The Fragility and Robustness of Trust*

Jim Engle-Warnick† Robert L. Slonim‡

August 22, 2001

Abstract

Although it is well known that trust is an important component of the fulfilment of incomplete contracts, less is known regarding how robust it is to past experiences. We present an experiment in which trust is required for transactions to occur, and nature provides a series of shocks along the path of play. Although the shocks have a short-term impact, we find that trust is surprisingly robust in the long-term. We argue that trust, though fragile in one way, is in another way more robust and stable over time than previously known. The results shed light on the resilience of economic institutions with incomplete contracts.

JEL classification: $C72, C91$

Keywords: Trust, Repeated Games, Experimental Economics

*We thank David J. Cooper, John Duffy, Ellen Garbarino, Susan Helper, John Miller, Jim Rebitzer, Alvin E. Roth for useful comments and discussions. We also received valuable suggestions from participants at the 2000 Economic Science Association Meetings and the University of Oxford Department of Economics Experimental Economics Workshop. The authors gratefully acknowledge financial support from the Economic Departments at the University of Pittsburgh and the Weatherhead School of Management, Case Western Reserve University.

†Corresponding Author: Economics Group, Nuffield College, Oxford University, Oxford OX1 1NF; Tel: 011-44-(0)1865 278–618; Fax: 011-44-(0)1865 278–621; email: jim.engle-warnick@nuf.ox.ac.uk; web page: http://www.nuff.ox.ac.uk/economics/people/engle-warnick.htm.

‡Department of Economics, Weatherhead School of Management, Case Western Reserve University, Cleveland, OH, 44106; email: rls18@po.cwru.edu http; web page: http://weather.cwru.edu/slonim.

1 Introduction

It is well known that people have many behavioral predispositions that affect the fulfilment of incomplete contracts. Since many real-world contracts are incomplete (Williamson 1985), understanding the determinants and the effects of these behaviors is important for understanding real-world outcomes. Since it is difficult to isolate individual behaviors and incentives in the field, a fruitful avenue to study them involves the use of laboratory experiments (e.g., Fehr and Gaechter, 2000). Experiments show, for example, that concerns for fairness can affect bargaining behavior (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999; Roth, 1995), trust can affect investment behavior (Berg, Dickhaut, and McCabe, 1995), cooperation can affect contributions to public goods (Andreoni, 1988) and reciprocity can affect gift giving (Fehr, Kirchsteifer, and Reidl, 1993).

What is less well known is how sensitive these behaviors are to players' experiences as games are repeatedly played, i.e., it is not known how sensitive they are to the history of the game. Understanding the effect of a game's history on behavior is important to better understand markets.[1] If, for example, a behavior that is crucial for transactions to occur is fragile with respect to a game's history, then careful design of a market institution in its early stage may be critical to its ultimate success. Since in the field the history of the game is both difficult to study and impossible to control, we report a new set of economics experiments.

This paper experimentally studies the effect of the history of a game on trust. In our experiment subjects play an indefinitely repeated trust game many times. In an indefinitely repeated game the players know the distribution, but not the realization, *a priori*, of the random length of the repeated game (i.e., the players do not know how many times a stage game will be played). The stage game requires an action that cannot be contractually guaranteed. The players are not permitted to communicate, identify each other, or in any way build a reputation that can carry through to future relationships. Thus the experiment provides a particularly tough institutional setting for the robustness of trust.

In this setting we find that trust declines within repeated games, but resets at the start of each new repeated game when subjects play against new opponents.[2] We also observe, similar to behavior in one-shot trust games, substantial trust and reciprocity at the start of each repeated game. Thus, within an individual relationship trust is sensitive to the history of the repeated game, but across relationships hope springs eternal.

What is really new is that by examining *many repetitions* of indefinitely repeated games, we can vary the realized lengths of the repeated game in a way that leads to a better understanding of trust across relationships. Though the realized length of an indefinitely repeated game is irrelevant with respect to the calculation of equilibria, theories of adaptive learning predict both short-term and long-term effects from unusually long and short games. We refer to these unusually long and short games, which differ from the expected length of the game, as positive and negative shocks.

We find that the level of trust in the short-term is sensitive to the realized length of the previous repeated game. Consistent with theories of adaptive learning and, in particular the Law of Effect (Thorndike, 1928; Erev and Roth, 1998), we find that there is significantly more trust and reciprocity immediately after positive shocks and significantly less after negative shocks. This shock-reset effect is remarkable in view of the fact that trust declines within each repeated game; although there is less trust at the end of longer than shorter repeated games, trust resets to higher levels after longer repeated games and resets to lower levels after shorter ones. Thus, a realization of uncertainty that is irrelevant with respect to computation of the repeated game's equilibria affects trust and reciprocity in the population. In our experiment, in the short-term, trust really is fragile.

We examine the long-term consequence of the fragility of trust by manipulating the history of the game so that all subjects are exposed to exactly the same lengths of the repeated games, but in different orders. Theories of adaptive learning and, in particular the Power Law of Practice (Blackburn, 1936; Erev and Roth, 1998), suggest that subjects will be less trusting and reciprocating the earlier they are exposed to the negative shocks and more trusting and reciprocating the earlier they are exposed to the positive shocks. However, we

find that the order of the shocks does not affect the ultimate level of trust or reciprocity. Thus, despite decreasing trust within relationships and despite negative shocks from nature, trust prevails. In our experiment, in the long-term, trust really is robust.

In our experiment, trust, though fragile in the short-term, is ultimately robust. The paper is organized as follows. We first discuss previous trust research and then describe our experimental design. We next present our hypotheses followed by the empirical results. We then conclude and discuss areas for further research.

2 Why Study Trust?

The literature stresses that trust, or more broadly social capital (Putnam, 2000), may lead to better functioning institutions and facilitate economic transactions. Arrow (1972) suggests that most transactions conducted over a period of time involve an element of trust, and refers to trust as an "important lubricant of a social system" (Arrow, 1974). Fukuyama (1995) argues that the level of trust influences the degree of economic success in markets. Empirical work focuses on how factors such as individual experiences and community characteristics affect trust (Alesina and La Ferrara, 2000) and the positive relation of trust to economic growth (e.g., Knack and Keefer, 1997; Zak and Knack, 2001).

Experimental research has explored the existence, effects and determinants of trust. Berg, Dickhaut, and McCabe (1995) and Ortmann, Fitzgerald, and Boeng (2000) find substantial levels of trust in one-shot relationships. In the context of a market institution, Fehr and Gaechter (2000) find that introducing partially enforceable contracts may actually crowd out trust in one-shot relationships. Glaeser, Laibson, Scheinkman, and Soutter (2000), bridging a gap between empirical and experimental data, find that standard attitudinal survey questions about trust actually predict trustworthy behavior, but past behavior is a better predictor of trust. Fershtman and Gneezy (2001) study an empirical phenomenon (ethnic discrimination in Jewish Israeli society) in the laboratory using one-shot lab trust games.

This paper differs from the existing experimental literature by examining trust when

people interact with each other many times.[3] Similar to much of the empirical and experimental literature, we are interested in examining how factors, specifically how shocks, affect the level of trust. We study this problem in the context of a repeated game in which there are multiple equilibria, some of which involve strategies that include trust and some of which do not (Kreps, 1990). Thus the level of trust obtained in this institution can be thought of as the selection of a subset of the equilibria of the game; we study the effect of shocks on the equilibrium selection process in order to better understand the role of trust in the fulfilment of incomplete contracts.

3 The Experimental Design

The experimental design consists of many plays of a repeated trust game. A repeated game, henceforth called a *supergame*, is a sequence of rounds (i.e., stage games) played between the same two players. We refer to the number of rounds within a supergame as the supergame's *length*. In the experiment, subjects play 40 supergames; the first 20 have indefinite length (i.e., an uncertain number of rounds) and the last 20 have definite length (i.e., a number of rounds that is known with certainty). We do not discuss definite length games in this paper.[4]

3.1 The Trust Stage Game

We examine the trust stage game shown in Figure 1. In this game there are two players who make decisions sequentially. We call the first and the second mover the Trustor and the Trustee. At the start of the stage game each player receives \$0.40. The Trustor then chooses between two actions: *Send* and *Don't Send*. If she plays *Don't Send*, the stage game ends and both players receive their \$0.40 endowment. If she plays *Send*, her \$0.40 doubles and is given to the Trustee; the Trustor now has \$0.00 and the Trustee now has \$1.20. At this point in the stage game, the Trustee chooses between two actions: *Return* and *Keep*. If she plays *Return*, the \$1.20 is split evenly between the players, so both receive \$0.60 and the

stage game ends. If she plays *Keep*, then she receives the entire \$1.20, the Trustor receives \$0.00 and the stage game ends.

We refer to the actions *Send* and *Return* as trust and reciprocity behavior.[5] The unique subgame perfect equilibrium of the stage game is for the Trustee to play *Keep* if the Trustor plays *Send* and, consequently, for the Trustor to play *Don't Send*. Thus, in the stage game, trust and reciprocity are not part of the subgame perfect equilibrium. Trust increases the total payoff to the players by 50%, reflecting the importance of trust. If the Trustee does not reciprocate, however, then trust also causes redistribution of income away from the Trustor. Thus, while trust is necessary to increase the total payoff, it is risky because reciprocity cannot be guaranteed contractually.

3.2 The Indefinitely Repeated Trust Game

In the indefinitely repeated supergames, there is a constant and independent probability of $p = 0.8$ after each round (i.e., stage game) is played that every Trustor-Trustee pair will remain together to play another round, and a probability of $1 - p = 0.2$ that the round just ended will be the last one in the supergame. When the supergame ends, subjects are randomly and anonymously re-paired with new opponents to play the next supergame. The $p = 0.8$ continuation probability models an infinitely repeated game with a discount factor of 0.8.

There are many equilibrium strategies in the repeated game (see Engle-Warnick and Slonim (2001) for a discussion and for evidence of them in the data). At one extreme, the Trustor will never trust (i.e., she will play *Don't Send*) and the Trustee, if the Trustor did trust, would not reciprocate (i.e., she would play *Keep*). At the other extreme, by the Folk Theorem of Repeated Games (see Fudenberg and Maskin, 1986) the minimum continuation probability for which trust (*Send*) and reciprocity (*Return*) can occur in equilibrium is 0.75. Thus, at $p = 0.8$, trust and reciprocity behavior are also consistent with equilibrium strategies.

3.3 The Lengths of the Indefinitely Repeated Games

With a continuation probability of $p = 0.8$ the expected supergame length is five rounds. The first two rows of Table 1 show the realized supergame lengths for the two treatments. The lengths were drawn from the distribution described above prior to running the first session. The treatments are denoted *Long-short* and *Short-long*; in Long-short the average length of the first 10 supergames is 6.5 rounds (i.e., long), and the average length of the last ten supergames is 3.6 rounds (i.e., short). The order of lengths in Short-long is the Long-short order played in reverse (i.e., the length of supergame 20 in Short-long equals the length of supergame 1 in Long-short, 19 maps to 2, 18 to 3, etc.). By running two treatments that vary only in the order of the supergame lengths, we can examine the persistence of negative and positive shocks on trust and reciprocity in a tightly controlled manner; in the end, all subjects experienced exactly the same shocks.

3.4 Experimental Procedures

We ran five Long-short and four Short-long sessions. We eliminated one Long-short session from the analysis because some of the data were lost when a computer failed. Of the remaining eight sessions, three Short-long sessions had 14 subjects and one had 10, and two Long-short sessions had 14 subjects and two had 10. Subjects participated in one session only. All decisions were anonymously made on computers at the Experimental Economics Lab at the University of Pittsburgh.[6]

At the start of each session each subject was randomly assigned to be a Trustor or Trustee for the entire session. At the start of each supergame subjects were randomly and anonymously paired with a partner who was different than the person who they were paired with in the previous supergame.[7] Subjects only observed the decisions of the person they were paired with. They received a $5 show up fee plus their payoffs from six supergames randomly selected at the end of the session. Subjects were not told how many supergames they would play during the session, but knew that sessions would last at most two hours. All

indefinite supergames ended at least 40 minutes before the time limit. All these procedures were common knowledge.

4 The Effect of Realized Supergame Lengths

This section discusses how positive and negative shocks can affect the level of trust and reciprocity in repeated games. Since the continuation probability is sufficiently high, there are multiple equilibrium strategies in the indefinitely repeated game. Among the equilibrium strategies are cooperative ones in which players play *Send* and *Return* in every round as well as a non-cooperative one in which *Don't Send* is played every round. Since the theory of repeated games is silent with respect to not only which equilibrium will be played, but also how the shocks affect the selection process, we turn to theories of adaptive learning to investigate how the shocks may affect play.

We first examine how one tenet of adaptive learning, the Law of Effect (Thorndike, 1998), may influence the level of trust and reciprocity immediately after positive and negative shocks and then examine how a second tent of adaptive learning, the Power Law of Practice (Blackburn, 1936), may determine whether the shocks have a transient or permanent effect.

4.1 The Short-term Effect of Shocks

According to the Law of Effect, strategies that result in good outcomes are more likely to be repeated. Thus, to apply the Law of Effect to repeated games, we must specify at least two strategies for each player. For simplicity in exposition, we assume each player type plays one of two equilibrium strategies of the repeated game. Specifically, we assume Trustors play either a *Don't Send Always* strategy or a *Grim Trigger* strategy. The Grim Trigger strategy plays *Send* in Round 1, plays *Send* in subsequent rounds as long as Trustees play *Return*, and plays *Don't Send* forever if the opponent ever plays *Keep*. We assume Trustees play either a *Return Always* or *Keep Always* strategy. The strategy pairs *Don't Send Always, Keep Always* and *Grim Trigger, Return Always* form equilibrium strategies of the repeated

game. The former pair produces no trust along the equilibrium path and the latter produces trust and reciprocity. These strategies are strongly supported in the data (Engle-Warnick and Slonim, 2001).

With these strategies, the Law of Effect predicts that longer supergames will lead to more trust and reciprocity in subsequent supergames. To see this, consider each player type separately. For Trustees, the monetary reward for reciprocating in Round 1 (i.e., playing the *Return Always* strategy) rather than not reciprocating (i.e., playing the *Keep Always* strategy) is a trade off between a lower first round payoff and a greater payoff in subsequent rounds. Assuming the Trustors play the *Grim Trigger* strategy, the Trustees receive less monetary reward ($0.60 less) in Round 1 playing *Return Always* than *Keep Always*, but receive a greater monetary reward ($0.20 more) in every subsequent round playing *Return Always* than *Keep Always*.[8] Thus, the monetary reward for reciprocating in Round 1 increases the longer the supergame.

For Trustors, the monetary reward for trusting in the first round (i.e., playing *Grim Trigger*) rather than not trusting (i.e., playing *Don't Send Always*) is a trade off of first round risk for subsequent round rewards. If Trustees are not trustworthy (i.e., *Keep Always* types), then Trustors receive a smaller monetary payoff ($0.40 less) in Round 1 by trusting rather than not trusting and an identical payoff of $0.40 in every subsequent round. However, if Trustees are trustworthy (i.e., *Return Always* types) then Trustors receive a greater monetary reward (an extra $0.20) in every round by trusting rather than not trusting; so the longer the supergame, the greater the monetary reward for trusting than not trusting in Round 1. Thus, if Trustees and Trustors are adaptive learners who react supergame by supergame to realized payoffs, then negative shocks will lead to less trust and reciprocity and positive shocks will lead to more trust and reciprocity.

4.2 The Long-term Effects of Shocks

Applying the Power Law of Practice to repeated games, we predict that the shocks will have a larger impact on the level of trust and reciprocity the earlier the shocks occur in play of the

repeated games. The Power Law of Practice states that learning curves are initially steep and then flatten out over time; it thus implies that shocks will have a bigger influence on trust and reciprocity the earlier they are experienced. Our treatments were designed with this prediction in mind. If players initially experience many negative shocks (the Short-long sessions), then these shocks will induce less trust and reciprocity over time, and although later positive shocks will occur, they will not contribute as much to learning as the initial negative ones. Likewise, if players initially experience many positive shocks (the Long-short sessions), then these shocks will induce more trust and reciprocity over time. Although later negative shocks will occur, they will not contribute as much to learning as the initial positive ones. Comparing the treatments, if subjects are adaptive learners who have learning curves that flatten out over time, then early shocks will have a larger effect on behavior and consequently we should observe less trust and reciprocity over time in the Short-long than Long-short sessions.

5 The Experimental Results

The context in which we examine our hypotheses can be summarized as follows: (1) a majority of players trust and reciprocate in the first round, (2) trust and reciprocity levels rapidly decrease across rounds within each supergame, (3) first round opponent actions strongly affect behavior in later rounds within each supergame and (4) trust and reciprocity reset to high levels at the start of each new supergame. We find support for our hypothesis that negative (positive) shocks decrease (increase) the level of trust and reciprocity in the population, yet we do not find support for our hypothesis that the shocks have a long-term impact on behavior. Thus, we find that trust and reciprocity are surprisingly resilient to problems with past opponents (regularities 2 and 3) as well as negative shocks from nature.

5.1 Descriptive Results

This section describes the data and the following sections present tests of our hypotheses. Figure 2 presents the proportion of times that Trustors play *Send* and Trustees play *Return* by round. Throughout our description of the data, we compute proportions by first calculating averages within each session and then averaging across sessions. In Figure 2 we calculate the proportions for each round separately. Figure 2 shows that the rates of trust and reciprocity begin high and decrease across rounds. The figure shows that the rate of playing *Send* monotonically decreases from nearly 90% in Round 1 to just over 30% in Round 10.[9] It also shows that the rate of playing *Return* decreases from over 80% in Rounds 1 and 2 to under 70% in Rounds 4 and 5. The relative levels of trust and reciprocity in Round 1 are an extension of one-shot game results found in Berg et al. (1995) to repeated games.

This decrease in trust and reciprocity is at least partially explained by players responding to their opponents' actions. For example, conditional on Trustors and Trustees playing *Send* and *Return* in Round 1, the proportion of times Trustors play *Send* in Round 2 is 0.949 (580/611) and the proportion of times Trustees play *Return* in Round 2 is 0.885 (513/580). However, conditional on Trustees playing *Keep* in Round 1, Trustors play *Send* in Round 2 at a rate of only 0.155 (20/129): Trustors trust over 80% less often after Trustees play *Keep* than *Return*. And conditional on Trustors playing *Don't Send* in Round 1, Trustees play *Return* in Round 2 at a rate of only 0.682 (45/66): Trustees reciprocate almost 20% less often (in Round 2) after Trustors play *Don't Send* than *Send* (in Round 1). This behavior highlights how opponents' actions affect the subsequent actions of both Trustors and Trustees, and stresses the importance of getting relationships right in the first round.[10]

Figure 2 also indicates that when comparing behavior across supergames with different lengths, it is necessary to address the fact that later rounds have lower rates of trust and reciprocity. When comparing behavior across supergames with different lengths, we therefore limit the analysis to *common rounds*. The number of rounds common to a set of supergames is simply defined as the length of the shortest supergame in the set. For example, the first

supergame in the Long-short treatment is of length twelve and the first supergame in the Short-long treatment is of length six; the number of common rounds in this set of supergames is thus six. The third row of Table 1 shows the number of common rounds by the order in which the supergames were played.

Figures 3 and 4 present the proportion of times Trustors play *Send* and Trustees play *Return* by common rounds for each of the twenty supergames. These figures show that the same pattern of play aggregated across supergames (see Figure 2) also occurs within supergames. Figure 3 reveals two additional features of the data for the Trustors. First, across supergames trust resets each time a new one begins (e.g., in Round 5 of Supergame 5 trust fell to under 50%, yet trust reset to 85% in Round 1 of Supergame 6). Second, trust gradually resets at higher levels over time. For example, in Round 1 of Supergame 1 Trustors play *Send* under 80% of the time but during Round 1 of Supergames 10-18 trust hovers around 90%.

Figure 4 reveals two additional features of the data for Trustees. First, similar to trust behavior, across supergames reciprocity generally resets each time a new one begins. Second, reciprocity jumps around across rounds within supergames. For instance, from Round 1 to 7 of Supergame 11 Trustees play *Return* 90%, 80%, 90%, 56%, 78%, 100% and 55% of the time. This irregular pattern is a consequence of Trustor behavior and the experimental design. First, since Trustors play *Send* less often as each supergame progresses we observe fewer Trustee actions in later rounds. Second, the Trustee actions we observe in later rounds are a biased sample of the Trustees.[11] Despite the unsteady graph, Figure 4 shows that Trustees within most supergames generally play *Return* less often in later than earlier rounds, and that there is a strong tendency for reciprocity to reset with each supergame regardless of how little trust occurred late in the previous supergame.

5.2 Result 1: Shocks Affect Play in Subsequent Supergames

To test whether supergame lengths affect trust and reciprocity as implied by the Law of Effect, we examine changes in the proportion of times Trustors play *Send*, p_s, and Trustees

play *Return* q_s, in Round 1 across all consecutive supergames s and $s+1$. Specifically, we examine how often $p_{s+1} > p_s$, $p_{s+1} = p_s$ or $p_{s+1} < p_s$ and how often $q_{s+1} > q_s$, $q_{s+1} = q_s$ or $q_{s+1} < q_s$ when the length of supergame s is *short* (1-3 rounds), *medium* (4-5 rounds) and *long* (6-14 rounds).[12] We only examine Round 1 actions because the minimum length supergame in the set of all supergames played is one round (i.e., the number of common rounds in the set of all supergames played is one), and because, as we have shown, Round 1 behavior critically affects trust and reciprocity in subsequent rounds. We assume that observations of sessions, but not individual subjects, are statistically independent because subjects played against the same opponent many times in each session (there were 20 supergames but only either five or seven possible opponents for each subject).

Figure 5 shows that after short supergames Trustors play *Send* less often in 62.5% of the sessions (i.e., in five of the eight sessions) and never play *Send* more often. In contrast, after long supergames Trustors play *Send* more often in 75% (6/8) of the sessions and play *Send* less often in only one session. A chi-square test shows that the difference in behavior both (1) after short and long, and (2) after short, medium and long supergames, is significant ($p < .02$). Thus, the data reject the null hypothesis that supergame lengths have no effect on trust in favor of the alternative that shorter supergames reduce trust and longer ones increase trust.

Figure 5 also shows that after short supergames Trustees play *Return* less often in 75% (6/8) of the sessions and only play *Return* more often in only one session. After long supergames, however, Trustees play *Return* more often in 75% (6/8) of the sessions and play *Return* less often in the other two sessions. A chi-square test shows that the difference between behavior both (1) after short and long, and (2) after short, medium and long supergames is significant ($p < .01$). Thus, the data reject the null hypothesis that supergame lengths have no effect on reciprocity in favor of the alternative that shorter supergames reduce reciprocity and longer ones increase reciprocity.

The effect of supergame length on trust and reciprocity is stronger and more important than it may first appear. The effect is stronger in that trust and reciprocity reset to higher

levels after longer supergames despite the fact that longer length supergames cause lower levels of trust and reciprocity to occur just prior to the reset. The effect is more important in the sense that differences in first round behavior understate the overall impact of the effect of supergame length: as mentioned previously, first round play strongly affects trust and reciprocity in subsequent rounds. In sum, longer supergames, despite ending with lower rates of trust and reciprocity than shorter supergames, lead to a higher reset in trust and reciprocity while shorter supergames lead to a lower reset.

5.3 Result 2: The Impact of Shocks on Play is Not Permanent

To test the long-term impact of supergame length on play we compare play (in the common rounds) of the Long-short and Short-long treatments. Figures 6 and 7 show the rate at which Trustors play *Send* and Trustees play *Return* by blocks of consecutive supergames for each treatment.[13] The averages are first calculated across common rounds within each session and then averaged across the sessions. Tables 2a and 2b report the averages for each session for the Short-long and Long-short treatments. The tables contain one row for each of the four sessions and a fifth row for the treatment average. The left and right half of each table present results for Trustors and Trustees. The columns are divided into consecutive blocks of supergames. For example, the upper-left cell in Table 2a indicates that Trustors in Session 2 of the Short-long treatment play *Send* at a rate of 0.551 in the common rounds of the first block of supergames (1 and 2). Table 2c reports Mann-Whitney rank order test results that indicate whether the averages in the Long-short and Short-long sessions are significantly different. Low p-values indicate the rejection of the null hypothesis that behavior across treatments did not differ.

We first note that there are no significant initial differences in behavior across treatments; Trustors play *Send* insignificantly more often in the Long-short than in the Short-long sessions during the first two supergames, and Trustees play *Return* insignificantly more often in the Long-short than in the Short-long sessions during the first two supergames.[14]

Figures 6 and 7 and Table 2c show strong support for the immediate and significant effect

of supergame length on trust and reciprocity, but no support for a permanent effect. During Supergames 3-5 Trustees *Return* significantly more often (13%) in Long-short than Short-long sessions.[15] During Supergames 6-10 Trustees continue to *Return* significantly more often in Long-short than Short-long sessions; in fact, Trustees in every Long-short session play *Return* more often than Trustees in every Short-long session during this time. This behavior is what we expected since during this time longer supergames have been occurring in the Long-short than Short-long sessions.

The difference in Trustee behavior during Supergames 3-5 and 6-10 is not only statistically significant, but is also economically important. Although the difference between the rate at which Trustors play *Send* between treatments is insignificant during Supergames 3-5 (only 5%), this difference swells to 18% during Supergames 6-10 and is statistically significant; during Supergames 6-10 Trustors in every Long-short session play *Send* more often than Trustors in every Short-long session. Further, the difference in behavior between treatments remains significant during Supergames 11-15; it continues to be the case that in every session Trustors play *Send* and Trustees play *Return* more often in the Long-short than Short-Long sessions.

During Supergames 11-15, the types of shocks across the two treatments reverse. If the shocks have a permanent effect on behavior, i.e., if the Power Law of Practice predicts behavior, then the early shocks will have affected behavior more than the later ones, and so behavior across the treatments will remain apart. However, the reversal in shocks quickly eliminates any significant difference in trust or reciprocity across treatments. During Supergames 16-20, Trustors play *Send* only 3% more often in Long-short sessions than in Short-long sessions (exactly the same difference as in the first two supergames), and Trustees play *Return* an insignificantly 7% more often in the Long-short sessions than in the Short-long sessions (during the first two supergames this difference was 9%). Thus, although subjects in Long-short sessions, who initially experience relatively long supergames, play *Send* and *Return* relatively more often than Short-long subjects during Supergames 6-15, the difference disappears soon after the relative order of the positive and negative shocks

start to reverse. Thus, the effect of shocks on trust and reciprocity is not permanent.

The appendix examines Trustee behaviour in more detail. It shows, consistent with the Law of Effect and our premise, that Trustees receive greater reinforcements playing *Keep* than *Return* during the early supergames of the Short-long but not Long-short sessions. Despite the greater reinforcement Trustees receive playing *Keep* early in Short-long sessions, we find no evidence that the level of reciprocity differs permanently across the treatments. In an effort to understand why this may be, we consider how Trustors react to Trustees playing *Keep*. If Trustors in early supergames do not harshly punish Trustees who play *Keep*, but learn to, then reinforcements for playing *Keep* will diminish over time, perhaps eventually overriding the effects of the negative shocks in early supergames. Consistent with this hypothesis, we find that the Trustors behave consistently with the *Grim Trigger* strategy in 59.8% of the first ten and 68.6% of the last ten supergames.[16] Thus, one reason that the effect of early negative shocks in the Short-long supergames may not have persisted is because Trustors were adapting a stronger punishment strategy over time.

6 Conclusion

This paper presents an experiment that tests the effect of the history of the game on trust in a population. Despite the inability of subjects to build a reputation, despite receiving negative shocks from nature, despite having negative experiences with opponents, and despite the rate of trust decreasing within supergames, we repeatedly observe a behavioral reset in which high levels of trust occur in the first round of play of each new relationship. Across supergames we find, consistent with the Law of Effect, that the lengths of the supergames have an immediate short-term impact on trust; short games lead to less trust and long games lead to more trust. However, we find, inconsistent with the Power Law of Practice, that the order of lengths of the supergames has no long-term effect on trust.

This experiment is the first to document the persistence of trust in repeated games. This evidence for the persistence of trust is important because an empirical regularity in one-

shot games is that trust is not fully reciprocated; two unanswered questions were whether this would occur in the repeated game and, if so, whether trust would nevertheless persist. Similar to one-shot games, less than full reciprocation did occur in the repeated game; yet trust persisted.

To understand the robustness of trust and reciprocity despite the effect of negative shocks, we noted that Trustor behavior becomes increasingly consistent with a harsh punishment strategy over time. The increasingly smaller payoffs that Trustees receive by not reciprocating in the presence of this growing punishment threat may override the early benefits that not reciprocating may have given them.

The results from the repeated trust game presented in this paper suggest many opportunities for further research. Understanding the factors that determine the existence and severity of punishment norms may be fundamental to understanding trust in markets; in our experiment if the Trustors had been initially too forgiving, then the Trustees may have learnd not to reciprocate, which may have lead to less trust in the long-term. The determinants of social norms and the types of enforcement that exist and evolve in population games (Kandori, 1992) can be studied by varying parameters of this experimental design such as the discount factor, the population size, the matching mechanism and the information available across relationships. Institutions for overcoming contractual problems when information is scarce (Greif, 1993; Milgrom, North, Weingast, and Barry, 1990) can be studied by adding mechanisms for transmitting information regarding past player behavior.

The experimental results add force to the use of trust to explain economic relationships in markets with incomplete contracts. Not only does trust exist in a repeated game market in which it is impossible for participants to build a reputation, it is robust to early negative shocks. Had trust been too fragile, the drop in trust could have persisted. Instead, subjects behaved as if they were experiencing a reset at the start of each supergame, even though they played the game against the very same group of people whose behavior had just changed, many times for the worse.

7 Appendix

7.1 Trustees are Reinforced to Play *Keep* in the Short-long Treatment

Consistent with our premise, we find that Trustees are monetarily reinforced to play *Keep* in short but not long supergames, thus leading to relatively less reciprocity in the Short-long than Long-short sessions during most of the supergames. Figure 8 presents scatter plots of average payoffs per round for each Trustee vs. the average payoff that her opponents receive against her. The four plots show the first and last ten supergames in the Short-long and Long-short sessions separately.

The upper left plot shows the first ten supergames of the Short-long treatment. During these supergames (1) over half the Trustees achieve average earnings that are greater than the cooperative equilibrium payoff of \$0.60 per round, (2) the average earnings range from 50 to 85 cents and (3) the points on the graph are remarkably arrayed in nearly a straight line with a visible negative slope; the less their opponents earn, the more Trustees earn. Since it is more profitable to be a Trustee whose opponents earn less, it is profitable to play *Keep*. In contrast, the lower left plot shows a much less pronounced relationship between Trustees and their opponents' payoff in the first ten supergames of Long-short sessions. These results are precisely what we expected; it pays for Trustess to play *Keep* when supergames are short.

Table 3 presents linear regressions that confirm the relationship visible in Figure 8. We regress Trustee average pay on opponent average pay during the first ten supergames and include fixed effects to control for session dependencies. For Short-long, the coefficient on opponent average pay is significantly negative (-0.63) and the adjusted R-squared is 0.85. Thus, if a Trustee's opponent's average pay decreases by one cent, then her pay increases on average by 0.63 cents. This simple model surprisingly explains 85% of the variance of the Trustee's average pay in the Short-long treatment. It really paid to play *Keep* in these sessions with relatively short supergames.

During the first ten supergames of Long-short, however, the effect of opponent's average payoff is much lower (only -0.174). Though the effect is marginally significant ($p = 0.061$), the magnitude is much less than in Short-long. Perhaps an even more dramatic indication of the weaker relationship between opponent's average pay and Trustee's average pay in Long-short than Short-long is that the simple model only explains 16% of the variance in Long-short (as opposed to 85% in Short-long). Thus, the reinforcement to play *Keep* is much stronger at the beginning of Short-long than Long-short.

Jim Engle-Warnick
Nuffield College, Oxford University
Robert L. Slonim
Case Western Reserve University

Notes

[1] For example, Greif (1993) analyzes an institution in the context of a repeated game that developed to enable agents to commit to act on behalf of merchants in eleventh-century Mediterranean trade.

[2] Johnson (1996) noted, "Second marriages are the triumph of hope over experience." This notion appears equally descriptive in repeated trust games.

[3] Trust has been studied in the laboratory using investment, gift-exchange, and centipeed games (for a survey see Camerer, 2001 and the references therein). In this paper we use a simple stage game that provides a crisp measure of trust in order to study the characteristics of trust in the repeated game.

[4] Repeated game strategies are inferred from the actions of players in both the definite and indefinite length games in Engle-Warnick and Slonim (2001). Many plays of repeated games have also been experimentally examined by Selten and Stoeker (1986) and Van Huyck, Battalio, and Walters (2001).

[5] As in Alesina and La Ferrara (2000) we abstract from the semantic issue of whether the terms trust and reciprocity are applicable when cooperation is achieved through reciprocal threat. Fehr and Gaechter (2000) use an alternative definition of reciprocity.

[6] Instructions are available from the authors.

[7] Subjects were told that they could think of the draws determining the end of the supergame by imagining a hat with one white and four red chips. At the end of each round, the computer would randomly pick one chip from the hat and then replace it.

[8] The payoff Trustees receive if Trustors play *Don't Send Always* is \$0.40 no matter what strategy Trustees play. Since Trustees are reinforced identically whenever Trustors play *Don't Send Always*, it will not affect which strategy Trustees will use, so we ignore this case.

[9] The decrease in trust within repeated games is consistent with the patterns observed in repeated play of social dilemmas with different matching mechnisms (see Andreoni, 1988; Cooper, DeJong, Forsythe, and Ross, 1996; Clark and Sefton, 2001).

[10] Given how strongly players appear to react to opponent behavior, it is interesting that Figure 2 also shows that playing *Send* increases during the last four rounds and that playing *Return* increases from Rounds 4 to 6, 8 to 9 and 10 to 13. Some of this behavior may be due to the decreasing number of observations in later rounds. However, for the Trustees this behavior most likely reflects selection bias rather than any other explanation; the Trustees who play *Keep* in early rounds are less likely to be sent to in later rounds, thus later rounds may be increasingly (decreasingly) reflecting the behavior of Trustees with a high propensity to play *Return* (*Keep*).

[11] For instance, in Round 2 Trustors send 95% of the time if Trustees played Return in

Round 1, but only 15% of the time if Trustees played Keep, thus in Round 2 we are much more likely to observe the choices of Trustees who played Return in Round 1.

[12] Separating supergames in this manner divides the data as evenly as possible; there are 7, 5 and 8 short, medium and long supergames. To compute the proportion of times Trustors play *Send* after short supergames we proceed as follows. Consider Session 1, a Long-short session with seven Trustors and seven Trustees. In this session there are seven short supergames (they are supergames 3, 4, 13, 15, 17, 18 and 19; see Table 1). We first compute the average number of times the seven Trustors play *Send* during round 1 of these seven short supergames. We next compute the average number of times these same seven Trustors play *Send* during Round 1 across the seven supergames following each short one (these are supergames 4, 5, 14, 16, 18, 19 and 20). We then compare these two averages to determine if Trustors play *Send* more often, the same amount or less often after short supergames. We then repeat this computation for every session to determine, across the eight independent sessions, how often Trustors send more often, the same amount or less often after the relatively short supergames. We then repeat this exercise for medium and long supergames.

[13] The specific supergame blocks were chosen to reflect the differences in supergame lengths across treatments. The first block includes only the first two supergames; there should be no difference between treatments at the start of the session. In Supergames 3-5, Long-short subjects previously experienced a relatively long supergame (Supergame 2, 8 rounds) whereas Short-long subjects experienced the shortest possible supergame (Supergame 2, 1 round). Thus in this block we expect play across treatments may diverge. For Supergames 6-10, we expect differences to increase across treatments since Long-short subjects will gain more experience with relatively long supergames and Short-long subjects will gain more experience with relatively short supergames. For Supergames 11-15, the difference in supergame lengths across treatments changes; now subjects receive more experience with relatively long supergames in Short-long than Long-short. Although Short-long subjects may learn to trust and reciprocate more often than Long-short during Supergames 11-15, if the learning curve is steep initially, as the Power Law of Practice predicts, then Long-short subjects should continue to trust and reciprocate more often than Short-long subjects during Supergames 16-20.

[14] We aggregate across the first two supergames to examine initial behavior. Behavior should not diverge after Supergame 1 because supergames are long in both treatments, but may diverge after Supergame 2 because Short-long subjects experience a short supergame and Long-short subjects experience a long one.

[15] We aggregate behavior during Supergames 3-5 since supergame lengths diverge even further between the two treatments. The remaining divisions divide the data evenly across the remaining supergames.

[16] Evidence for Trustors increasingly playing trigger strategies across supergames is even stronger when repeated game strategies are inferred from similar data; see Engle-Warnick and Slonim (2001) for details.

References

ALESINA, A., AND E. LA FERRARA (2000): "Who Trusts Others?," Centre for Economic Policy Research Discussion Paper No. 2646.

ANDREONI, J. (1988): "Why Free Ride? Strategies and Learning in Public Goods Experiments," *Journal of Public Economics*, 37, 291–304.

ARROW, K. (1972): "Gifts and Exchanges," *Philosophy and Public Affairs*, 1, 343–362.

——— (1974): *The Limits of Organization*. Norton, New York, NY.

BERG, J., J. DICKHAUT, AND K. MCCABE (1995): "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 10, 122–142.

BLACKBURN, J. (1936): "Acquisition of Skill: An Analysis of Learning Curves," IHRB Report No. 73.

BOLTON, G., AND A. OCKENFELS (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *The American Economic Review*, 90, 166–193.

CAMERER, C. (2001): "Behavioral Game Theory: Experiments on Strategic Interaction," In press.

CLARK, K., AND M. SEFTON (2001): "The Sequential Prisoner's Dilemma: Evidence on Reciprocation," *The Economic Journal*, 111, 51–68.

COOPER, R., D. DEJONG, R. FORSYTHE, AND T. ROSS (1996): "Cooperation Without Reputation: Experimental Evidence from Prisoner's Dilemma Games," *Games and Economic Behavior*, 12, 187–218.

ENGLE-WARNICK, J., AND R. SLONIM (2001): "Inferring Repeated Game Strategies from Actions:Evidence from Trust Game Experiments," Nuffield College Working Paper 2001-W13.

EREV, I., AND A. ROTH (1998): "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *The American Economic Review*, 88, 848–881.

FEHR, E., AND S. GAECHTER (2000): "Do Incentive Contracts Crowd Out Voluntary Cooperation?," Institute for Emperical Research in Economics University of Zurich Working Paper No. 34.

FEHR, E., G. KIRCHSTEIFER, AND A. REIDL (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics*, 108, 437–459.

FEHR, E., AND K. SCHMIDT (1999): "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 114, 817–868.

FERSHTMAN, C., AND U. GNEEZY (2001): "Discrimination in a Society: An Experimental Approach," *Quarterly Journal of Economics*, 116, 351–377.

FUDENBERG, D., AND E. MASKIN (1986): "The Folk Theorem of Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 125, 533–554.

FUKUYAMA, F. (1995): *Trust: Social Virtues and the Creation of Prosperity.* Free Press, New York, NY.

GLAESER, E., D. LAIBSON, J. SCHEINKMAN, AND C. SOUTTER (2000): "Measuring Trust," *The Quarterly Journal of Economics*, 125, 811–846.

GREIF, A. (1993): "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition," *The American Economic Review*, 85, 525–548.

JOHNSON, S. (1996): The Columbia World of Citations no. 31728.

KANDORI, M. (1992): "Social Norms and Community Enforcement," *Review of Economic Studies*, 59, 63–80.

KNACK, S., AND P. KEEFER (1997): "Does Social Capital Have an Economic Payoff?," *Quarterly Journal of Economics*, 112, 1251–1288.

KREPS, D. (1990): "Corporate Culture and Economic Theory," in *Perspectives on Positive Political Economy*, ed. by J. Alt, and K. Shepsle, pp. 90–143. Cambridge University Press, Cambridge.

MILGROM, P., D. NORTH, B. WEINGAST, AND R. BARRY (1990): "The Role of Institutions in the Revival of Trade: The Medieval Law Merchant, Private Judges, and the Champagne Fairs," *Economics and Politics*, 2, 1–23.

ORTMANN, A., J. FITZGERALD, AND C. BOENG (2000): "Trust, Reciprocity, and Social History: A Re–examination," *Experimental Economics*, 3, 81–100.

PUTNAM, R. (2000): *Bowling Alone: The Collapse and Revival of American Community.* Simon and Schuster, New York.

ROTH, A. (1995): "Bargaining Experiments," in *The Handbook of Experimental Economics*, ed. by J. Kagel, and A. Roth, pp. 253–348. Princeton University Press, Princeton, NJ.

SELTEN, R., AND R. STOEKER (1986): "End Behavior in Sequences of Finite Pirsoner's Dilemma Supergames," *Journal of Economic Behavior and Organization*, 65, 47–70.

THORNDIKE, E. (1928): "An Experimental Study of Representative Factors in Monkeys," *Journal of Comparative Psychology*, 8, 197–236.

VAN HUYCK, J., R. BATTALIO, AND M. WALTERS (2001): "Is Reputation a Substitute for Commitment in the Peasant–dictator Game?," Unpublished manuscript. Texas A&M University.

WILLIAMSON, O. (1985): *The Economic Institutions of Capitalism.* Free Press, New York, NY.

ZAK, P., AND S. KNACK (2001): "Trust and Growth," *The Economic Journal*, 111, 295–321.

Table 1: Supergame Lengths and Common Rounds Between Treatments

Supergame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Treatment																				
Long-Short	12	8	1	2	4	7	5	5	14	7	7	4	3	6	2	4	2	1	1	6
Short-Long	6	1	1	2	4	2	6	3	4	7	7	14	5	5	7	4	2	1	8	12
Common Rounds	6	1	1	2	4	2	5	3	4	7	7	4	3	5	2	4	2	1	1	6

Table 2a: Proportion of Send and Return Actions by Supergame Block: Long-short Treatment

	Frequency Trustors Play Send					Frequency Trustees Play Return				
Supergame Block	**1-2**	**3-5**	**6-10**	**11-15**	**16-20**	**1-2**	**3-5**	**6-10**	**11-15**	**16-20**
Session 2	0.551	0.551	0.571	0.565	0.633	0.519	0.593	0.655	0.783	0.694
Session 3	0.735	0.796	0.680	0.728	0.816	0.694	0.641	0.750	0.785	0.813
Session 5	0.857	0.800	0.657	0.667	0.614	0.633	0.750	0.739	0.743	0.721
Session 8	0.592	0.673	0.592	0.687	0.724	0.759	0.697	0.713	0.762	0.775
Average	0.684	0.705	0.625	0.662	0.697	0.651	0.670	0.714	0.768	0.751

Table 2b: Proportion of Send and Return Actions by Supergame Block: Short-long Treatment

	Frequency Trustors Play Send					Frequency Trustees Play Return				
Supergame Block	**1-2**	**3-5**	**6-10**	**11-15**	**16-20**	**1-2**	**3-5**	**6-10**	**11-15**	**16-20**
Session 4	0.837	0.918	0.918	0.952	0.857	0.878	0.911	0.888	0.907	0.905
Session 6	0.743	0.686	0.752	0.771	0.643	0.692	0.708	0.747	0.790	0.756
Session 7	0.657	0.657	0.819	0.800	0.686	0.783	0.783	0.919	0.917	0.771
Session 9	0.612	0.776	0.755	0.741	0.724	0.600	0.763	0.802	0.817	0.831
Average	0.712	0.759	0.811	0.816	0.728	0.738	0.791	0.839	0.858	0.816

Table 2c: Significance Results by Supergame Block Between Treatments

	Frequency Trustors Play Send					Frequency Trustees Play Return				
Supergame Block	**1-2**	**3-5**	**6-10**	**11-15**	**16-20**	**1-2**	**3-5**	**6-10**	**11-15**	**16-20**
p-values*	0.3429	0.4429	0.0143	0.0143	0.2429	0.2429	0.0286	0.0286	0.0143	0.1714

Notes: * Wilcoxian-Mann_Whitney One-sided Exact Test, Ho: Long-short = Short-long, Ha: Long-short > Short-long

Table 3: Trustee Average Pay Regressions

	Short-long	Long-Short
Constant	88.79**	70.91**
	(35.43)	(17.22)
Opponent Average Pay	-0.63**	-0.17*
	(11.51)	(-1.96)
Adjusted R-squared	0.85	0.16

Notes: t-statistics in parentheses

*estimates are statistically significant at the 10% level

**estimates are statistically significant at the 1% level

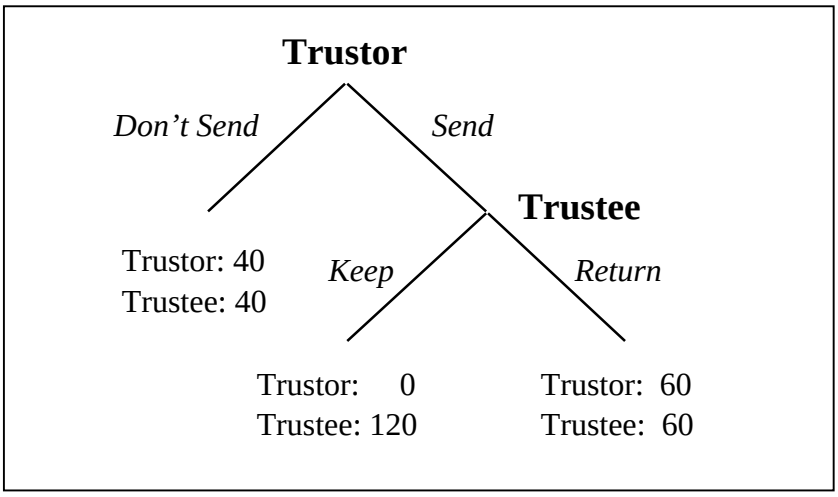

Figure 1: The Trust Stage Game

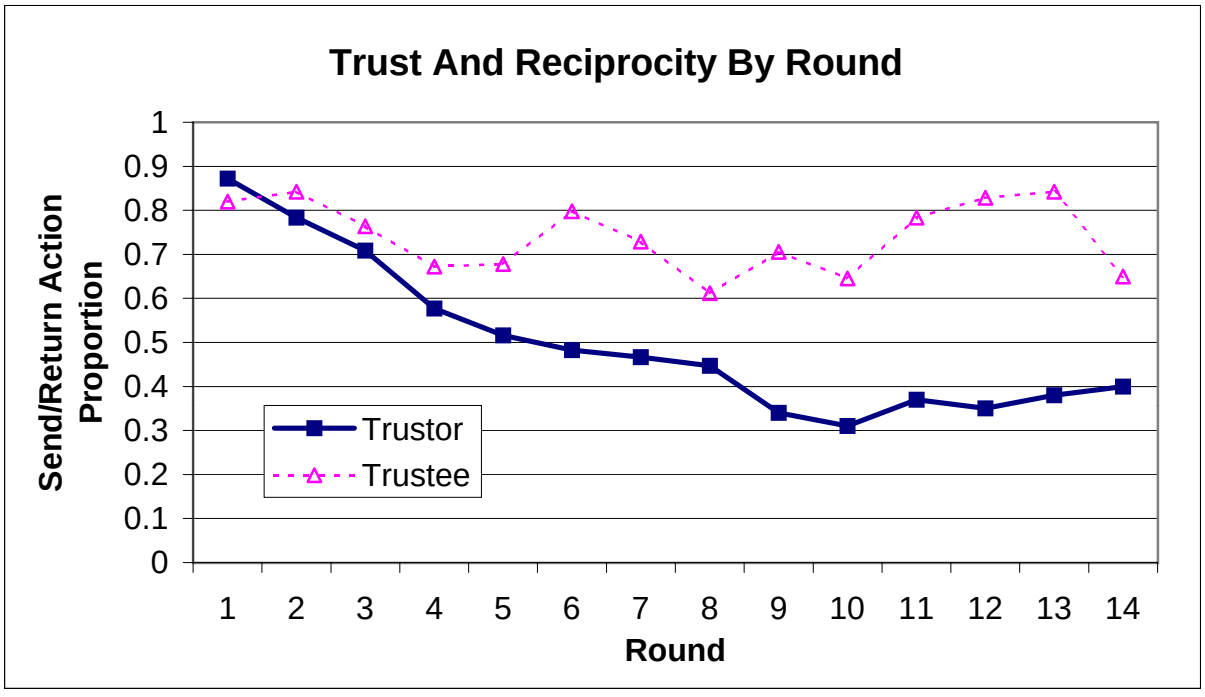

Figure 2: Proportion of *Send* and *Return* Actions by Round

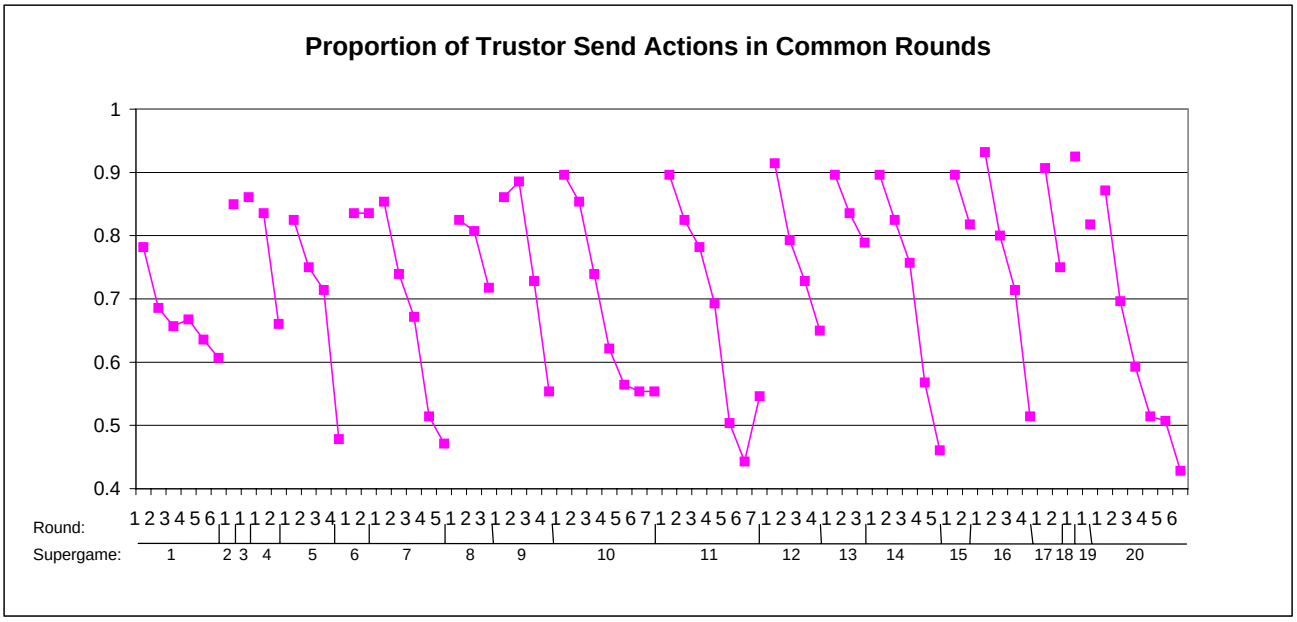

Figure 3: Proportion of Send Actions In Common Rounds By Supergame

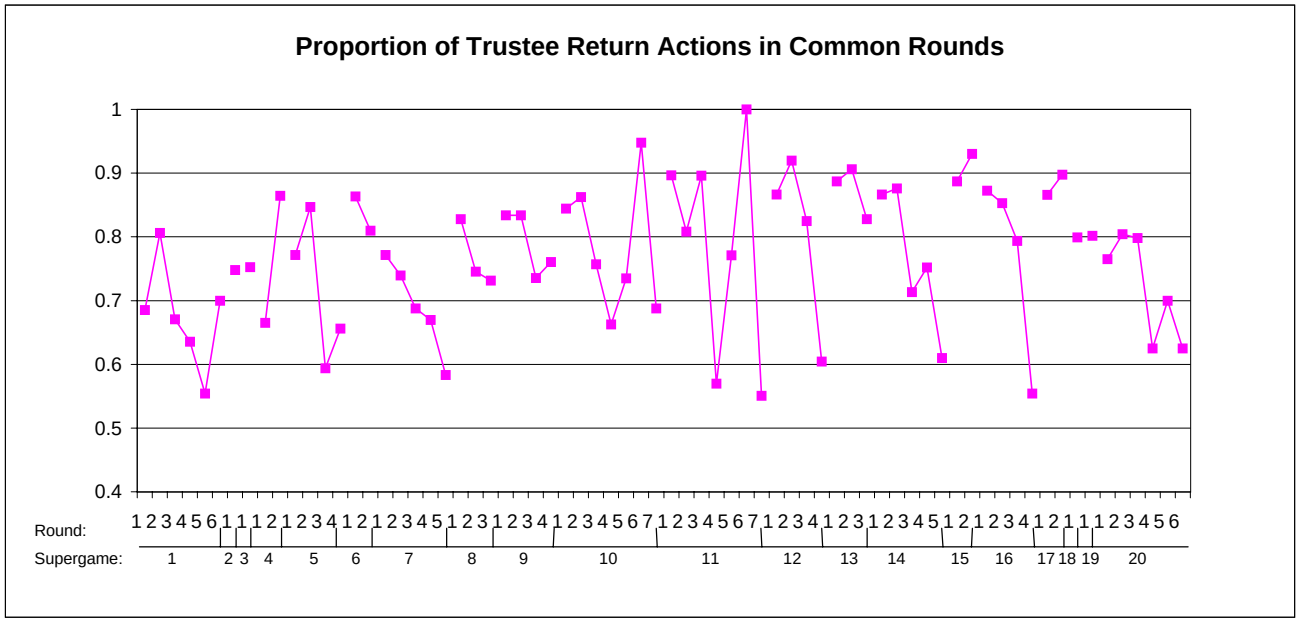

Figure 4 Proportion of Return Actions In Common Rounds By Supergame

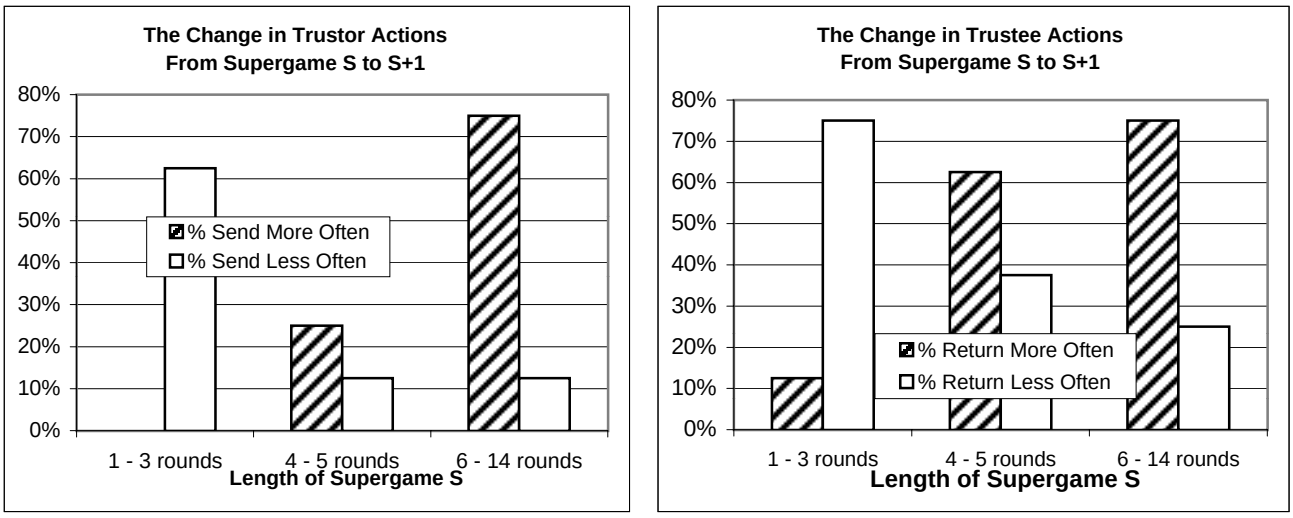

Figure 5: The Effect of Supergame Length on Round 1 Behavior

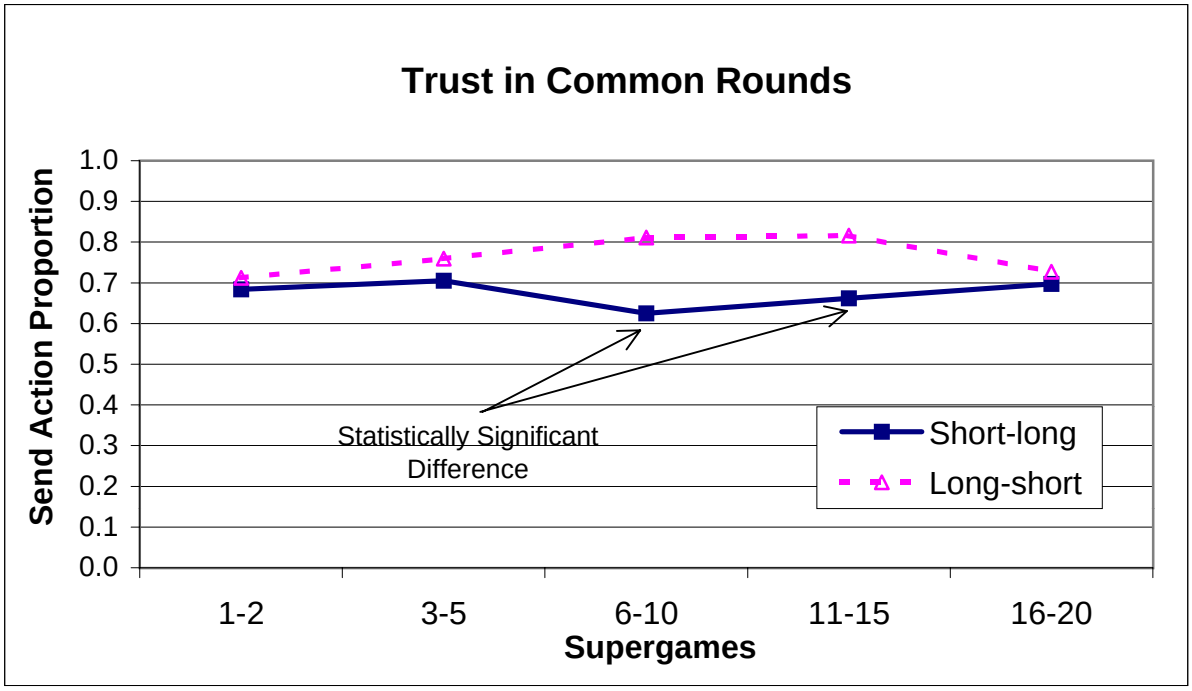

Figure 6: Proportion Send Actions By Supergame Block

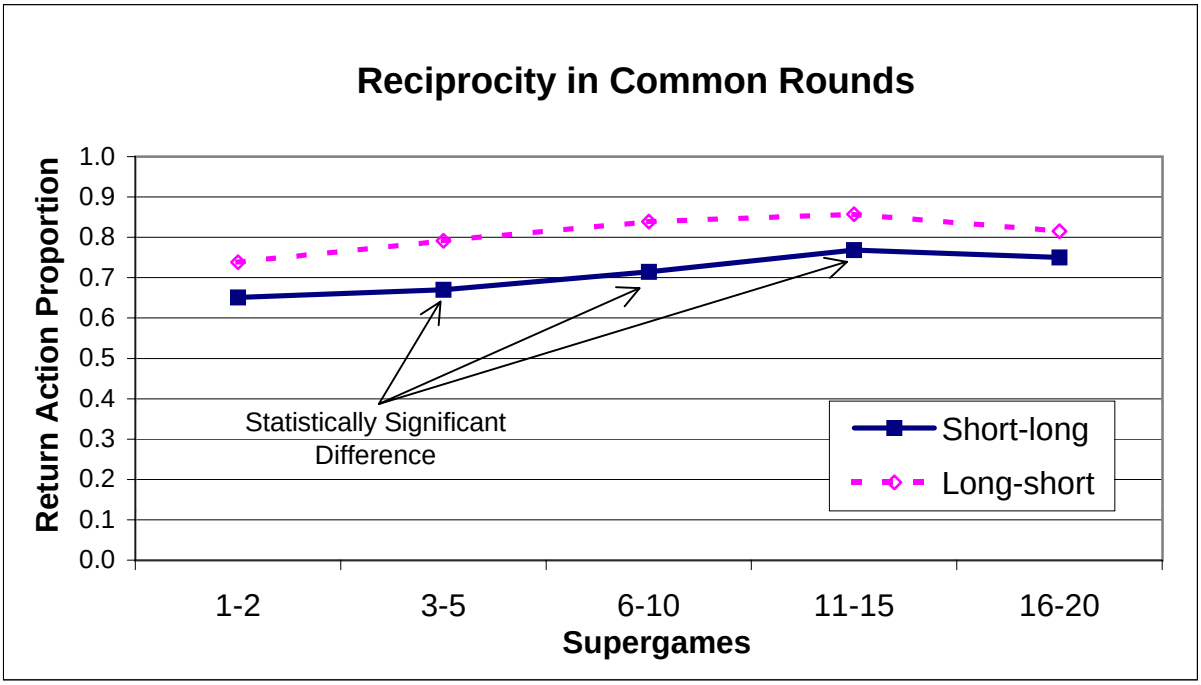

Figure 7: Proportion Return Actions By Supergame Block

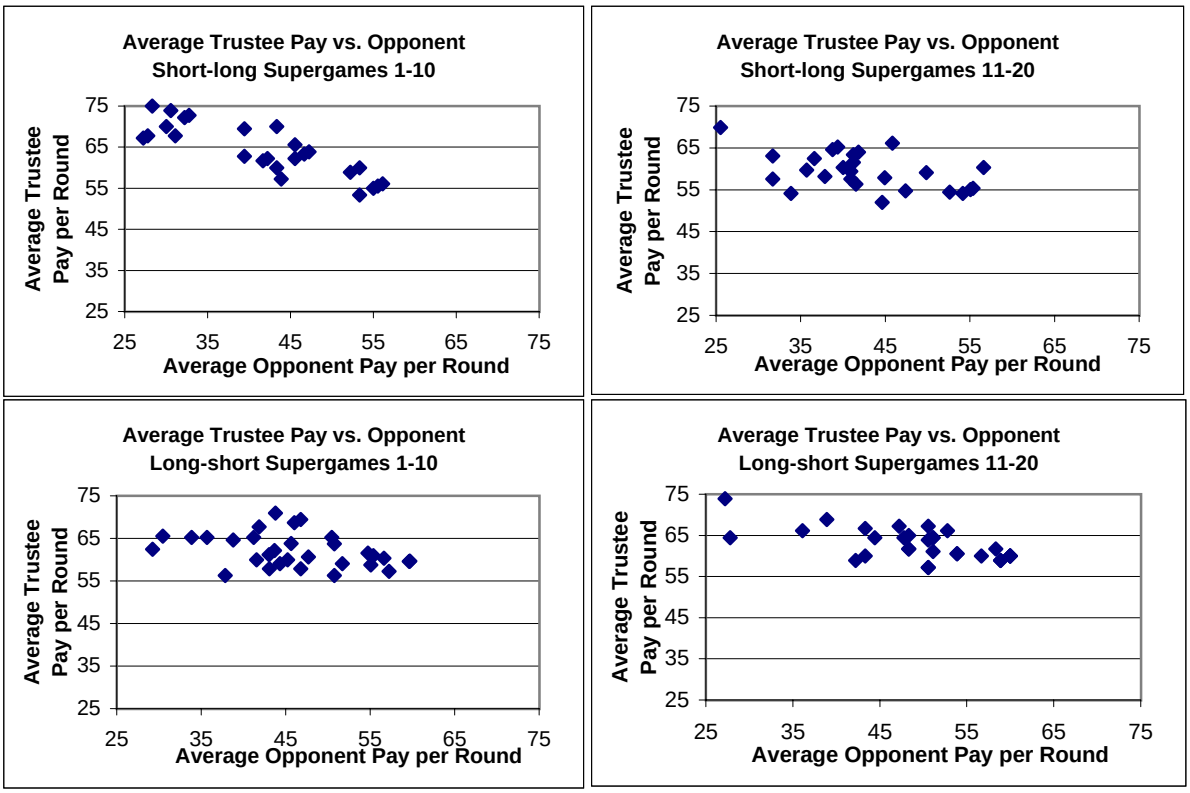

Figure 8: Average Trustee Pay as a Function of Average Opponent Pay