TRUST BUILDING AMONG STRANGERS*

Teck-Hua Ho
University of California, Berkeley
Berkeley, CA 94720

and

Keith Weigelt
University of Pennsylvania
Philadelphia, PA 19104

hoteck@haas.berkeley.edu
weigelt@wharton.upenn.edu

August, 2004

* We thank seminar participants at the ESA meeting and the University of Pennsylvania for helpful suggestions, Colin Camerer for his insightful comments on an earlier draft, and Christophe Van den Bulte for his suggestions and advice on data analysis. We are especially grateful to Karen Jehn for her early collaboration and ideas. This research is partially supported by NSF Grant SBR 9730187 and the Reginald Jones Center of Management at the Wharton School.

TRUST BUILDING AMONG STRANGERS

Abstract

The trust building process is basic to social science. We investigate it in a laboratory setting using a novel multi-stage trust game where social gains are achieved if players trust each other in each stage. And in each stage, players have an opportunity to appropriate these gains or be trustworthy by sharing them. Players are strangers because they do not know the identity of others and they will not play them again in the future. Thus there is no prospect of future interaction to induce trusting behavior. So, we study the trust building process where there is little scope for social relations and networks.

Standard game theory, which assumes all players are opportunistic, untrustworthy, and should have zero trust for others is used to construct a null hypothesis. We test whether people are trusting or trustworthy and examine how inferring the intentions of those who trust affects trustworthiness. We also investigate the effect of stake on trust, and study the evolution of trust.

Results show subjects exhibit some degree of trusting behavior though a majority of them are not trustworthy and claim the entire social gain. Players are more reluctant to trust in later stages than in earlier ones and are more trustworthy if they are certain of the trustee's intention. Surprisingly, subjects are more trusting and trustworthy when the stake size increases. Finally, we find the sub-population who invests in initiating the trust building process modifies its trusting behavior based on the relative fitness of trust.

1. INTRODUCTION

Graham Greene aptly captured the centrality of trust to social relations when he wrote, "It is impossible to go through life without trust: that is to be imprisoned in the worst cell of all, oneself" (1943). Trust is one of a handful of concepts to transcend territorial barriers of social disciplines; sociologists, anthropologists, economists, and psychologists study it. The social sciences view trust as an emergent process fundamental in social interaction and market efficiency (Creed & Miles, 1994; Etzioni, 1988). Without trust, opportunism rules and social exchange is costly. Despite the costs of trusting when others are untrustworthy, establishing trust can greatly benefit those involved (Axelrod, 1984). Some psychologists claim trust is the hallmark of social adjustment (Gurtman, 1992; Erickson, 1963) and without it neuroses prevail. Sociologists generally believe in producing trust, social relations and obligation are more important than morality, contracts, or institutions (Granovetter, 1985; Shapiro, 1987; Uzzi, 1996). Most economists view trust rationally and posit people only trust when it pays to do so (Camerer and Weigelt, 1988; Berg, Dickhaut, and McCabe, 1995).

Arguably, trust delivers its greatest social value as a linchpin to low-cost cooperative behavior. It holds immense strategic value for any size group or organization. Monitoring and incentives can motivate cooperative behavior, but at a higher cost. Organizational researchers theorize and empirically show trust reduces managerial monitoring, decreases coordination costs, and increases individual effort. (Ring & Van de Ven, 1992; McAllister, 1995; Chiles & McMackin, 1996, Uzzi, 1996). Trust often results in social gains, prompting some to suggest it leads to prosperity in society (Fukuyama, 1995).

Like religion, trust requires a leap of faith. As Bradach and Eccles (1989) note, [The faith to believe] "where opportunism might be rationally expected, trust prevails." Some social

biologists believe humans are genetically programmed not to trust or be trustworthy except for kin (Wright, 1995). This is echoed by a dominant assumption in game theoretic equilibrium analysis of self-interested and not trustworthy individuals (Myerson, 1991). Clearly, if people trust and others are not trustworthy, they exact a heavy toll from the trusting. Consequently, the existence of trusting and trustworthy behaviors is problematic.

Though problematic, we observe trusting and trustworthy behavior in others: A colleague who borrows a book and returns it. Or, a pedestrian crosses the street in the face of oncoming traffic because of a "stop sign". Neoclassical economists use institutional mechanisms (Alchian and Demsetz, 1973; Williamson, 1975; Rosen, 1982) to help explain trusting and trustworthy behavior by self-interested people within the classical framework of utility maximization. They rationalize that mechanisms such as explicit contracts, incentive schemes, and hierarchical authority can induce individual behavior consistent with trust and trustworthiness. In this world, trust is used not to explain behavior, but to label it (Craswell, 1993).[1] To trust or not is purely calculative; institutional mechanisms help individuals act as if they do trust, usually because these mechanisms increase the costs of opportunistic behavior. As Granovetter (1985) aptly notes, such mechanisms do not produce trust but are functional substitutes for it.

On the other hand, Granovetter (1985) posits that concrete personal relations are an important determinant of trusting behavior. This so-called "embeddedness" argument emphasizes the role of social networks and reputation arising from interpersonal contacts. The higher the embeddedness of the social relation, the more likely (on average), trust is extended, and returned. Granovetter suggests embeddedness is a necessary (though not sufficient) condition for generating trusting behavior. General population-level reputation is significantly

[1] Malhotra and Murnighan (2000) report an interesting study to show that contracts make development of trust difficult or block preexisting trust.

less important than individualized reputations generated by direct prior contacts, because direct contacts provide more reliable information of others' intentions. Granovetter states (p. 506) the inclusion of such relations in the trust decision does not call for the abandonment of a rational decision-making process. Instead, the degree of embeddedness should be factored into such models, for example, as with a matching protocol (e.g., Ellison, 1993).

In this paper, we create an experimental setting where there is minimal scope for social embeddedness. In our experiments, subjects interact anonymously and only play each other once. Contracting is not possible, and in fact, subjects never find out with whom they play the trust game.[2] The novel game has additional desirable features. It allows us to untangle trusting and trustworthy behavior; it has multiple stages, which allow subjects to detect the intention of others; and mutual trusting behavior leads to sizable social gains. Delineating trusting from trustworthy behavior is important because they are fundamentally different, and most previous research on Prisoners' dilemma entangles the two. The ability to detect intention of others allows us to study an important construct: "pseudo trust", which has not been investigated much in prior studies. Pseudo trust occurs when one trusts and gives up property rights in early stages of the game in order to regain larger ones in the later stages. This pseudo trust is to be differentiated from the trusting behavior in later stages where there is no chance to regain property rights.

This paper aims to answer five basic research questions. First, can anonymous interactions (e.g., like exchanges between participants of e-bay auctions) generate trust and trustworthy behavior? Our experimental design allows us to quantify the extent to which these behaviors exist in our population. Second, how does the potential of future social gains influence current trusting behavior? This potential varies across the stages of our trust game. Third, does

[2] We use student subjects. It is possible that some students may know each other prior to the experiment. Thus, they are strictly not strangers. We define strangers here to mean people who engage in anonymous and one-shot

better knowledge of others' trusting behavior make people more trustworthy? Within our game, the uncertainty of intent of trusting players is high in early stages and low in later stages. Fourth, is trusting behavior dependent on stake size? In three sessions, we increase the stake size tenfold to see whether this increase affects trusting behavior. The dramatic increase in stake size allows us to remove any concerns on whether subjects are sufficiently motivated and to test whether stake size increases the extent of rational behavior. Fifth, do evolutionary paths of trust building differ across sub-populations? Specifically, do the trust dynamics of the sub-population who invests and initiates the trust building process differ from the one with the first decision rights to terminate the process? We answer the above five questions by collecting data in three different countries: US, Singapore, and China. This cross-cultural data collection effort helps to check for robustness of the basic results and increase external validity.

The paper is organized as follows. In Section 2, we describe our trust building game. Section 3 develops hypotheses using theories of trusting and trustworthy behavior from sociology, economics, and psychology. Experimental design is presented in Section 4 and results in Section 5. Section 6 discusses the methodological issues, explores theoretical implications of results, and suggests future research directions.

2. A TRUST BUILDING GAME

Our trust building game is shown in FIGURE 1.[3] The game consists of 4 decision stages, and is played between two players; one designated as "RED", the other as "BLUE". Akin to the

interaction.

[3] While our trust building game looks like the centipede game (see Rosenthal, 1982; McKelvey and Palfrey, 1992), there is an important difference. In our game, the player can choose to be either trustworthy or not while in the centipede game, players are forced to be not trustworthy.

real world, individual property rights are voluntarily given up in exchange for potential social gain.

RED moves first. She can either "trust" BLUE and pass, or choose not to trust and select down. If down is chosen, RED exercises her property rights and unilaterally decides how to split the payoff of 4 (rewarding herself with 4a ($0 \leq a \leq 1$), while BLUE receives 4(1-a)), and the game is over. If RED trusts, she passes, thereby giving up her property rights in hopes of sharing future social gains. Social gains occur because the payoff doubles, though BLUE now owns the property rights. That is, BLUE must now make a similar decision of whether to trust or not. If BLUE trusts in decision stage 2, he passes. If BLUE does not trust, he takes and splits the payoff so he receives 8 b ($0 \leq b \leq 1$) and RED receives a payoff of 8 (1-b). Similar to Glaeser et al. (2000), we measure BLUE's level of trustworthiness by the amount returned to RED. The higher the returned amount, the more trustworthy is BLUE. The game is over if BLUE does not trust. If BLUE trusts, and passes, we again double social gain to 16 in stage 3.

Then RED again must decide to trust or not. If RED does not trust, she takes, and must decide whether to be trustworthy. Again, RED's level of trustworthiness is measured by the amount she returns to BLUE. If RED trusts, she passes to BLUE and the payoff doubles to 32. In Stage 4, BLUE can only choose his level of trustworthiness.

In sum, we measure the degree of trusting behavior by the proportion of "passes" in stages 1-3. We denote these proportions by x, y, and z respectively. The degree of trustworthiness at each stage is represented by the proportions a, b, c, and d respectively.[4]

[4] We implicitly assume the measure of trustworthiness is stage–independent. That is, b=0.6 and c = 0.6 represent equivalent trustworthiness behaviors in stages 2 and 3. An alternative view could allow trustworthiness measures to be stage-dependent. For example, RED in stage 3 may feel she is partly responsible for the larger payoff of 16 by passing to BLUE in stage 1 and therefore is entitled to a larger proportion of the payoff. Hence, the level of trustworthiness in stage 2 and 3 could be equivalent even if c > b. This latter view suggests that players will take more for themselves in later stages for the same degree of trustworthiness. Thus, our proportions c and d provide a lower bound for the degree of trustworthiness in stage 3 and 4. We find no evidence to support a stage-dependent

The game has several desirable features for examining the trust-building process. Unlike some previous trust games (e.g. Bolle, 1995, Yamagishi, Cook, and Watabe, 1998), our game allows an individual to exhibit both trusting and trustworthy behavior. Subject decisions in each stage consist of two components – strategy choice (to trust or not)[5] and payoff choice (to be trustworthy or not). Strategy choice determines whether a subject trusts (pass) or not (take). Payoff choice determines whether the subject is trustworthy.[6] Hence, one can be trusting but not trustworthy.

Second, our trust game allows subjects to reveal their trustworthiness. Previous games force subjects to choose between two payoff choices, which might not coincide with their most preferred choice (for example in McKelvey and Palfrey 1992, subjects are forced to claim 80% of the property rights. The same thing can be said about most studies of repeated prisoner's dilemma). So subjects may choose the better of two evils. Because subjects in our game can divide the social gain in any way, they choose their most preferred level of trustworthiness.

With the multi-stage design, we can study how future social gains might influence the trust building process. For example, if higher future social gains induce more trusting behavior, we will see more passes in earlier decision stages. The design also allows us to test process-based trust, a widely identified trust process in sociological work (Zucker, 1986), while controlling for institutional and person-based trust.

Our game shares similarities with other trust games – it is a two-person game and trusting

---

view because players do not seem sensitive to history within a game round (to be discussed in Section 5).

[5] Mayer, Davis, and Schoorman (1995) suggest one can cooperate but not trust, for the use of external mechanisms can compel others to cooperate. For example, a team leader can punish those who won't cooperate. Members may not trust each other, but they prefer cooperation to the punishment for non-cooperation. No external mechanisms or authorities enforce punishment in our design. To trust or not is solely an individual's decision.

[6] Strictly speaking, RED's payoff choice in the first stage is not an indication of her trustworthy behavior because she is endowed with the payoff, and it is not a consequence of social gain from trust. Also, BLUE's strategy choice in stage 4 is not an indication of trust because property rights are last exchanged in stage 3. We discuss this further in Section 3.2. Because RED knows BLUE will not give back control of property rights, RED is clearly hoping to

results in social gains. It also differs from other games in several important ways. For example, our trust game is different from a prisoner's dilemma (PD) game with an option to exit (PDO). In our game, "trusting" at each stage means giving up the property rights – the truster has no control over the payoff realization if the trustee chooses to end the game at the next stage. In the PDO game, choosing to play does not mean the trustee has no influence over the expected payoffs. A player who chooses to play the game or "trust" can still guarantee herself a nonnegative payoff by choosing defection (as in Orbell and Dawes 1991; 1993). Second, our game gives those who are trustworthy a complete freedom to divide the property rights as he wishes. The outcomes of the PDO game are determined by the experimenter, not subjects[7].

Similarly, there are other trust games closely related to ours (e.g., Berg, Dickhaut, and McCabe, 1995, Bolle, 1995; Glaeser, et al., 2000). These dictator-type trust games also give trustees complete control in division of social welfare. However they are single-staged. The trustee (the dictator) is never asked to relinquish control once given to her. Consequently, they can only study initial stages of the trust building process.

3. HYPOTHESES

3.1 RATIONALITY AND SOCIAL UNCERTAINTY

A standard assumption in game-theoretic equilibrium analysis is that individuals are rational (they are self-interested and monetary maximizing). Consequently, they cannot be entrusted with property rights if there is no chance for penalizing untrustworthy behavior (either through

get a "fair split" of the social gains. She must believe that BLUE is trustworthy.

[7] This "selective play" paradigm literature shows that social welfare increases when subjects have a freedom to choose between playing or not playing a particular Prisoner's Dilemma game. These increases occur because intending cooperators are more likely to enter such games relative to intending defectors, resulting in a better chance of cooperative outcomes. Consequently, the welfare of intending cooperators is higher than that of intending defectors (Orbell, Schwartz-Shea and Simmons, 1984; Orbell and Dawes, 1991; 1993; Morikawa, Orbell and Runde, 1995; Hayashi and Yamagishi, 1998).

formal institutional mechanisms or informal sanctions of social relations). This game-theoretic reasoning is as follows: If stage 4 is reached in our game, a rational BLUE player will keep the property entrusted to him for himself and take 32 (i.e., $d = 1$). A rational RED player, knowing this, will neither trust BLUE nor be trustworthy. Hence, she will take the entire payoff of 16 in stage 3. Anticipating this action of RED in stage 3, a rational BLUE player exhibits this behavior in Stage 2, takes 8, and leaves nothing for RED. Completing this backward induction argument, the rational RED player takes 4 in the first stage. Hence all players are neither trusting nor trustworthy.

Sociologists also believe little trust will develop where social uncertainty is high, there is little commitment to relations, and social embeddedness is insignificant. Kollock (1994) and Yamagishi, Cook and Watabe (1998) indicate social uncertainty exists when 1) the trustee has an incentive to impose harm or cost on the truster 2) the truster is uncertain about whether the trustee will actually impose this harm or cost. Our design imposes high levels of social uncertainty on subjects by decreasing the probability of committed relations. Committed relations reduce social uncertainty in two ways: 1) committed partners accumulate information about each other (i.e., mutual monitoring) and 2) reciprocity is possible (i.e., mutual hostages or Tit-For-Tat) (Yamagishi and Yamagishi, 1994). Our experimental design makes any degree of commitment in a relation almost impossible. In fact, it provides an estimate of the degree of trusting behavior in a population of strangers where social embeddedness is insignificant. Subjects play against each other at most once (i.e., they are strangers), each subject only assumes the role of either a RED or BLUE player, RED and BLUE players are in different rooms, and players are not told with whom they are matched. Consequently, we expect subjects to have a high level of social uncertainty. In our experiments, there is neither a way to reveal the identity

nor a disincentive to penalize untrustworthy individuals. The act of being untrustworthy is known only to the untrustworthy subject and the one whose trust was betrayed. And, the two will never meet again in the future. So there are no formal or informal mechanisms available to punish an untrustworthy player. The lack of monitoring and penalty increases the probability of opportunism. Consequently, we expect less trustworthy behavior.

The standard theories in economics and sociology lead to the predictions that people will exhibit neither trusting nor trustworthy behavior. Consequently, we have following hypotheses:

Hypothesis 1a: Proportions b = c =d =1. That is, individuals are not trustworthy (i.e., they are opportunistic).

Hypothesis 1b: Proportions x = y = z = 0. That is, individuals exhibit zero trust; they do not give up property rights.

Since there is an increasing evidence to suggest that humans are boundedly rational and care about fairness (see Camerer, 2003 for an excellent review), the above hypotheses serve as our baseline hypotheses. They allow us to compare our results with those reported in the literature and quantify the degree of deviation from the standard rational hypotheses.

3.2 INFORMED RECIPROCITY

There is a subtle difference between trusting behaviors in the first two stages and the third. A BLUE player who receives the property rights from a RED player in stage 3 can infer unambiguously the intention of RED. By passing, RED clearly indicates her belief of the BLUE's trustworthiness, for the only decision facing BLUE is one of trustworthiness. The same

thing cannot be said for trusting behavior in stage 1 or 2. We cannot unambiguously infer players who relinquish their property rights in these earlier stages believe others are trustworthy. Players, who give up their property rights in stages 1 and 2, have the opportunity to regain future larger property rights. Consequently, there is greater social uncertainty in the earlier stages for the trustee. Do subjects give up property rights because they believe others are trustworthy, or for cultivating pseudo trust? Pseudo trust occurs when one trusts to increase property rights for the possibility of claiming them in the later stages. Because players are better informed about intentions of trusting subjects in stage 3 relative to stages 1 and 2, we expect a greater level of trustworthiness in stage 4. Since BLUE can unambiguously detect and interpret the intention of RED in stage 3, he can selectively be more trustworthy to those who trust not for own but for social gain. This informed reciprocity is characterized as "translucent" (Gauthier, 1986; Orbell and Dawes, 1991), which is the degree to which one can recognize another's intention. Consequently, individuals who are more certain of the intentions of others' trusting behaviors are more trustworthy. That is,

Hypothesis 2: Proportion $d < (b + c)/2$.

3.3 THE POTENTIAL VALUE OF TRUST

In each stage of the trust building game, those who trust must give up property rights in exchange for potential social gain. We define trust's potential value as the difference between the expected future return from trusting and the property rights currently held. The expected future return at any stage is computed given the percentages of passing and taking and the mean percentage of property rights claimed in future stages. Necessarily, the expected return for earlier stages is higher than for later ones. Hence, we expect to see less trusting behavior in later

stages of the game than earlier ones. For example, a RED player in stage 1 can take and receive a payoff of 4, but is losing the opportunity to share a potential future payoff of 32. In stage 3, this same player can take and realize a gain of 16, versus the opportunity to share a future payoff of 32. Consequently, individuals are less trusting as the potential value of trust decreases.

Hypothesis 3: Proportion $x > y > z$.

3.4 HIGH-STAKE TRUST BUILDING

While relatively little research shows how the magnitude of payoffs affects trusting behavior – though stories of soldiers who sacrifice themselves to save comrades are as old as war itself (Holmes, 1985) – the effect of financial incentives on experimental subject behavior has attracted substantial attention among social scientists. For some reviews see Smith and Walker (1993), Bonner et al., (1996), Hertwig and Ortmann (1998), Jenkins, et al. (1998) and Camerer and Hogarth (1999). While these authors examine hundreds of experiments, no definitive answer arises from them, though there is consensus on general trends. The majority of reviews find increased financial incentives have no significant impact on mean subject behavior, though they do reduce the variance across choices (Smith and Walker, 1993; Bonner et al., 1996; Camerer and Hogarth, 1999). Hertwig and Ortmann (1998) find incentives slightly affect behavior (though most often when increased from zero). Jenkins et al. (1998) find the strongest link between incentives and performance, though most of the studies they review are one-person output tasks (e.g., coding) that measure performance on a purely quantitative scale.

Another consensual finding is that incentives matter more in non-strategic than in strategic tasks. A strategic task is one where a subject's payoff depends on her actions and those of others. Non-strategic ones examine one-person decisions like the recalling of items,

predicting future states, or output per time period. For example, Camerer and Hogarth (1999), report the following. They categorize fifty-eight studies as whether an increase in financial incentives did or did not significantly change mean performance. In thirty-one (53.4%) studies, they did. Of these thirty-one studies, seven (22.5%) were strategic experiments and twenty-four were non-strategic. In contrast, of the twenty-seven studies where mean performance did not change, twenty (74.1%) were strategic. Furthermore, if we categorize the 34 strategic studies into bargaining [trust building] or non-bargaining (e.g., signaling, auctions, tournaments) games, all 9 bargaining games fall into the category of increased incentives having no effect on mean performance. Camerer and Hogarth (1999) state, "… results in Table 1 [summarizing the above results we discuss] suggest little reason to think the effects of very large incentives will be substantial." They conclude this finding is applicable to strategic decisions especially in bargaining and auctions. Consequently, we hypothesize the level of trust is not dependent on stake size.

Hypothesis 4: The level of trust does not depend on the stake size. Proportions x, y, and z are not functions of stake size.

3.5 POPULATION LEARNING AND THE EMERGENCE OF TRUST

The logic behind hypothesis 1 is grounded in full rationality of players. Recent research in economics increasingly suggests players are boundedly rational with adaptive expectations (Simon, 1957, Fudenberg and Levine 1998, Camerer and Ho, 1999; Camerer, Ho, and Chong, 2002). These models assume subjects' actions are path dependent and self-correcting. Subjects stick to strategies, which worked well in the past, or experiment with previously unchosen strategies likely to improve performance. Alternatively, one can view trust building in an evolutionary perspective. For instance, Hoffman, McCabe, and Smith (1998) suggest that their subjects exhibited reciprocity even in single-shot games (counter to rational hypothesis) because

humans have evolved mental algorithms for identifying and punishing cheaters in social exchange. In evolutionary dynamics terms, subjects adjust their behavior by mimicking the person that is the most successful or "fittest." Thus, future trusting behavior depends on its current success or relative fitness. The higher the number of trusting and trustworthy individuals in the population, the more likely trusting and trustworthy behavior increases both the payoffs of the collective and individuals. We expect trusting behavior in the population increases as its expected value (or fitness) increases. While our focus is on the evolution dynamics of group behavior, other studies suggest group dynamics can be derived from individual learning models in which subjects modify their propensity of choosing a strategy based on its performance (see for example Friedman, 1991; Borgers and Sarin, 1997).

Hypothesis 5: The emergence of trust in a population is governed by evolution dynamics. There is more trust in the population if trust has a higher relative fitness.

4. EXPERIMENTAL METHOD

A total of 386 subjects participated in the experiment. Subjects were recruited from undergraduate courses at National University of Singapore in Singapore (n=106), University of Pennsylvania in the United States (n=100), and from Beijing University in the People's Republic of China (n=180).

A typical experimental session consisted of twenty subjects who met in a room at a specified time. As they entered the room, they chose a bingo ball from a cage containing 20 balls numbered 1-20 (a ball chosen was not replaced).[8] The number on the ball represented his subject number. Then the instructions (Appendix I) were publicly read. After questions,

[8] Two sessions in Singapore used 52 and 54 subjects and required a larger number of bingo balls.

"BLUE" subjects (those with odd subject numbers) were asked to go to another room. "RED" subjects remained in the room. Subjects in each room were positioned so no subject could see the worksheet of another.

Our multi-stage trust building game consisted of four stages. RED subjects made choices in Stage 1 and 3 and BLUE subjects in Stage 2 and 4. Although the game could end in any stage (when subjects chose "take"), subjects recorded their choices in both stages. We asked subjects to record their contingent strategy for two reasons. First, we reduced the data collection time and hence the potential for boredom. Second, contingent strategies help us understand subject behavior in all stages even though the actual game might end earlier.[9]

Each session consisted of ten rounds and lasted for about one hour. In each round a BLUE subject was paired with a RED subject (this was her pair member). Subject choices were privately made and recorded by an administrator in the room. After all subjects in both rooms made their decisions, the two administrators met outside the rooms and recorded all choices. They then went back to their respective rooms, privately informed each subject of the choices of the pair member and the generated payoff. The administrator then announced the end of that round and the beginning of the next round. Each BLUE subject was paired with a RED subject only once, so each subject had a new partner in each round. This procedure of matching subjects rules out any possibility of a shadow of the future across rounds.[10] A RED player clearly cannot influence the action of another RED player because they never interact with each other. Also, RED cannot influence a BLUE player because she only plays with BLUE once and BLUE players never play against each other. A similar line of reasoning applies to BLUE players. This

[9] Consequently, in Figure 2, the total choice count in stage 3 is 1178 and in stage 4 is 662.

[10] This matching protocol rules out even the triadic effects. That is, player A may want to be nice to player B because player B will be playing player C whom player A will play in the future. Clearly this is not possible in our

matching method however does not rule out a shadow of the future across stages within a game round.[11] After 10 rounds subjects add up their points from each round, and multiply this by a conversion rate which transforms the points into a monetary payoff. The subjects were explicitly told they would be paid in cash after the experiment, so they understood earnings were a function of decision choices.[12]

Subjects seemed to understand the instructions and the payoff associated with each of their actions. This was evidenced in two ways. First, after the experiment and during the debriefing, subjects' comments suggested that they did not have trouble writing down contingent strategies (e.g., "Many of my choices in stage III or IV did not matter.") and the matching protocol (e.g., "I could not build relationship with a person because we did not match with the same person more than once.") Second, almost 100% of the choices in Stage IV were Take (see Figure 2), which is a dominant choice. This result indirectly suggests subjects did understand the game.

We conducted over 16 experimental sessions, five in the US, nine in China, and two large sessions in Singapore to check the robustness of our findings. Six China sessions were low-stake, 3 were high-stake.[13] High-stake sessions had a tenfold increase in the conversion rate. Because we collected data in different countries, we took the following precautionary controls. Instructions in a country were written in its primary language – English in Singapore and the US and Chinese in China. We used the reverse translation method to ensure Chinese instructions

---

design because players of the same color never play each other.

[11] We believe our simultaneous social exchange protocol deters subjects from perceiving a shadow of the future within a game round because subjects must submit a full contingent strategy without knowing the strategies of others. As discussed in Section 5, we test this belief by running additional experiments using a sequential social exchange protocol. In the sequential protocol, subjects choose after knowing the strategies of others in prior stages within a game round. Since the results from the two exchange protocols are virtually identical, we conclude that subjects did not perceive a shadow of the future within a game round.

[12] As suggested by Kee and Knox (1970), such meaningful incentives allow subjects to clearly recognize the dangers and rewards of extending trust.

were equivalent to those in English. That is, an independent translator translated the instructions from English into Chinese and another independent translator from Chinese back into English. Then, one of the authors who is fluent in both Chinese and English checked both translations. The conversion rate was chosen to normalize average payoff so that subjects were paid about twice the minimal wage for the country. On average, US subjects earned US$11, Singapore subjects earned about S$10, and Chinese subjects about 24 and 320 Yuan Renminbi in the low-stake and high-stake games respectively. This latter earning is approximately equivalent to two-month wages for a blue-collar worker.

5. RESULTS

Altogether 386 subjects participated in the experiment (193 RED and 193 BLUE subjects). Each subject made choices over 10 rounds. Figure 2 summarizes their choices in each of the four decision stages. As you can see, trusting behavior monotonically declines across decision stages while the degree of trustworthiness monotonically increases. Recall we examine trusting behavior in stages 1-3 and trustworthiness in stages 2-4. In each stage, there is a substantial decline in the proportion of trusting choices. On the other hand, the degree of trustworthiness increases marginally for stages 2-3 with a substantial increase in stage 4.

Seventy-two percent of the RED choices initiate trusting building in stage 1. Correspondingly, 41% of the BLUE's choices and 20% of the RED's choices are of the trusting nature in stage 2 and 3 respectively. In stages 2 and 3, those who did not trust were also not trustworthy; they kept 95% and 94% of the property rights. Interestingly, in stage 4 only 78% of the property rights were claimed.[14] We formally test our hypotheses below.

[13] Our research budget did not allow us to conduct high-stake sessions in Singapore and US.
[14] Obviously, these percentages are generated only those who trusted in stage 2.

We test whether subjects are trustworthy by examining the percentages of property rights claimed by BLUE players in stage 2. We focus on the trustworthiness of BLUE subjects who choose not to trust and immediately terminate the trusting building process. A more trustworthy BLUE should return a higher entrusted property rights to RED. A fully opportunistic BLUE takes 100% of the property rights. Table 1 reports the average proportion of property rights claimed along with the 95% confidence interval. It clearly shows BLUE subjects are largely opportunistic in stage 2. The average proportions claimed exceed 95% in all three countries. Overall, these results support hypothesis 1a.

In spite of this lack of trustworthy behavior, we find some support for trusting behavior in Table 1. We test hypothesis 1b by the percentage of RED subjects who initiate the trust building process by choosing pass in stage 1. About 70% of RED subjects trust by giving up their property rights in stage 1. For example, 73% of the Singaporean RED subjects chose pass in the first stage. Clearly, a significant percentage of subjects initiate the trust building process. Thus, we find no support for hypothesis 1b.

Hypothesis 2 examines the effect of behavioral translucency on trustworthy behavior. It states trustworthiness is more likely when the intention of the trustee is less ambiguous. We test this hypothesis by comparing trustworthy behaviors in stage 4 relative to earlier stages. A BLUE subject in stage 4 is certain about the trusting intention of RED because there is no opportunity for RED to regain property rights. The same cannot be said for earlier stages. For example, a BLUE subject in stage 2 cannot be sure why a RED subject passes her property rights in stage 1. She may pass selfishly so as to increase property rights for the possibility of claiming them in stage 3 or she may pass for the collective social gain. In stage 4, there is no such uncertainty.

Table 2a shows the average proportions claimed are 95.85%, 93.96%, and 77.4% in

stages 2, 3 and 4 respectively. Across the three countries, the 95% confidence intervals for stage 4 do not overlap with those in stages 2 or 3. In fact, in the first 5 rounds, the modal proportion claimed in stage 4 is 50% (131 out of the 354 subjects who take in stage 4)[15]. Clearly, subjects are more trustworthy in stage 4. Thus we find some support for hypothesis 2.

Hypothesis 3 states trust increases with the difference in the expected return from trusting and the size of property rights currently held. Since this difference decreases over stages, we expect less trusting behavior in latter stages. Table 2b clearly supports this hypothesis. There is a significant decrease in trusting behavior from stage 1 to stage 2, and from stage 2 to stage 3. This is true across all countries. The average percentages of subjects trusting are 72%, 40% and 20% in stages 1, 2 and 3 respectively.

Hypothesis 4 predicts that stake size will not affect trusting behavior. Figure 4 shows subject behavior in the three high-stake trust games. Comparing Figures 2 and 4, it is obvious that subjects exhibit significant more trust in the high-stake games. For example, in stage 3, RED subjects pass 80.3% as compared to 20.4% in low-stake games (a simple proportion test indicates this difference is highly significant ($p < 0.001$)). This higher level of trust in stage 3 is also matched by a higher level of trustworthiness in stage 4. For example, 167 out of 270 choices in the high-stake games were $d = 0.5$ whereas only 197 out of 660 choices in the low-stake games were $d = 0.5$ (i.e., divide the property rights evenly).

We examine hypothesis 5 by testing whether changes in subjects' behaviors are consistent with evolution dynamics. For each RED player in every round, there are three possible trusting behaviors: 1) take in stage 1 (T), 2) pass in stage 1 and take in stage 3 (PT), or 3) pass in stage 1 and pass in stage 3 (PP). The fitness (or payoff) for T is 4. The fitness for PT

[15] The number of subjects who chose a 50:50 split in stage 1,2, 3, and 4 across all 10 rounds were 11 (out of 453), 23 (out of 968), 38 (out of 938), and 197 (out of 660) respectively.

is the sum of two terms: 1) the percentage of "Take" in stage 2 times 8 (1-b) and 2) the percentage of "Pass" in stage 2 times 16. For example, in Figure 3, the expected potential payoff for PT in round 1 is 0.547*(1- 0.8838)*8 + 0.453*16 = 7.76. Similarly, the fitness for PP is a sum of two terms: 1) the percentage of "Take" in stage 2 times 8 (1-b), and 2) the percentage of "Pass" in stage 2 times the percentage of "Take" in stage 4 times 32(1-d). In Figure 3, this becomes 0.547*8*(1-0.8838) + 0.453*1.00*32*(1-0.6767) = 5.20. These fitness measures are given in Table 3. Hypothesis 5 suggests the expected use of a strategy in the next round depends on its use in the current round and increases with its current fitness relative to the fitness of other strategies. Mathematically, we model the proportion of strategy j in round t to be:

$$P_j(t) = \alpha P_j(t-1) + (1-\alpha)\frac{e^{\beta EV_j(t-1)}}{\sum_K e^{\beta EV_k(t-1)}}$$

The model suggests the proportion of subjects choosing strategy j in a round is a function of the proportion of subjects choosing the same strategy in the previous round and the fitness of strategy j in the previous round relative to other strategies. $\alpha > 0$ implies subjects show a tendency to repeat their previous choice regardless of the fitness from their previously chosen strategies. This could be inertia or due to insufficient experimentation. $\beta > 0$ suggests subjects are more likely to pick a strategy j if they receive a higher fitness from the strategy relative to others. That is, trusting behavior is driven by fitness (i.e., subjects modify their behavior in response to fitness changes if $\beta > 0$). Such behavior can be interpreted as subjects either imitating a more successful other or as merely selecting a more successful action with a higher probability regardless of whether it was chosen or not in the current round (Camerer and Ho, 1999).

We test hypothesis 5 by examining whether α and β are significantly greater than zero. We run a nonlinear regression using data from Table 3 and report the results in Table 4. As shown, both RED and BLUE players tend to repeat their previously chosen strategies ($\alpha = 0.5262$ (RED), $p < 0.0001$; $\alpha = 0.8209$ (BLUE), $p < 0.0001$). In addition, results suggest RED players are sensitive to the relative fitness of trust ($\beta = 0.2794$, $p < 0.0001$). They trust if rewarded for the trust. However, we do not observe similar calculative behavior for BLUE players. Overall, these results suggest that both RED and BLUE exhibit status-quo bias and only RED seems to be sensitive to payoffs and engage in imitative behavior.

The behavioral difference in RED and BLUE players is interesting. We note that RED players are the initiator of the trust building process and own the original property rights of 4. BLUE players, on the other hand, do not explicitly invest in the trust building process. Hence, RED players may expect a higher payoff from the trust building process than the BLUE players. Consequently, they care more about the expected payoff than BLUE players.

Our results suggest the two sub-populations follow a different evolutionary trust building path. The evolutionary path of RED is governed by the relative fitness of trust while that of BLUE is not. We conjecture this is due to RED being endowed with an initial payoff of 4 before the trust-building process begins and hence this is a natural reference point to compare her actual payoff with. Consequently they are likely to be more conscious of the actual payoff. Conversely, BLUE does not have this initial endowment. And, BLUE appears to have a commitment to reward a trusting RED when he is certain of her trusting intention. Alternatively, one may interpret that RED has more power than BLUE in the sense that she can choose not to initiate the trust process without any inputs from BLUE. This difference in perceived power can cause her to be more sensitive to actual payoffs.

6. DISCUSSION

This research is designed to examine trust building in communities with anonymous exchanges. We define an anonymous exchange as one in which individuals interact with each other only once and do not know the identity of their partners. We disentangle trusting and trustworthiness behaviors using a novel experimental design. We find players in the population are not trustworthy when they are uncertain about others' trusting intentions. Interestingly, even with a total lack of formal mechanisms, and little (if any) informal sanctions due to the minimum scope of social relations, we find evidence of trusting behavior. Stake size appears to increase both trusting and trustworthy behaviors. Sub-populations however may follow different evolutionary paths of trusting behaviors. Some may trust based on its relative fitness and others may not. Results suggest sub-populations with the decision rights to initiate the trust building and those who hold the first rights to continue the process follow different evolutionary paths.

6.1 Methodological Issues

The experimental design allows us to differentiate and examine trust and trustworthiness through the choice set of subjects. Choices consist of two components – strategy (trust or not) and division of social gains (trustworthiness). Many previous studies fail to distinguish between these components of the trust process because they test trust using a sequential prisoner's dilemma design (Heckathorn,1988). Results suggest there is value in separating these components since we do see divergence in behavior. For example, in the initial stage of our

game there is significantly more trusting behavior than trustworthiness. We find a decrease in trusting behavior as the potential value of trust drops.

Our design uses a social exchange protocol in which players must decide what to do without knowing the choices of others (i.e., contingent strategy). Game theory predicts that using this design should not alter the strategy choices of subjects. However, previous research in prisoner's dilemma games suggests that subjects are more likely to cooperate if they know the other is willing to cooperate (Kiyonari et. al (in press); McCabe, Smith and LePore, 2000). If this finding generalizes to our trust game then we should see more trust building behavior when subjects are provided with the choices of others in previous stages.

We test this empirically.[16] We ran two additional 20-subject experiments in the United States using a sequential exchange protocol to determine whether social exchange protocol has an impact on trusting building. That is, subjects were not asked for contingent strategies. In each stage, they were told the choice of their pair member in the previous stage. For example, if RED passed in Stage 1, BLUE was told this and then asked what they wanted to do in Stage 2. If BLUE passed in Stage 2, we went back to RED and asked what she wanted to do in Stage 3, and so on. Figure 5 reports the results of these experiments. If we compare Figure 5 with Figure 2, we see that subject behaviors are alike. At each of the four stages, the degree of exhibited trust and trustworthiness is very similar. It is as if subjects can predict the behavior of others without knowing their action choices.

6.2 Theoretical Implications

Prior explanations for trust-building generally revolve around the promise of future

[16] In addition to testing behaviors in different exchange protocols, this manipulation provides an indirect test of whether subjects can predict others' intentions. Our results suggest they can in this setting.

interactions. Alexrod (1984) claims this shadow of the future helps to develop trust. Our experiment design allows us to test this formally. There is no shadow of the future; subjects are told they will not interact with one another for more than one round.[17] Verbal or non-verbal cues are not possible because subjects do not play face-to-face and play is anonymous. Yet, we see a significant level of trust and non-negligible level of trustworthiness develop.

We conjecture the following. In our experimental setup, trusting may lead to higher payoffs, so it is not surprising to see people trust. In fact people seem to trust more when the likelihood of a greater payoff is higher. For example, RED subjects who trust in stage 1 have a higher probability of receiving a payoff greater than the non-trusting one of 4 relative to that of the trusting RED subjects in stage 3 receiving a payoff greater than the non-trusting one of 16. Hence it is the promise of social gain that drives this trusting building process.

It is however more difficult to rationalize the significant level of trustworthiness behavior in the last stage. As discussed before, subjects at stage 4 exhibit the greatest level of trustworthiness because they are certain about the trusting intention of others. Some have suggested that our trust game can be interpreted as a multi-stage dictator game. Clearly, our game at stage 4 is a single-stage dictator game (Kahneman, Knetsch and Thaler, 1990; Forsythe, et. al, 1994; Hoffman, McCabe, Shachat, and Smith, 1994; Hoffman, McCabe and Smith, 1996; Malhotra and Murnighan, 2000).[18] Thus it might be interesting to compare our results with those

[17] We take steps to ensure that subjects believe in the instructions. For example, they are explicitly told that they will only interact with one another once. Subjects are also told there are 10 RED subjects and 10 BLUE subjects in each room so that we have the required number of subjects to match them with each other for each round. We also have shown subjects our scoring sheet before the experiment which clearly shows the different pairings in each round.

[18] Kiyonari and Yamagishi (1997) devised an interesting variant of the dictator game in which the respondent has an option whether or not to trust the proposer for a division. In their studies, the respondent could opt out for a sure payoff, which is less than half of the proposer's property rights. They found more than 80% of their subjects trusting and the average division of the proposer was 42%. Compared to our results from stage 3 and 4, these were significantly higher trusting and trustworthiness behaviors. One reviewer suggests that this difference could be due to the fact that a truster's payoff is totally dependent on a trustee's action in their game whereas the same action only accounts for one-tenth of a truster's payoff in our design.

from single-stage dictator games. Payments to the respondents (or in our case the RED subjects) are often interpreted as altruistic behavior (Camerer, 2003). At stage 4 of our game, this interpretation may not be appropriate because our stage 4 also contains trusting history from prior stages and subjects may want to reward this past trusting behavior.

We compute the mean proportion of payment to respondents in three similar studies that allow proposers to divide the property rights continuously from 0 to 100% (Forsythe, et. al, 1994; Hoffman, McCabe, Shachat, and Smith, 1994; Hoffman, McCabe and Smith, 1996). The mean payment to respondents is 17% and the modal response is 0% across 7 of the 9 experimental conditions.

Our mean proportion is 22.6%. This is higher than that from the above three dictator game studies. We speculate that some BLUE subjects initially reward RED's trusting behavior because they are certain about their trusting intention. However, this effect wears off gradually over time. This conjecture is supported by the fact that our modal response in rounds 1-5 was 50% whereas in rounds 6-10 it was 0%.[19]

Overall, our results suggest a significant number of subjects exhibit "pseudo trust" when interacting anonymously. This pseudo trust is cultivated in order to harness a return in exchange for the risk taken in giving up property rights. As long as everyone takes this risk, it benefits the population because this pseudo trust increases social gain. This expectation becomes a self-fulfilling prophecy or a population-level behavioral norm. For such a behavioral norm to develop, it is crucial that each player has the opportunity to regain the increased property rights. This is the case in the first 3 stages of the game. Consequently, players who want to stop trusting are also not trustworthy. Hence, it is possible to have a population who exhibits "pseudo trust",

[19] In rounds 1-5, of the 354 observations of BLUE subjects passing in stage 2 and taking in stage 4, 131 times (37.0%) they took 50% and 110 times (31.1%) they took 100%. In rounds 6-10, they took 50% 66 times (21.6%)

but is not trustworthy.

To sustain the pseudo trust behavior of BLUE in stage 2, there must be a non-negligible proportion of RED subjects who trust in stage 3. These RED subjects clearly trust for social gains because they will not regain the property rights. Figure 2 shows about 20% of RED exhibit this trusting behavior. The result that 20% of our population trust for social gain without any formal enforcement mechanism or informal sanctions mirrors results of earlier studies. For example, similar to the dictator-type games discussed above, Camerer and Weigelt (1988) find about 17% of their subjects exhibit this type of trust. These findings suggest about 15-20% of the subject population willingness to trust even in anonymous exchanges.

Our findings have implications for Granovetter's theory of embeddedness (1985). Granovetter suggests the network of social relations is an important determinant of trust between individuals. Macy and Skvoretz (1998) measure the degree of embeddedness in social relations by the percentage of interactions with neighbors (i.e., individuals with which a person has prior contact). For example, if embeddedness equals .5, then half of future interactions are with neighbors, and half with strangers. They examine trust-building in environments with degrees of embeddedness between 1 and .5. Using this measure, our experiment has a degree of embeddedness of 0. Since each subject only interacts with another for a maximum of 1 round, future interactions are always with strangers. Yet, despite the lack of embeddedness, we observe trusting and some degree of trustworthy behavior.

6.3 Future Research

We observe subject trusting and trustworthy behaviors are similar across all three countries in the early stages. For example, the proportion of trusting behavior in stage 1 is around

---

and 100% 177 times (58.0%) out of 306 times.

70% in all three countries (see Table 1). Similarly, the proportion of property rights claimed in stage 2 is around 95% in all three countries. Table 2a shows there is also little difference across countries in trustworthiness in the later stages (except in stage 3 where the mean proportion claimed by China subjects appears to be lower, though it is above 90%). Table 2b shows the level of trust in the later stages is different across the three countries. In particular, China subjects appear to exhibit a slightly higher level of trust than the other 2 countries in stages 2 and 3. Future research should investigate this difference more formally.

It is worth noting that Chinese subjects exhibited significantly more trusting and trustworthy behaviors in high-stake games. Put differently, subject behaviors deviate more from standard game theory predictions in these games. Future high-stake experiments in other countries would help to examine the generalizability of this finding.

An interesting research route is to modify the exchange context to eliminate the use of a credible signal by RED and a detection strategy by BLUE in stage 4. In its current form, the trust relationship ends in stage 4 with certainty. If this exchange only ends probabilistically, the ability of RED to send a credible signal and BLUE to detect the signal is considerably weakened. In this modified design, we expect less trustworthiness in stage 4.

Our results on stake size suggest another route of future research -- why does stake size increase the level of trust (c.f. McKelvey and Palfrey, 1992)? One possible explanation is that stake size changes the nature of social interaction. Subjects may feel a greater sense of reliance for each other in the high-stake games because they want to realize the huge increase in social gains. The penalty for non-trusting behavior is more costly. Consequently, they are more willingly to trust and reward trust. Such behavior can occur if subjects develop an initial expectation for it, which later is confirmed by observed behavior (i.e., a self-fulfilling

expectation).

Another path to pursue is to modify the rate of social gain (see for example McKelvey and Palfrey, 1992). In our study, social gain doubles every time a subject exhibits trusting behavior. What kind of trusting behavior might emerge if the rate of social gain is increasing in the amount of trust developed? Such a pattern of social gain seems plausible and reflects a greater efficiency among players. For example, the property rights could be 8, 24, 96 in Stage 2, 3, and 4 respectively.

Lastly, it may be interesting to extend this research is to manipulate social embededness.[20] We can easily do this by varying the number of times each pair of players interact in each round and the probability of their rematch in the future rounds. In this study, each pair of RED and BLUE players interact once in each round and their rematch probability is zero in future rounds.

[20] Ellison (1993) shows mathematically that the rate at which dynamic behaviors converge to the Pareto efficient equilibrium in a coordination game depends on the structure of the interaction. A faster rate of convergence can be achieved if the interaction is confined only to nearest neighbors compared to the interaction among strangers.

7. REFERENCES

Alchian, Armen and Harold Demsetz. 1973 "The property rights paradigm." *Journal of Economic History 33*, 16-27.

Axelrod, R. 1984. *TheEvolution of Cooperation.* New York: Basic Books.

Berg, Joyce, John, Dickhaut and Kevin McCabe. 1995. "Trust, reciprocity, and social history", *Games and Economic Behavior 10*: 122-142.

Bonner, Sarah, Young, Mark, and Reid Hastie. 1996. "Financial incentives and performance in laboratory tasks: The effects of task type and incentive scheme type", unpublished manuscript, University of Southern California Department of Accounting.

Bolle, Friedel. 1995. "Does Trust Pay?" Europa-Universitat Viadrina Working Paper.

Bradach, J. and R. Eccles. 1989. "Price, authority, trust: From ideal types to plural forms", *Annual Review of Sociology 15*: 97-118.

Borgers, T. and R. Sarin. 1997. "Learning Through Reinforcement and Replicator Dynamics, *Journal of Economic Theory* 77 (1997), 1-14.

Camerer, Colin. *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton University Press, 2003.

Camerer, Colin and Teck-Hua Ho. 1999. "Experience-Weighted Attraction Learning in Normal-form Games," *Econometrica*, 67, 4:827-874.

Camerer, C. Teck-Hua Ho, and Juin-Kuan Chong, "Sophisticated Learning and Strategic Teaching," *Journal of Economic Theory*, 104 (2002), 137-188.

Camerer, Colin and Robin Hogarth. 1999. "The Effects of Financial Incentives in Economic Experiments: A Review and Capital-Labor Production Framework," *Journal of Risk and Uncertainty*, 1999, pp. 7-42.

Camerer, Colin and Keith Weigelt. 1988. "Experimental Tests of a Sequential Equilibrium Reputation Model," *Econometrica*, 56:1:1:37.

Chiles, T.H. and J.F. McMackin. 1996. "Integrating variable risk preferences, trust, and transaction cost economics." *Academy of Management Review*, 21: 73-99.

Craswell, Richard. 1993. "On the uses of "trust": Comment on Williamson, "Calculativeness, trust, and economic organization", *Journal of Law & Economics, 36*: 487-500.

Creed, W.E.D. and R. E. Miles. 1994. "Trust in organizations: A conceptual framework linking organizational forms, managerial philosophies, and the opportunity costs of controls." In R.M. Kramer and T.R. Tyler *Trust in Organizations: Frontiers of Theory and Research*, 16-38.

Ellison, Glenn. 1993. "Learning, local interaction, and coordination." *Econometrica* 61:1047-71.

Erickson, E.H. 1963. *Childhood and society*. New York: Norton.

Etzioni, A. 1988. *The moral dimension: Toward a new economics*. New York: The Free Press.

Forsythe, Robert, Joel Horowitz, N. Z. Savin, and Martin Sefton. 1994. "Fairness in simple bargaining experiments," *Games and Economic Behavior*, 6, 347-369.

Friedman, D. 1991. "Evolutionary games in economics." *Econometrica*, 58, 637-666.

Fudenberg, D. and D. Levine. 1998. *The theory of learning in games*. Cambridge: MIT Press.

Fukuyama, F. 1995. *Trust: The social virtues and the creation of prosperity*. New York: The Free Press.

Gauthier, David. 1986. *Morals by agreement*. Oxford, England: Clarendon.

Glaeser, Edward, David, Laibson, Jose Scheinkman, and Christine Soutter. 2000. "Measuring trust," *Quarterly Journal of Economics*, pp. 811-846.

Granovetter, M. 1985. "Economic action and social structure: The problem of embeddedness." *American Journal of Sociology 91*: 481-510.

Greene, Graham. 1943. *The Ministry of Fear*. New York: Penguin Books.

Gurtman, M.B. 1992. "Trust, distrust, and interpersonal problems: A circumplex analysis." *Journal of Personality and Social Psychology 62*: 989-1002.

Hayashi Nahoko and Toshio Yamagishi. 1998. "Selective Play: Choosing Partners in an Uncertain World." *Personality and Social Psychology Review*. Vol. 2, 4:276-289.

Heckathorn, Douglas. 1988. "Collective sanctions and the emergence of prisoner's dilemma norms," *American Sociological Review 55*: 366-384.

Hertwig, Ralph and Andreas Ortmann. 1998. "Experimental practices in economics: A methodological challenge for psychologists," Unpublished manuscript, Max Planck Institute for Human Development, Berlin.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. 1994. "Preferences, property rights and anonymity in bargaining games," *Games and Economic Behavior*, 7, 346-380.

Hoffman, Elizabeth, Kevin McCabe, and Vernon Smith. 1996. "Social distance and other regarding behavior in dictator games," *American Economic Review*, Vol. 86, No. 3, pp. 653-660.

Hoffman, Elizabeth, Kevin McCabe, and Vernon Smith. 1998 "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," *Economic Inquiry*, Vol. 36, No. 3, pp 335-352.

Holmes, R. 1985. *Acts of War: The Behavior of Men in Battle*, New York: The Free Press.

Jenkins, Douglas Jr.; Mitra, Atul,; Gupta, Nina; and Jason Shaw. 1998. "Are financial incentives related to performance? A meta-analytic review of empirical research", *Journal of Applied Psychology* 83, p. 777-787.

Kahneman, Daniel, Jack, L. Knetsch, and Richard Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy* 98, No. 6. , pp. 1325-1348.

Kee, H. and R.T. Knox. 1970. "Conceptual and methodological considerations in the study of trust and suspicion", *Journal of Conflict Resolution 14:* 357-365.

Kiyonari, Tohko and Yamagishi, T. 1997. "A new game of trust: Letting the partner be the dictator." The Seventh International Conference on Social Dilemmas.

Kiyonari, Tohko, Shigehito Tanida, and Toshio Yamagishi. In Press. "Social Exchange and Reciprocity: Confusion or a Heuristic?" *Evolution and Human Behavior*.

Kollock, P. 1994. "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust." *American Journal of Sociology 100*: 313-45.

Macy, M. and J. Skvoretz. 1998. "The Evolution of Trust and Cooperation Between Strangers: A Computational Model," *American Sociological Review 63*: 638-660.

Malhotra, Deepak and J. Keith Murnighan. 2000. "Self versus social impression management in dictatorship games," Unpublished manuscript, Northwestern University.

Mayer, R.C., Davis, J.H. and F.D. Schoorman. 1995. "An integrative model of organizational trust," *Academy of Management Review 20*: 709-734.

McAllister, D.J. 1995. "Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations." *Academy of Management Journal 38:* 24-59.

McCabe, K., Smith, V., & LePore, M. (2000). Intentionality detection and "mindreading": Why does game form matter? *Proceedings of the National Academy of Science of the United States of America*, *97*, 4404-4409.

McKelvey, R. and T. Palfrey. 1992. "An experimental study of the centipede game," *Econometrica 60*: 803-836.

Morikawa, T., Orbell, J. M. and Runde, A. S. 1995. "The Advantage of Being Moderately Cooperative", *American Political Science Review*, 89, 601-611.

Myerson, Roger. 1991. *Game Theory: Analysis of Conflict*. Cambridge: Harvard University Press.

Orbell, John; P. Schwartz-Shea, and Simmons, R. T. 1984. "Do cooperators exit more readily than deflectors?" *American Political Science Review*, 78, 147-162.

Orbell, John and Robin Dawes. 1991. "A `cognitive miser' theory of cooperators' advantage." *American Political Science Review* 85: 515-28.

Orbell, John and Robyn Dawes. 1993. "Social welfare, cooperators' advantage, and the option of not playing the game," *American Sociological Review 58*: 787-800.

Ring, P.S. and A.H. Van de Ven. 1992. "Structuring cooperative relationships between organizations." *Strategic Management Journal 13*: 483-98.

Rosen, Sherwin. 1982. "Authority, control and the distribution of earnings, " *Bell Journal of Economics 13*: 311-323.

Rosenthal, R. 1982. "Games of perfect information, predatory pricing, and the chain-store paradox," *Journal of Economic Theory 25*: 92-100.

Shapiro, S.P. 1987. "The social control of impersonal trust." *American Journal of Sociology 93*: 623-58.

Simon, H. A. 1957. *Models of man: Social and rational*. New York: Wiley.

Smith, Vernon. and Walker, James, M. 1993. "Rewards, experience, and decision costs in first price auctions." *Economic Inquiry* 31, 237-244.

Snijders, C. and G. Kerem, 1999 "Determinants of Trust" in (eds.) D. Budescu, I. Erev, and R. Zwick. *Games and Human Behavior*, New York: LEA Publishers.

Uzzi, Brian. 1996. "Embeddedness and Economic Performance: The Network Effect." *American Sociological Review 61*: 674-698.

Williamson, Oliver. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: The Free Press.

Wright, Robert. 1995. *The Moral Animal*. New York: Vintage Books.

Yamagishi, T., K.A. Cook, and WM. Watabe, 1998. "Uncertainty, Trust, and Commitment Formation in the United States and Japan." *American Journal of Sociology 104*: 165-94.

Yamagishi, T., and M. Yamagishi, 1994. "Trust and Commitment in the United States and Japan." *Motivation and Emotion* 18: 9-66.

Zucker, L. 1986. "Production of trust: Institutional sources of economic structure, 1840-1920", in *Research in Organizational Behavior 8*, JAI Press, p. 53-111.

APPENDIX I

NAME _____ SUBJECT NO. _____

INSTRUCTIONS

This is an experiment in decision-making. The instructions are simple; if you follow them carefully and make good decisions, you could earn a considerable amount of money which will be paid to you in cash before you leave today. Different subjects may earn different amounts of cash. What you earn today partly depends on your decisions, partly on the decisions of others, and partly on chance.

There are 20 subjects in this room. Each subject has been randomly assigned a subject number. When we finish reading these instructions, 10 subjects will be told to go to another room. When this experiment begins, you will always be paired with a subject in the other room. You will never play with a subject in your room. The experiment will consist of 10 decision rounds. In these 10 decision rounds, you will be paired with each of the 10 subjects in the other room for 1 round. So you will always be paired with a subject in the other room, and you will never play with any subject more than once.

It is important that you do not look at the decisions of others, and that you do not talk, laugh, or exclaim aloud during the experiment. You will be warned if you violate this rule the first time. If you violate this rule twice, you will be asked to leave, and you will not be paid. That is, your earnings will be $0.

Experimental procedure

Subjects will be randomly assigned to one of two rooms. Subjects in this room will belong to the "RED" group; subjects in the other room to the "BLUE" group. Whether you belong to the RED or BLUE group will be determined as follows. An administrator will circulate among you with a box containing numbered balls. As the administrator passes by, you select a ball from the box. If the ball you select is an even number then you will be a member of the RED group, and you will remain a RED member for the entire experiment. All RED members will go to the other room. If the ball is an odd number, you are a member of the BLUE group, and you will remain a BLUE member for the entire experiment. All BLUE members will remain in this room.

In this experiment you will play the following game. First, you will be randomly matched with a subject of the opposite color. Hence a BLUE member will always be paired with a RED member and vice versa. The following Figure shows how the game will work.

[SEE FIGURE 1]

RED moves first in Stage 1. If RED chooses "take" then the RED player can choose to take some percentage (a) of 4 points. This percentage can range from 0% to 100%. The BLUE player receives the difference (1 - a). For example, if the RED player chooses an a of 70%, the RED player receives 2.8 points (70% X 4), and the BLUE player receives 1.2 points ((1 - .7) X 4). Or, if the RED player chooses an a of 20%, the RED player receives .8 points (20% X 4), and the BLUE player 3.2 points ((1 - .2) X 4). If the RED player passes then the BLUE player must choose between "pass" or "take". In Stage 2, the pool of points doubles. So the BLUE player must choose to take some percentage (b) of 8 points. For example, if the BLUE player chooses 90% then the BLUE player receives 7.2 points (90% X 8), and the RED player receives .8 points ((1 - .9) X 8). If the BLUE player "passes" we move to Stage 3. At Stage 3, the pool of points doubles again to 16 points, and the RED player must decide whether to "take" or "pass". If the RED player "takes", the player must decide what percentage of 16 points she will keep, and what percentage to give to the BLUE player. If the RED player "passes", we move to Stage 4. At Stage 4, the pool of points doubles to 32. Here, if the BLUE player "takes" the player must decide how to divide the 32 points between himself and the RED player. If the BLUE player "passes", the game ends with both players receiving 0 points.

Once the game reaches Stage 4, or a player "takes", subjects are awarded the points they earn for that round, and they move to the next round. In the next round they will play the same game, but <u>they will be paired with a new player</u>. Again, you will never play another subject more than once. We will continue this procedure for 10 rounds. Since we have two groups of 10 subjects and members of a group do not play against each other, <u>none</u> of your pair members have played against each other before.

In each round, to save time, you are asked to indicate what you would do if you choose "pass" in earlier stages of the game. For example, if a RED player chooses "pass" in stage 1, we will ask the player what she would do in stage 3. Similarly, if a BLUE player chooses "pass" in stage 2 we will ask what they would do in stage 4. Note that the game may actually end earlier so your choice may be immaterial. For example, if a RED player chooses to "pass" in stage 1, and his BLUE pair member chooses to "take" in stage 2, then the choice of the RED player in stage 3 is meaningless. The game has ended in stage 2. We ask you to tell us your future choices to save time in passing messages back and forth. After each round we will tell you your payoff for that round, and the choice of your pair member.

Payoffs

Your dollar earnings for the experiment are determined as follows. First, we will sum up your point earnings over the 10 rounds. Then we will multiply this sum by $0.10. We will pay you this amount when you leave the experiment. Note the more points you earn, the more money you will receive.

SCORING SHEET FOR RED PLAYER

ROUND	STAGE 1	IF TAKE, a =	STAGE 3	IF TAKE, c =	PAIR CHOICE	EARN
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
TOTAL						

SCORING SHEET FOR BLUE PLAYER

ROUND	STAGE 2	IF TAKE, b =	STAGE 4	IF TAKE, d =	PAIR CHOICE	EARN
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
TOTAL						

Figure 1: The Trust Building Game

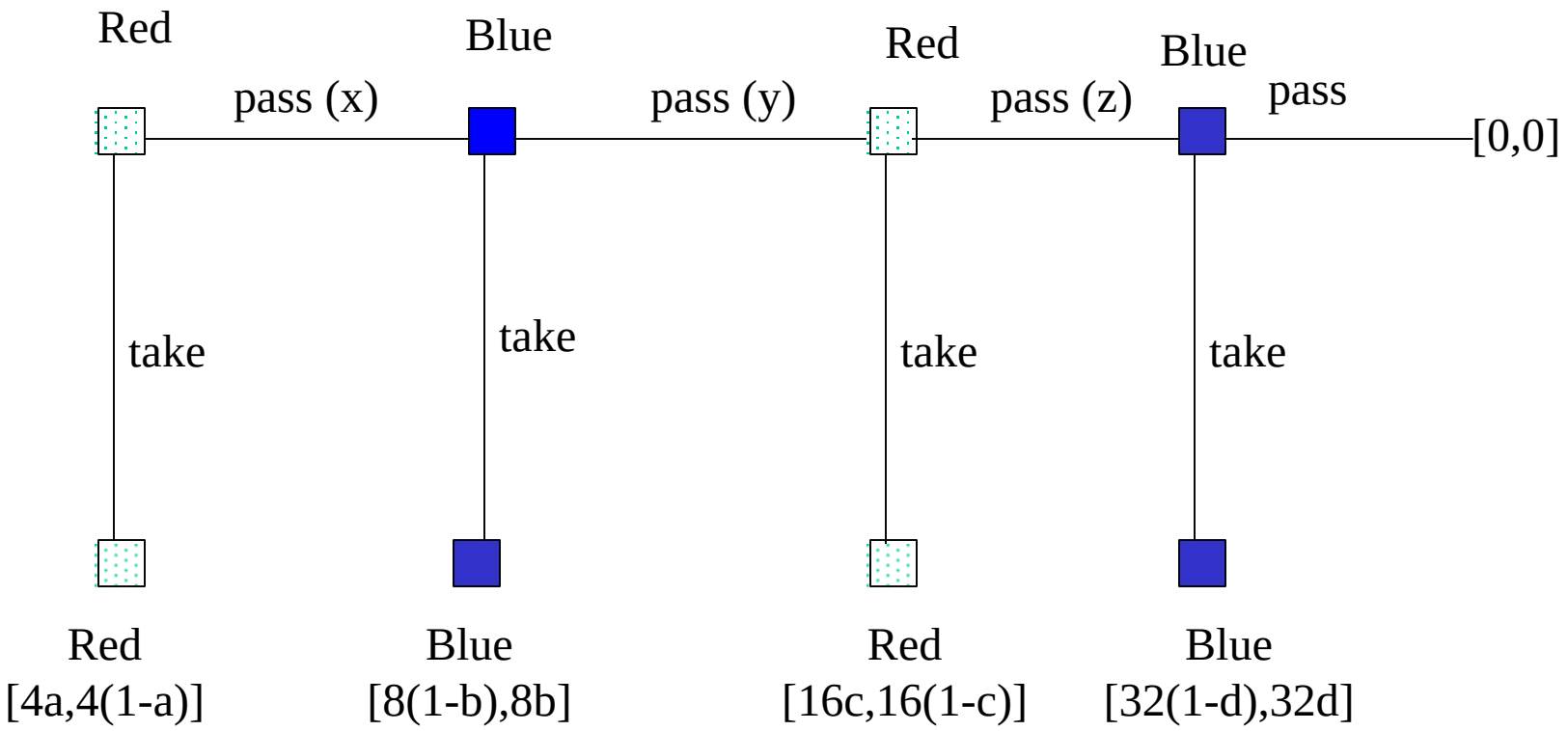
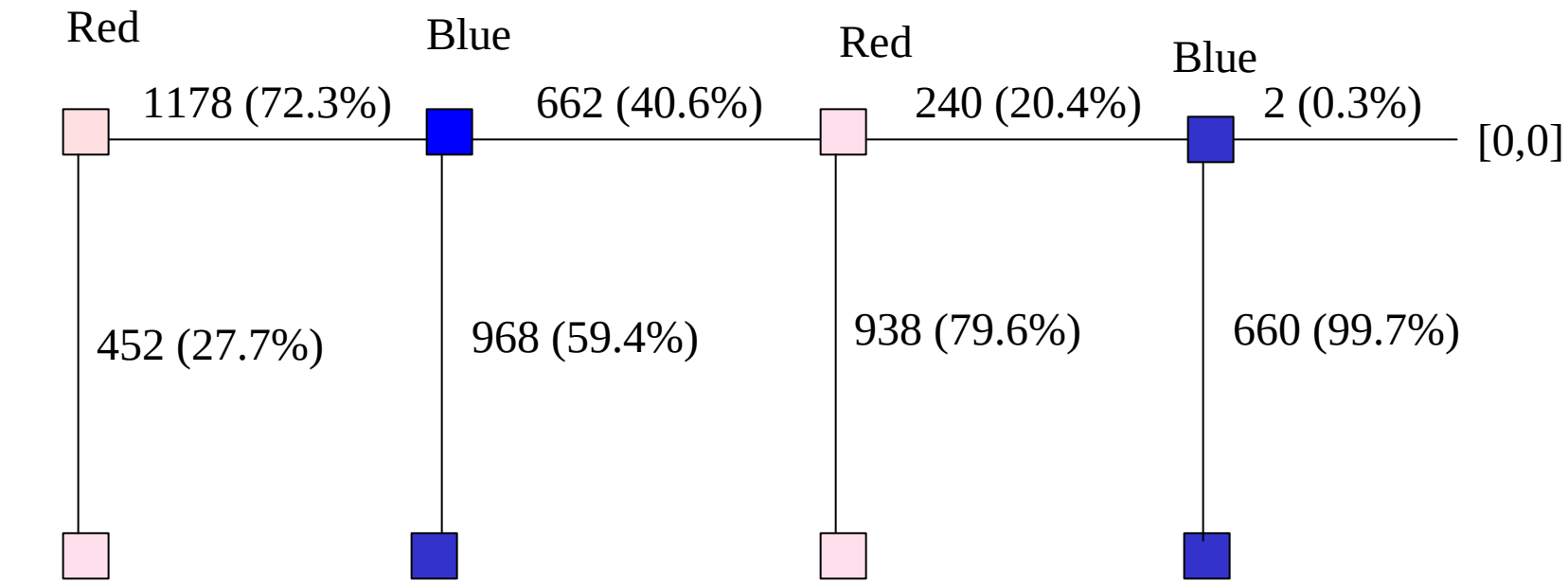


Figure 2: Descriptive Statistics of Players' Choice
(pooled across 13 experimental sessions and 10 rounds)



a = 0.9645 b = 0.9520 c = 0.9378 d = 0.7753

Figure 3: Calculation of Fitness (Expected Payoff) of Trust
(For Singaporean Experiment in Period 1)

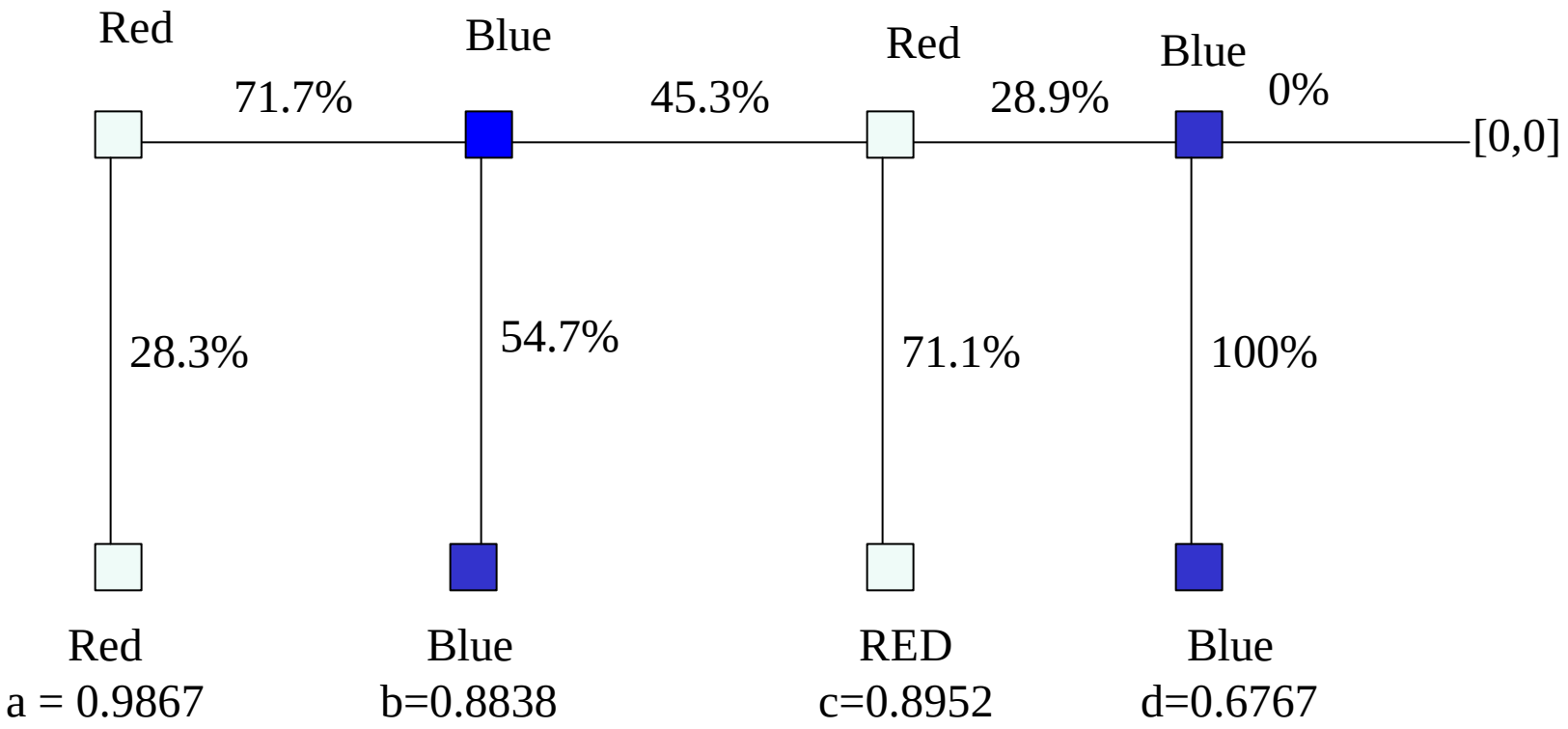


Figure 4: Descriptive Statistics of Players' Choice in the High-stake Games

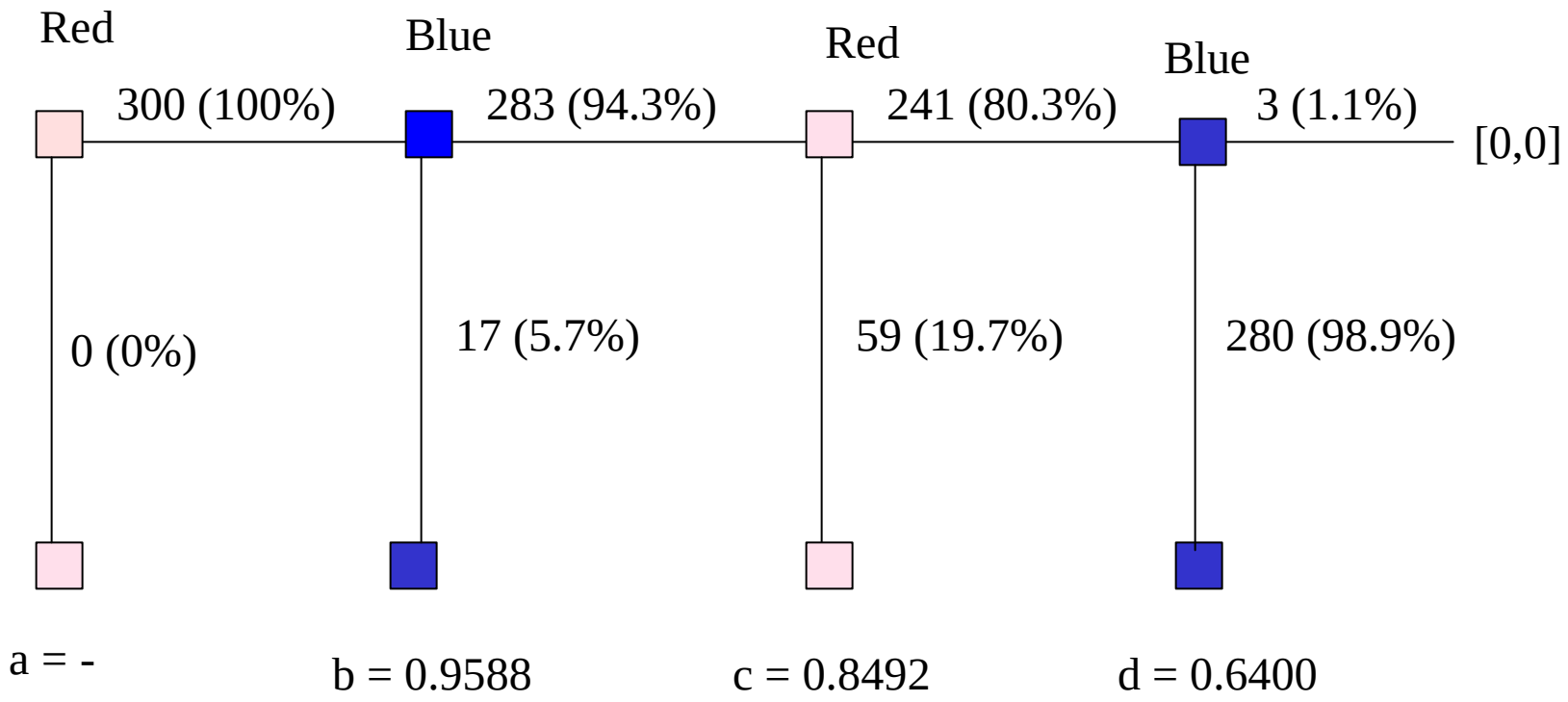
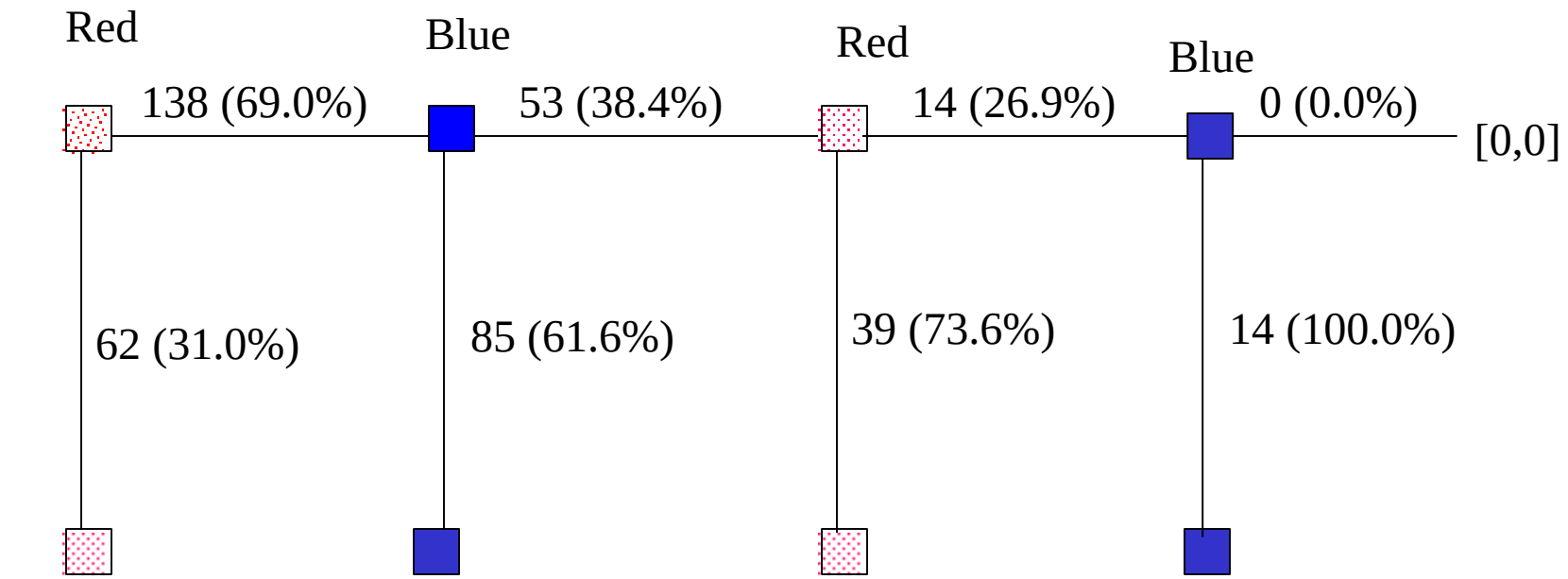


Figure 5: Descriptive Statistics of Players' Choice
(Sequential Exchange Protocol)



a = 0.9756 b = 0.9741 c = 0.9053 d = 0.7920

Table 1: Percentages of Subjects Showing Trust in Stage 1 and Trustworthy Behavior in Stage 2

		Country	
	China	Singapore	USA
Mean Proportion of Trust in Stage 1	73.3%	73.0%	69.8%
95% Confidence Interval	[71.7%, 74.9%]	[71.3%, 74.7%]	[68.0%, 71.7%]
Mean Proportion of Trustworthiness in Stage 2	96.04%	96.44%	95.06%
95% Confidence Interval	[93.47%, 96.67%]	[95.16, 97.71%]	[93.47%, 96.66%]

Table 2a: A Test of Trustworthiness (Percentages of Property Rights Claimed)

		Decision Stages		
Country	1	2	3	4
China				
Mean Proportion	-	96.04%	91.77%	79.17%
95% Confidence Interval		[93.47%, 96.67%]	[89.29%, 93.06%]	[76.44%, 81.91%]
Singapore				
Mean Proportion	-	96.44%	96.39%	79.30%
95% Confidence Interval		[95.16, 97.71%]	[95.18%, 97.62%]	[76.07%, 82.52%]
USA				
Mean Proportion	-	95.06%	93.71%	73.73%
95% Confidence Interval		[93.47%, 96.66%]	[92.10%, 95.33%]	[70.45%, 77.01%]

Table 2b: A Test of Trusting Behavior (% of Subjects who pass at each stage)

		Decision	Stages	
Country	1	2	3	4
China				
Mean Percentages	73.3%	48.3%	25.8%	-
95% Confidence Interval	[71.7%, 74.9%]	[46.3%, 50.3%]	[24.0%, 27.6%]	
Singapore				
Mean Percentages	73.0%	34.2%	20.2%	-
95% Confidence Interval	[71.3%, 74.7%]	[30.9%, 37.5%]	[18.6%, 21.8%]	
USA				
Mean Percentages	69.8%	38.2%	13.8%	-
95% Confidence Interval	[68.0%, 71.7%]	[36.1%, 40.3%]	[12.6%, 15.0%]	

Table 3: Relative Fitness Of Different Trusting Behavior Paths Based On Observed Choices

Singapore										
RED	1	2	3	4	5	6	7	8	9	10
T	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
PT	7.76	7.21	7.87	5.88	5.06	4.66	5.17	3.67	3.71	5.53
PP	5.20	4.61	4.23	2.21	2.37	1.97	1.50	0.69	0.84	0.88
BLUE										
T	5.75	6.04	6.79	6.49	6.19	6.34	5.28	5.58	4.70	5.28
PT	10.12	6.80	10.14	5.99	9.51	7.76	4.19	3.79	2.68	4.00
USA										
RED	1	2	3	4	5	6	7	8	9	10
T	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
PT	9.38	8.27	7.51	6.67	8.81	5.28	4.90	4.02	4.38	4.39
PP	6.00	5.42	4.35	4.49	4.94	2.05	2.15	2.16	2.23	1.00
BLUE										
T	6.51	5.58	6.41	6.45	5.98	6.43	4.99	4.71	4.50	4.71
PT	7.84	7.31	8.06	4.80	3.26	4.37	3.68	3.62	2.71	3.37
China										
RED	1	2	3	4	5	6	7	8	9	10
T	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
PT	7.43	7.38	10.45	6.06	7.78	7.58	8.65	7.69	8.80	8.00
PP	4.17	4.34	5.60	3.66	3.30	2.78	3.32	3.04	2.77	1.71
BLUE										
T	5.79	6.32	5.36	6.52	6.33	6.05	5.76	6.01	5.79	5.58
PT	9.16	9.88	10.11	9.27	9.64	7.06	10.00	8.12	9.10	8.69

Table 4: A Test of Evolutionary Trust Building Paths Acorss Sub-populations

RED			
	Estimate	t-statistic	p-value
α	0.5261	5.83	0.0001
β	0.2794	8.01	0.0001
Adj. R-Sq	0.8156		
BLUE			
	Estimate	t-statistic	p-value
α	0.8208	6.07	0.0001
β	-0.0591	-0.32	0.7489
Adj. R-Sq	0.2136		