seem conditionally trusting, trustworthy, and cooperative. These people take clues from the institutional structure of the interaction, the payoffs of various actions, the perceived intentions of their counterparts, and the social context in which the interactions are framed. These differences in behavior are the focus of much of this book.

Elinor Ostrom's essay on the development of a behavioral theory of trust and reciprocity (Chapter 2) is a masterful synthesis of the literature, especially her discussion of bounded rationality (pp. 40–49). In the first part of this chapter, Ostrom catalogues six empirical regularities seen in many social dilemma experiments on which to base a behavioral theory. She notes that (1) initial participant behavior falls between the social optimal and the subgame perfect Nash equilibrium of most social dilemma experiments, (2) cooperation decays slowly with repetition, (3) communication increases cooperation, (4) the Nash prediction is rarely seen at the individual level, even at the end of many experiments, (5) participants do not seem to use backward induction, and (6) participants are often better at solving second-order dilemmas to provide rules to govern first-order behavior than they are at simply solving the first-order problem.

In the second part of the chapter, Ostrom outlines a model of behavior in which agents (1) learn from interactions with others, (2) apply a process, now commonly called image scoring (see Nowak and Sigmund, 1998), in which they remember those who have been cooperative or trustworthy, (3) use their memories and other clues about the likely trustworthiness of a counterpart (e.g. in-group status, emotional clues) when deciding whether to trust, (4) invest in reputations for being trustworthy, (5) punish free riders at some personal cost, and (6) have time horizons that extend past immediate interactions. So where do these characteristics come from? Ostrom's argument is much subtler than most who use evolutionary reasoning. In her view, these behavioral heuristics are not linked directly to some reciprocity or trust gene, but are due to the interaction of evolved human cognitive capabilities and socialization. Specifically, Ostrom (and many of the other contributors to this volume) contends that we are born with the capacity to solve social dilemmas using the tools listed above, but will be unable to do so if we do not learn how to use these tools.

Russell Hardin's contribution (Chapter 3) decomposes trust. The chapter is useful because he is specific about what trust is and how one should measure it. He focuses on defining and understanding trust in three situations. One-way trust is a situation in which only one person in a dyad must trust the other (e.g. the classic hold-up situation of investing in firm-specific human capital). This sort of situation is captured by the (Berg et al., 1995) investment game in which one person can invest any portion of her US$ 10 show-up fee in an asset that triples the value of the investment but is owned exclusively by another person. This second person can return any part of the gross investment to the first person, but would never do so in the subgame perfect equilibrium. The second trust situation is called mutual trust and is captured by the simultaneous-move prisoner's dilemma. Here both people in the dyad have to trust the other. The last situation is based on the idea of "thick" relationships which implies that interactions occur within a relatively intimate group so that asocial behavior might be reported to others. This is obviously linked to the idea of indirect reciprocity and image scoring which is an important part of Ostrom's theory.

Hardin is more focused on the equally important topic of trustworthiness. To him, the important question is: why do people behave trustworthy in these three situations? According to Hardin, we behave trustworthy in one-way situations that are repeated because we care about future interactions (i.e. we fear trigger strategies). In mutual situations we are trustworthy because we do not want the other party to pull out of a beneficial relationship, and in thick relationships we worry about our reputations. Hardin goes onto discuss experiments that are consistent with these motives for being trustworthy.

Part 2 begins with Rob Kurzban's summary of the evolutionary psychological foundations of trust and reciprocity (Chapter 4). This is probably one of the best summaries of this viewpoint that exists. Kurzban explains Robert Trivers' idea of reciprocal altruism in which socially efficient outcomes can be maintained in repeated social dilemmas based on conditional strategies such as tit-for-tat. For economists, reciprocal altruism, essentially, focuses on the subset of Pareto efficient outcomes that can be supported by the folk theorem and appropriately patient agents. While reciprocal altruism looks like nothing more than enlightened self-interest to a game theorist, to an evolutionary psychologist, the argument is more nuanced. What is important is not the logic of the strategy to which behavior is observationally equivalent, it is the evolved proximate mechanisms that are part of our biology and let us know when the payoffs constitute a social dilemma and when someone will defect.

Kurzban goes on to identify the preconditions for the evolution of reciprocal behavior in humans and links these conditions to the evolution of our cognitive capacities (e.g. our memories to keep tract of good and bad partners and our ability to detect cheating). He then makes an interesting link to trust by extending this theory to situations in which resources are not shared simultaneously. These are the situations that dominated evolutionary history and surely coincide with the development of human trust. In a sentence, Kurzban links our evolved capacity to anticipate greater future gains (along with the other evolved traits like cheater-detection) to the willingness of humans to trust each other. That is, trust evolves when humans become patient.

The second chapter in Part 2 is authored by Frans de Waal who studies the behavior of non-human primates (Chapter 5). One way to theorize about the evolutionary foundations of prosocial behavior humans is to exploit the link between humans and chimpanzees with whom we share most of our genes. The basic idea is straight-forward. Because chimps are like less cognitively evolved versions of humans (although they have followed a different evolutionary path), if we identify reciprocal and trusting behavior among them, we have some idea of the early evolved mechanisms that directed the development of our own brains and behavior.

This chapter is particularly interesting in the context of the chapters by Kurzban and Ostrom who both call attention to the importance of our evolved capacity to remember the actions of others. In his chapter, de Waal describes his observational work with chimps in which he has discovered that chimps remember favors they owe to others (e.g. being groomed) for about 2 hours (which is longer than most of my students). In addition, he shows that chimps understand and use tit-for-tat, or in this case I guess we should call it I will scratch your back if you scratch mine (as long as more than 2 hours does not pass in-between).

The two contributions in Part 3 describe the links between cognition and trust. While the two papers have a common theme, their methods could not be more different. I remember telling Kevin McCabe in 1998 that neuroeconomics was a research agenda that only a tenured professor could undertake because it seemed so risky. I am glad to see his pioneering work has paid off. So what is neuroeconomics? According to the Center for Neuroeconomics Studies' website, neuroeconomics "investigates the neurophysiology of economic decisions. Its researchers draw on economic theory, experimental economics, neuroscience, endocrinology, and psychology to develop a comprehensive understanding of human decisions."

McCabe's chapter (Chapter 6) is a perfect example of what we can learn from this research. He demonstrates the cognitive and biological foundations of human decision-making by conducting trust-related experiments while people are having their brains imaged by magnetic resonance imaging (MRI) scanners. MRI scanners determine differences in the density of blood flowing to different regions of the brain and therefore, one can see whether a region of the brain "lights up" when people are confronted with social dilemmas. Moreover, if the lights do come on in a region of the brain that researchers know corresponds to social reasoning (e.g. the prefrontal cortex), then we have linked reciprocal behavior in games to a part of the brain that has evolved along with our capacity to live socially with each other. This is exactly the sort of evidence provided by McCabe at the end of this chapter. My only criticism is that he does not tell us more.

McCabe's chapter is juxtaposed against Chapter 7 in which James Hanley and his coauthors John Orbell and Tomonori Morikawa report on simulations they conducted to learn about the role of conflict in assuring cooperation in social dilemmas. This simulation is clever because it demonstrates that cooperation can be assured when agents are given the choice to abandon the prisoner's dilemma for a hawk-dove game. The logic of why this choice is important is not obvious and therefore illustrates the power of agent-based simulation. Essentially, the hawk-dove game is an assortation device that forces defectors to be more likely to interact with each other than with cooperators. When defectors cannot exploit cooperators they wane in the population.

The experimental papers are collected in Part 4 of the book. In their contribution (Chapter 8), Karen Cook and Robin Cooper provide a survey of the social psychological research conducted on social dilemmas and trust. While their focus is on the social contextual determinants of trust which include mechanisms for communication among the players and the social roles of the players, they make a good point early in the chapter that much research confounds trust and cooperation. Highlighting an early study by Morton Deutsch the authors point out that Deutsch showed that people with cooperative predispositions are more likely to cooperate in a prisoner's dilemma and that he refers to this as making a "trusting choice" (Deutsch, 1960). This is problematic because treating a cooperative act as measuring trust means that trust is simultaneously a feature of the relationship between the players (and therefore determines player orientations) and a feature of behavior (cooperation = trust). That is, defining cooperation as a trusting act means that we have not learned anything about the link between trust and cooperative predispositions. They then describe a study by Orbell et al. (1984) which does establish this link by allowing players to decide whether or not they want to play a prisoner's dilemma before choosing whether to cooperate or not. Here, choosing to play is an act of trust which can then be correlated with whether or not one cooperates.

Chapter 9 by Catherine Eckel and Rick Wilson complements Kurzban's discussion of evolutionary psychology in Chapter 4. Evolutionary psychology is partially founded on the idea that our brains have developed specialized modules that deal particularly well with the prehistoric environments in which we evolved. One of the more popular notions is that humans developed the capacity to read the intentions of others or, more broadly stated, we have developed a "theory of mind" (Baron-Cohen, 1995). Being adept at reading each other's intentions is a handy trait because it provides another mechanism by which cooperators or trustors can discriminate among potential partners. As Skyrms (1996) and Ameden et al. (1998) show, assortative interactions solve social dilemmas.

Eckel and Wilson describe two experiments to test whether people can discriminate among partners by picking up on facial cues. While none of the experiments involve poker playing, the idea is similar, successful social dilemma players will be good at reading the intentions of other because their intentions are transmitted through facial expressions. This contribution demonstrates the sort of scientific inquiry we should all aspire to for one simple reason. Eckel and Wilson do not just show us the experiment that "worked" (i.e. that provided data consistent with their hypothesis), they also present results from an experiment that did not work. This is important because the "non-results" can tell how robust the result is. In a nutshell, the authors find, in their second experiment, that people do condition their trusting behaviors on clues they receive from others. One of the cues that seems to signal trustworthiness is a smile.

Kevin McCabe and Vernon Smith develop a model of goodwill accounting in Chapter 10 and survey a number of the experiments they have conducted as evidence supporting the model. The goodwill model is a novel and powerful idea which has the important feature that agents account for, and remember, the choices and intentions of others when deciding how to behave. At the same time, their model is also simpler and more tractable than many of the intentions-based models currently in the working paper stage of development. In terms of the evidence, McCabe and Smith demonstrate that experimental outcomes depend on whether games are played in extensive form or normal form which is consistent with the idea that players use information in the extensive form that is not available in the normal form and this information matters to their mental accounting.

As mentioned earlier in the discussion of Ostrom's chapter, the current thought among researchers is that evolution provides us with a trust and reciprocity switch and it is up to society to either turn the switch on or not. One way to assess the degree to which trust and reciprocity are learned behaviors is to have children of different ages play trust games. Chapter 11, by William Harbaugh, Kate Krause, Steven Liday, and Lise Vesterlund reports on just such an experiment. In this experiment, children from grades 3, 6, 9, and 12 play an extensive form trust game with an anonymous member from each of the other grades. They find that children are generally less trusting than adults but they do not find much variance among the grades, which they conclude suggests that the bulk of socialization must happen at a very young age or at some time in the future.

The chapter by T.K. Ahn, Elinor Ostrom, David Schmidt, and James Walker (Chapter 12) reports on prisoner's dilemma experiments conducted to determine the effects of changes in the relative payoff structure of the dilemma and in the length of interactions on the amount of cooperation. Two payoff comparisons are important in the prisoner's dilemma. Greed is the difference between the payoff to defecting on someone and the payoff to cooperating with her and fear is the difference between the payoff to defecting on a defector and being "suckered" by a defector. The hypotheses is that cooperation should be decreasing in fear and greed. In their one-shot treatment, cooperation rates are low (approximately 30 percent) and do not seem to vary significantly with changes in fear and greed. However, in the repeated setting, their hypotheses are confirmed. These results are particularly interesting because they suggest that payoff differences only matter when decision-makers can judge the trustworthiness of their partners based on past behavior.

Toshio Yamagishi summarizes his research on trust differences between Japan and the United States in Chapter 13. This is a well-structured review of this work. Like Hardin (Chapter 3), Yamagishi is very clear about what he means by trust. People responding prosocially in their own interests is what Yamagishi calls assurance and he differentiates this from people responding prosocially because they expect that others care about what happens to them, which he calls trust. For example, cooperating in a completely anonymous, one-shot social dilemma is trust but cooperating because you know that there are punishers out there who have the ability to reduce your payoff is assurance. Yamagishi debunks the common myth that the Japanese are more trusting than Americans. Instead, he shows that Japanese culture is simply better at constructing assurance providing institutions than American culture is. In this sense, many more prosocial acts in Japan are coerced.

Part 5 of the book consists of a gentle critique of the research by Margaret Levi (Chapter 14) and a summary of the book's major results by Elinor Ostrom and James Walker (Chapter 15). Levi brings up the perennial criticism of laboratory experiments—external validity, but also spends time being critical of the loose definitions of trust that many of us use in our own work. Levi also reiterates the importance of adding social context to these experiments (I return to this below).

I conclude by emphasizing two recurring thoughts I had while reading this remarkable book. The first thought has to do with social context and the second concerns the standard Berg et al. investment game used in much of the trust research in economics. Social context is put forth as a major concern by many authors in this volume and many interesting results concerning social context are highlighted in the book. However, one important component of social context is emphasized much less—framing. As illustrated in Kahneman and Tversky (2000), the way experiments are present to subjects matters just like the way survey questions are worded matters.

One might argue that experiments designed to test various aspects of game theory should be framed as neutrally as possible (e.g. action A versus action B), but this is not the purpose of the experiments that are meant to uncover the foundations of trusting behavior. Here, we hypothesize that people react to social cues when deciding how to behave and therefore to learn anything we need to provide these cues. For example, it is interesting that only the pointed trust questions (e.g. do you loan small amount of money to your friends) asked by Ed Glaeser and his colleagues predict behavior to any extent in their trust experiments (Glaeser et al., 2000). The point is that it is time for a series of experiments that systematically varies the framing of the instructions in trust experiments to learn more about the subtle social cues we use to decide whether to trust each other or not.

Much of the research mentioned in this book is based on a variant of the Berg et al. investment game. There seems to be tacit agreement among experimentalists that the investment game measures trust. Recent research should make us question whether the investment game is as clear to our participants as it is to researchers. For example, Jim Cox's recent work on the investment game (e.g. Cox, 1999) using a triadic design seems to indicate that trust is confounded by altruism. That is, people may send money without any expectation of it being returned. Likewise, the data described in Eckel and Wilson (2002) and Bohnet and Zeckhauser (2003) indicate that the investment game may measure the risk preferences of people as much as it measures their trust in others. In the future it will be interesting to see what residual behavior in this game can be cleanly identified as trust.

References

Ameden, H., Gunnthorsdottir, A., Houser, D., McCabe, K., 1998. A minimal property right system for the sustainable provision of public goods. Working Paper, Department of Economics, George Mason University.

Arrow, K., 1974. The Limits of Organization. Norton Press, New York, NY.

Baron-Cohen, S., 1995. Mindblindness: An Essay on Autism and Theory of Mind. MIT Press, Cambridge, MA.

Berg, J., Dickaut, J., McCabe, K., 1995. Trust, reciprocity and social history. Games and Economic Behavior 10, 122–142.

Bohnet, I., Zeckhauser, R., 2003. Trust, risk and betrayal. Working Paper, Kennedy School of Government.

Cox, J., 1999. Trust, reciprocity, and other-regarding preferences of individuals and groups. Working Paper, Department of Economics, University of Arizona.

Deutsch, M., 1960. The effect of motivation orientation upon trust and suspicion. Human Relations 13, 123–139.

Eckel, C., Wilson, R., 2002. Whom to trust? Choice of partner in a trust game. Working Paper, Virginia Tech Department of Economics.

Glaeser, E., Laibson, D., Scheinkman, J., Soutter, C., 2000. Measuring trust. The Quarterly Journal of Economics 65, 811–846.

Kahneman, D., Tversky, A. (Eds.), 2000. Choices, Values, and Frames. Cambridge University Press, Cambridge, UK.

Nowak, M., Sigmund, K., 1998. Evolution of indirect reciprocity by image scoring. Nature 393, 573–577.

Orbell, J., Schwarz-Shea, P., Simmons, R., 1984. Do cooperators exit more readily than defectors? American Political Science Review 78, 147–162.

Skyrms, B., 1996. Evolution of the Social Contract. Cambridge University Press, New York, NY.

Jeffrey P. Carpenter
Department of Economics, Middlebury College
Munroe Hall, Middlebury, VT 05753, USA
Tel.: +1-802-443-3241; fax: +1-802-443-2084
E-mail address: jpc@middlebury.edu
(J.P. Carpenter)

Available online 24 May 2004

doi: 10.1016/j.jebo.2003.07.005

The Gifts of Athena: Historical Origins of the Knowledge Economy

Joel Mokyr (Ed.); Princeton University Press, Princeton, NJ, 2003, 376 pages, Index (US$ 35.00)

Joel Mokyr's book is a summary of essays and lectures from the late 1990s addressing various aspects of scientific and technological knowledge and the historical evolution of