Trust and Social Exchange

Rick K. Wilson
Department of Political Science, MS 24
Rice University
Houston, TX 77005
rkw@rice.edu
713.348.3352

Catherine C. Eckel
School of Economic, Political and Policy Sciences
University of Texas at Dallas
Mail Code GR31
800 W. Campbell Rd.
Richardson, TX 75080
eckelc@utdallas.edu
972-883-4949
Trust and Social Exchange
Rick K. Wilson (Rice University)
Catherine C. Eckel (University of Texas at Dallas)

Abstract:

Trust is broadly construed as a willingness on the part of individuals to put their wellbeing in the hands of other persons. In survey research it is typically assessed at a societal level by asking questions about “others” in a community, while in experimental research, it most often refers to decisions made in situations involving two persons, where reciprocated trust has the potential to improve the monetary wellbeing of both participants. This dyadic trust is an exchange transaction. A decision to trust involves an evaluation of costs and benefits, including a strategic assessment of the trustworthiness of the trusted person. A decision about whether and how much to reciprocate is likely to be determined primarily by norms concerning fairness or obligation, though it can also be strategic if the transaction is to be repeated.

The earliest work on trust used survey methods; a set of general questions assessed a person’s perceptions about whether others in their communities were trustworthy or fair. We begin with a brief review of research based on this set of survey questions. More recently, there has been considerable growth in behavioral research on trust, consisting of laboratory and field experiments, with most studies using a version of the “investment game” devised by Berg, Dickhaut and McCabe (1995). This research has provided an important challenge to the survey-based research, as it shows only weak relationships between survey responses and behavior in the games. We survey this work, highlighting methodological concerns. Separate sections consider the correlates of trust, trust as a relationship between two people, and the impact of specific institutional factors.

Note: We appreciate the help provided by Seonghui Lee and Cathy Tipton. Useful comments were provided by Salvador Vázquez del Mercado, David Llanos, Skip Lupia and Jamie Druckman. We gratefully acknowledge support by the National Science Foundation SES-0318116 and SES-0318180. That agency is not responsible for the findings reported here.

Word Count: 8608 (excluding title page and abstract)
6464 (excluding references, title page and abstract)
Introduction

Trust is a key concept in political science. It typically is treated as a perception or belief held by individuals about the overall level of trust and trustworthiness of others in society, in other words, “generalized” trust. Trust between citizens is considered valuable. Some think that it leads to increasing social capital (Putnam 2000), others think that it strengthens political institutions (Rothstein 2000), while others argue it provides political leaders with the leeway to reach compromise in representative democracies (Bianco 1994). Uslaner (2002) concludes that trust is “the chicken soup of social life”, but like chicken soup, its power to cure all ills is oversold.

Our concern is with trust among citizens, as opposed to citizen trust in government. (For this latter issue, see the extended debates among Miller (1974), Citrin (1974), Hetherington (1998) and Hibbing and Theiss-Morse (2002)). The bulk of the work focused on citizen trust uses survey questions measuring generalized trust, the same questions have been used for many years, and the trends are disturbing, indicating that levels of trust are in decline (Alford 2002). The causes of this decline are in considerable dispute ((Jackman and Miller 1998; Putnam 2000; Levi and Stoker 2000; Sobel 2002; Nannestad 2008)). Seemingly disengaged from this debate is a rich body of experimental studies that point to considerable trust and trustworthiness among individuals. This raises the question of what laboratory-based bargaining games and observational studies can jointly contribute to understanding the complex nature of trust.

1. Generalized Trust

While there is widespread recognition that generalized trust is a central concept, there is substantial disagreement about its relationship with important political phenomena. To some it is seen as a critical foundation for stable political institutions and for the formation of social capital and civic engagement (Putnam 1993; Putnam 2000; Stolle 1998). Others focus on the extent to which particular aspects of political institutions, such as transparency, appropriate levels of monitoring, and credible threats of sanctions, can foster trust and cooperation (Ostrom 2003; Lupia and McCubbins 1998; Sztopka 1999; Knight 2001). Still others see trust as a social lubricant that reduces the cost of exchange, whether in reaching political compromise (Fenno 1978; Bianco 1994) or in daily market and nonmarket exchange (Arrow 1974). Trust affects and is affected by political institutions, and plays an important role in limiting or enhancing the effectiveness of those institutions, but the complex causal relationships between trust and institutions are difficult to disentangle from observational studies alone. While enormous insights have been gained from such studies, a number of important questions that remain that experimental research is well-suited to address.

Trust has been measured using standard survey questions in the General Social Survey and the American National Election Survey for more than 40 years. The main question asks: “Generally speaking, would you say that most people can be trusted or
that you can’t be too careful in dealing with people?” Two additional questions gauge perceptions of whether people are helpful or fair. Levi and Stoker (2000) and Nannestad (2008) note that it is unclear what these generalized trust measures capture. Both agree that the standard measures tackle a very general aspect of interpersonal trust; the items are ambiguous concerning the trust relationship, and most of the survey studies do not allow for causal claims. Concerns about the use of survey data led Levi and Stoker (2000) to call for additional measures, especially behavioral measures, in order to understand the degree to which trust is manifested in a society. The point is driven home by Glaeser et al. (2000) who compare standard survey questions with behavioral trust measures ranging from self-reported behavior to a version of the trust game noted below (see also Ben-Ner et al. (2007), Bellemare and Kröger (2007) and Naef et al. (2009)). Disturbingly, they find there is very little correlation between survey and behavioral measures. At best measures of generalized trust tend to predict only the trustworthiness side of the exchange.

We believe that laboratory experiments are uniquely situated to address three concerns. First, a primary purpose of experimental methodology is to tackle questions of causality. The key causal question is whether trust influences political institutions or whether (and how) political institutions foster increased rates of trust. Second, doubts have been raised about the ways in which trust has been measured in observational settings. Experiments are well suited for refining and calibrating of measures of trust. In the lab it is easy to stress-test measurement ‘instruments’ used for gauging trust, whether based on surveys or behavior. Finally, experiments provide a powerful tool for understanding strategic behavior. We argue below that strategic interaction is a core element of the trust relationship. By carefully manipulating the information that subjects have about one another, it is possible to shed light on the strategic considerations of trusters and trustees.

2. Dyadic Trust

At its core, trust involves a strategic relationship among two actors. In political science this includes negotiations among legislators to trade votes, the decision of a voter to give latitude to a representative, or the willingness of a citizen to comply with the decision of a public official. This concept of dyadic trust is similar to Hardin’s (2002) view of “encapsulated” trust, where two actors know something about one another, the context of exchange is clear, and what is being entrusted is well-defined. These conditions produce clear expectations about trusting and trustworthy behavior. In the case of legislators engaged in vote trading, one legislator faces the problem of giving up a vote with the future promise of reciprocity. Knowing the reputation of the other party is critical to this choice.

Both parties in a dyadic trust relationship have important strategic considerations. The problem for the truster is whether or how much to trust the trustee. The problem for the trustee is to decide, if trusted, whether and how much to reciprocate that trust. Each party, but especially the truster, makes decisions based on expectations about the other. Focusing on dyadic trust illustrates the missing element in discussions of generalized
trust: trust depends on who (or what) one is dealing with. Individuals do not “trust” in the abstract, but rather with respect to a specific target and in a particular context. While the decision about whether and how to reciprocate does not carry the same level of risk for the trustee, their decision also is made in a specific context, with attendant norms of responsibility or obligation. This strategic interaction is difficult to explore in the context of many observational studies, but is well-suited to the laboratory.

Since the mid 1990s more than 150 experimental studies have examined dyadic trust. The standard trust experiment, originally known as the “investment game” (Berg et al. 1995), has proved to be a valuable vehicle for subsequent research. It has given rise to new methodological innovations in experimental protocols, allowed researchers to examine the correlation between behavior and individual characteristics (including neuroscience innovations), provided an environment to study stereotyping and discrimination, and served as a platform for cross-cultural comparison. In addition, it has allowed researchers to examine existing institutional mechanisms and ‘testbed’ new institutions. In the remainder of this chapter, we examine these aspects of trust experiments and conclude with a set of unanswered questions.

3. The Trust Game

The trust game consists of a sequence of moves between two actors, where both are fully informed about its structure and payoffs. To illustrate, suppose there are two actors, Player A and Player B. Both are endowed with $10 by the experimenter. Player A has the right to move first, and can choose to keep the $10 or can pass any part of it to the second player. Any amount that is passed is tripled by the experimenter, and then delivered to Player B. (The tripling plays the part of a return on an investment in the game.) Player B now has her original $10 and the tripled amount passed to her, and is given the option to send some money back to Player A. The amount can range from $0 to the full tripled value. Player A’s move is “trust”, in that by sending a positive amount, he entrusts his payoff to Player B; Player B’s move is “trustworthiness” or reciprocity. From a game-theoretic perspective, a naive, payoff-maximizing Player B would retain anything sent to her: Player A, knowing this, will send her nothing. Thus the equilibrium of the game (assuming payoff-maximizing agents) is for Player A to send 0, rightly failing to trust in Player B’s trustworthiness.

The canonical implementation of this game has the following characteristics:

- Subjects are recruited from the general student population, and paid a nominal show up fee, usually $5
- Subjects are randomly assigned to the role of Player A or B
- Brief instructions are read aloud, followed by self-paced computerized instructions and a comprehension quiz
- Each player is endowed with an equal amount of money (usually $10)
- Partners are kept anonymous
- A brief questionnaire collects demographic and other information
Subjects are paid their actual earnings in cash, in private, at the end the experiment.

In contrast to the Nash equilibrium, a meta-analysis of results by Johnson and Mislin (2008) shows that, on average, trusters send 50.8% of their endowment (based on 84 experiments). Trust pays (barely), in that 36.5% of what is sent is returned (based on 75 experiments), just over the 33.3% that compensates Player A for what was sent. Contrary to game theoretic expectations, trust is widespread and it is reciprocated. Several methodological issues have been raised concerning the canonical experimental design and we turn to these before discussing more general findings.

Methodological Issues

As others in this volume have noted, political scientists are sometimes skeptical of what laboratory experiments can tell us. Experiments seem contrived, the sample is too limited and the motivations of subjects often seem trivial (see for example the discussions by Dickson (2009), Druckman (2009) and McDermott(2009)). A common complaint about laboratory experiments is that, even if subjects are paid, the stakes are insufficient to mimic what happens in natural settings. Johansson-Stenman et al. (2005) ask whether stakes matter in the trust game conducted in Bangladesh, with the highest-stake game being 25 times greater than the lowest-stake game. The high-stakes setting has US dollar price parity equivalent of $1683. They find that as the size of stakes increases, a slightly smaller percentage of the money is sent (38% in the high stakes condition compared to 46% in the middle stakes condition). In experiments carried out in Tatarstan and Siberia, subjects were given the equivalent of a full day’s wage (Bahry and Wilson 2004); 62 percent sent half or more of their endowment. While it is true that subjects are not completely insensitive to size of stakes, the data clearly show that people trust at high rates even when the stakes are high.

Another common complaint is that students coming into the laboratory are friends and/or anticipate post-game repeated play. High levels of trust may simply be due to subjects investing in reputation. Anderhub, Engleman and Guth (2002) and Engle-Warnick and Slonim (2004) find that reputational effects quickly appear when subjects repeatedly play the trust game. Isolating ‘trust’ from an investment in reputation is important and experimenters take considerable care to ensure that subjects do not know one another in the same experimental session. To ensure complete anonymity, we (2006) conduct experiments over the internet, with subjects matched with others at another site, more than 1000 miles away. When care is taken to verify for subjects that their counterparts indeed exist, the subsequent play of the game is within the range observed in other studies.

A final complaint is that the game doesn’t really measure trust, but rather some other thing such as other-regarding preferences. Glaeser et al (2000) ask subjects to report the frequency of small trusting acts – leaving a door unlocked, loaning money to a friend – and find positive correlations between these actions and the trust game. Karlan (2005) finds that the repayment of micro-credit loans is positively correlated with trustworthiness in the trust game, but not trust. The behavioral measures of trust have
weak correlation with generalized trust questions. Cox (2004) argues that behavior in the 
trust game includes elements of cooperativeness. He argues that trust is “altruism plus”, 
and designs experiments to isolate trust and trustworthiness from altruistic tendencies of 
subjects. He finds positive levels of trust even controlling for individual-level altruism.

While initial studies were designed to test game theoretic conjectures about 
exchange in settings without reputation, and found high levels of trust and reciprocity on 
average, more recent studies have tried to understand sources of heterogeneity in behavior across individuals. Our own work examines how trust and trustworthiness are 
affected as information about partners is revealed. Finally there is a body of work on the 
effect of institutional changes – changes in the ‘rules of the game’. We discuss each in 
turn.

4. Correlates of trust

Observational studies point to heterogeneity in generalized trust within a given 
population. Uslaner (2002), for example, finds that generalized trust is positively 
correlated with education, and that African-Americans report lower levels of trust. Just 
as survey researchers ask what is correlated with generalized trust, experimenters ask 
what factors are correlated with trust and reciprocity between individuals. The work 
augments the trust game by collecting demographic and attitudinal information, or 
nervescience information. We first examine observable individual characteristics and 
then turn to a separate discussion of underlying neural mechanisms.

Individual Characteristics

Trust experiments have examined the relationship between personal 
characteristics, such as gender and ethnicity, and behavior in the games. In a 
comprehensive survey of gender differences in experiments, Croson and Gneezy (2009) 
find considerable variation across 20 trust-game studies. Many find no difference in the 
amount sent, but among the twelve that do, nine find that men trust more than women. 
Among the eight studies finding a difference in trustworthiness, six show women 
reciprocate more. They argue that the cross-study variation is due to women’s greater 
response to subtle differences in the experimental protocols.

Trust may be rooted in other aspects of socio-economic status. While 
experiments with student subjects rarely find an effect of income on behavior, (although 
see Gachter, Hermann and Thoni (2004)), several recent studies use representative 
samples and find positive relationship between income and trust behavior (Bellemare and Kröger 2007; Naef et al. 2009). Age is also related to trust and reciprocity. Bellemare 
and Kröger (2007) find that young and the elderly have lower levels of trust, but higher 
levels of reciprocity than middle aged individuals, a result they attribute to a mismatch 
between expectations about trust and realized trust. Sutter and Kocher (2007) obtain a 
similar finding using six age cohorts ranging from 8 year-old children to 68 year-old 
subjects. They find two clear effects. Trusting behavior is nonlinear with age, with the 
youngest and oldest cohorts trusting the least, while the 22 and 32 year-old cohorts
contributing the most. However, reciprocity is almost linearly related to age, with the oldest cohort returning the most. These age cohort effects are similar to those reported by Uslaner (2002) using survey data.

Several studies examine religion and trust. Anderson and Mellor (2007) report little effect on trust or trustworthiness for religion, no matter the denomination. This is contrary to the findings by Danielson and Holm (2007) who find that churchgoers in Tanzania reciprocate more than their student sample. Johansson-Stenman, Mahmud and Martinsson (2009) match Muslims and Hindus both within religion and across religion, find no difference in any of their matching conditions and cautiously conclude that religious affiliation does not matter. Together these findings show that there are small effects for standard socio-economic status variables on behavioral trust.

Some have conjectured that trust is a risky decision, and that observed heterogeneity in trust may be due in part to variations in risk tolerance. In our own work we directly test this conjecture by supplementing the canonical experiment with several different measures of risk tolerance (2004). These range from survey measures to behavioral gambles with stakes that mirror the trust game. None are correlated with the decision to trust (and not related to reciprocity). By contrast Bohnet and Zeckhauser (2004) focus on the risk of betrayal. They use a simplified trust game with a limited set of choices, and implement a mechanism designed to elicit subjects’ willingness to participate in the trust game with varying probabilities of betrayal, relative to a sure outcome. They find considerable evidence for “betrayal aversion”, with trusters sensitive to the potential actions of the trustee. Trusters are less willing to accept a specified risk of betrayal by trustees than to risk a roll of the dice with the same probability of attaining a high payoff. In a later paper Bohnet et al. (2008) extend this study to six countries and find variation in betrayal aversion across societies. In a recent study incorporating a trust game into a large representative survey in two countries, the US and Germany, Naef et al. (2009) examine the relationship between trust and risk tolerance, measured using a survey question, and find a strong positive relationship between risk tolerance and trust. Whether trust is a risky decision seems to depend importantly on how risk is measured. Clearly this is an area where more work is needed.

Laboratory experiments and observational studies reach similar conclusions. Education and income matter, a point noted by Brehm and Rahn (1997). Generational differences matter, a point that scholars like Putnam (2000) and Inglehart (1997) make as a cultural explanation for trust.

Contributions from Biology and Neuroscience

A promising arena for understanding individual correlates of trust is linked with biological and neurological mechanisms. In principle observational studies are equally capable of focusing on these mechanisms. However, most scholars focusing on such issues have a laboratory experimental bent.

Several research teams focus on the neurological basis of trust (for an overview, see Fehr et al. (2005)). McCabe et al. (2001) find differences in brain activation in the
they speculate what the neural underpinnings might be for trusting behavior. Rilling et al. (2004) examine neural reward systems for a setting similar to the trust game, in which there is the possibility of mutual advantage. Delgado et al. (2005) focus on both reward and learning systems that follow from iterated play with multiple partners in the trust game. King-Casas et al. (2005) also use an iterated trust game and find not only reward and learning processes, but anticipatory signals in the brain that accurately predict when trust will be reciprocated. The neural system they isolate is clearly related to processing social behavior and not simply due to internal rewards. Tomlin et al. (2006) report similar results when subjects are simultaneously scanned in an fMRI while playing the trust game.

Several research groups show that the hormone oxytocin (OT) is an important basis for cementing trust. It is proposed that OT is stimulated by positive interactions with a specific partner. Zak, Kurzban and Matzner (2005) focus on a design to test for changes in OT levels for subjects playing the trust game with another human or playing with a random device. They find elevated levels of OT for trustees assigned to the human condition. Behaviorally they also observe an increase in reciprocation for those in the human condition. There is no difference in OT for trusters, indicating that the effect is absent for the trust decision. By contrast Kosfeld et al. (2005) use a nasal spray to administer either OT or a placebo. They find that OT enhances trust, but it is unrelated to trustworthiness.

There is also evidence for a genetic basis of trust. Cesarini et al. (2008) report on trust experiments conducted with monozygotic (MZ) and dizygotic (DZ) twins in the US and Sweden. While the distribution of trust and trustworthiness is heterogeneous, they find that MZ twins have higher correlations in behavior than their DZ counterparts. The estimated shared genetic effect ranges from 10-20 percent in their samples. As the authors admit, it is not all about genes. A significant component of the variation is explained by the twins’ environments.

The jury is still out concerning the biological and neural mechanisms that drive trust and trustworthiness. Trust and reciprocation involve complex social behaviors and the capacity to test the mechanisms that cause these behaviors remains elusive.

5. Trust and Stereotypes

Individuals vary systematically in their propensities to trust and to reciprocate trust, but another source of heterogeneity in behavior results from differences in the way individuals are treated by others. Those who have studied campaigns (Goldstein and Ridout 2004; Lau and Rovner 2009) or ethnicity and social identity (Green and Seher 2003; McClain et al. 2009) understand how important it is to control for specific pairings of individuals or groups. Voters respond differently when they have information about a candidate – such as gender, race, or age – than when they have abstract information about a candidate. This is partly because beliefs about others are based on stereotypes. Stereotyping is the result of a natural human tendency to categorize. Two possibilities arise. First, the stereotype may accurately reflect average group tendencies, and so
provide a convenient cognitive shortcut for making inferences about behavior. On the other hand, stereotypes can be wrong, reflecting outdated or incorrect information, and so bias decisions in a way that reduces accuracy in making such inferences. If trust leads to accumulating social capital, then decisions based on stereotypes will advantage some groups and disadvantage others, and negative stereotypes may become self-fulfilling prophesies.

We have extended the canonical trust game to study strategic behavior between subjects. While most of the experimental studies have gone through great efforts to ensure that subjects know nothing about one another, we provide visual information to subjects about their partners. In one design we randomly assign dyads and allow counterparts to view one another’s photograph. This enables us to focus on the strategic implications of the joint attributes of players. To eliminate reputation effects, we use subjects at two or more laboratories at different locations. Photographs are taken of each subject and then displayed to their counterparts. For example to study the effect of attractiveness on trust and reciprocity, we look at pairings in which the truster is measured as more (or less) attractive than the trustee. We show that expectations are higher for more attractive trusters and trustees: attractive trusters are expected to send more, and attractive trustees are expected to return more. The attractive truster fails to live up to high expectations; as a consequence the truster is penalized, and less is reciprocated (Wilson and Eckel 2006). In another study (2008) we show that skin shade affects expectations about behavior. Darker skinned trusters are expected to send less, but send more than expected, and they are rewarded for their unexpectedly high trust. The insight we gain is not just from the expectations, but from the response to exceeded or dashed expectations.

Our findings concerning stereotypes are not unusual. For example, trusters prefer to be paired with women, thinking women will be more trustworthy. On average they are a little more trustworthy, but not to the extent expected (see Croson and Gneezy (2009)). Trusters send more to lighter skinned partners, trusting them at higher rates, and beliefs about darker skinned partners are weakly supported (see also (Fershtman and Gneezy 2001; Haile et al. 2006; Simpson et al. 2007; Eckel and Petrie 2009; Naef et al. 2009)). These findings are often masked in observational studies, as subjects display socially-acceptable preferences when it costs them nothing to do so.

In our current work (2009) we introduce another change to the game by allowing subjects to select their partners. We allow trusters view the photographs of potential partners after the trust experiment is explained to them, but before any decisions are made. Subjects rank potential counterparts according to their desirability as a partner, from most desirable to least. To ensure that the ordering task is taken seriously, one truster is randomly drawn and given her first choice, then a second truster is randomly drawn and given his first choice from those remaining, etc. (see Castillo and Petrie (2009)). Not surprisingly, when trusters choose their partners, the overall level of trust is higher. At the same time, when subjects know they have been chosen, they reciprocate at higher rates. Giving subjects some control over the choice of counterpart has a strong positive impact on trust and trustworthiness.
Experiments that focus on the joint characteristics of subjects and that allow for choice among partners are moving toward answering questions about the importance of expectations in strategic behavior, and how expectations are shaped by characteristics of the pairing. Experiments are well suited to answer these questions because of the ability to control information about the pairings of the subjects.

6. Cross Cultural Trust.

An ongoing complaint about experiments is that they lack external validity. The concern is that the behavior of US university students is not related to behavior in the general population, within or across different cultures. Several recent experimental studies tackle the question of external validity by looking at population samples, and considerable work has taken place cross-culturally in recent years.

The impetus to behaviorally measure trust across cultures is partly driven by findings by Knack and Keefer (1997) and Zak and Knack (2001) who find that the level of generalized trust in a country is correlated with economic growth. These findings, derived from surveys and aggregate-level measures, mirror those by Almond and Verba (1963), who provide evidence that trust is correlated with democratic stability. Researchers using trust experiments have entered this arena as well.

Several studies focus on cross cultural comparisons of trust using volunteer student subjects. These studies replicate the high levels of trust found among US students, while finding some variability across cultures (see for example Yamagishi et al. (1998), Buchan et al. (2002) and Ashraf et al. (2006)).

Wilson (with Bahry, 2004) are the first to extend these studies to population samples in two Republics in Russia. This study drew from a large sample of respondents who were given a lengthy face-to-face interview. A subset of subjects was randomly drawn to participate in laboratory-like experiments in the field. While something was gained in terms of confidence in external validity, a price was paid in terms of a loss of control. Sessions were run in remote villages, usually in classrooms or libraries, and the quality and size of the facilities varied (as did the temperature).

Their findings reveal high levels of trust and reciprocity in these Republics, despite the fact that the political institutions are regarded with suspicion. On average 40.8 rubles were sent (51 percent of the truster’s endowment) and trustees returned 48.7 rubles (38.3 percent of what was sent and tripled), a result very close to average behavior among US students. These findings indicate that trust is widespread in an environment in which it is unexpected (for example, see Mishler and Rose (2005)). More importantly Bahry and Wilson (2004) point to strong generational differences across norms that lead to distinct strategic patterns when subjects engage in trust and trustworthiness – a finding that would have been unexplored without a population sample.

Others have also generated new insights when conducting trust experiments outside University laboratories. Barr (2003) finds considerable trust and little effect of ethnicity in Zimbabwean villages. Carpenter, Daniere and Takahashi (2004) use a sample of adult volunteers in Thailand and Vietnam and find that trust is correlated with
formation of social capital measured as owning a home, participating in social group and conversing with neighbors. Karlan (2005) obtains a similar finding in Peru. He notes that trust is related to social capital and trustworthiness predicts the repayment of microcredit loans. Cronk (2007) finds low levels of trust among the Maasai in a neutral (unframed) experimental condition, and even less trust when the decision is framed to imply a long-term obligation. There are numerous other studies that have focused on trust in Bangladesh (Johansson-Stenman et al. 2006), Paraguay (Schechter 2007) Kenya (Greig and Bohnet 2005), the US and Germany (Naef et al. 2009) and across neighborhoods in Zurich (Falk and Zehnder 2007). These studies are important in that they aim at linking the trust game to ethnic conflict, repaying loans and to the risk and patience of individuals.

Studies using culturally different groups have moved beyond student samples and begin to assure critics that concerns with external validity are misplaced. As measured by the trust game, trust and trustworthiness permeates most cultures. These studies are beginning to give us insight into cultural variation. How these studies are linked to key questions of support for democratic institutions or increasing political participation have not routinely been addressed.

7. Trust and Institutions

Political scientists have long been concerned with the relationship between interpersonal trust and political institutions. This tradition extends back to Almond and Verba (1963) who claim that there is a strong correlation between citizen trust and democratic institutions. However, the causal role of institutions in producing citizen trust is in dispute. Rothstein (2000) thoughtfully details the problem, arguing that there are two approaches to understanding how trust among citizens is produced. The first, largely advocated by Putnam (1993), takes a bottom up approach. Trust emerges when citizens participate in many different environments and in doing so experience trust outside their own narrow groups. This in turn provides for democratic stability in that citizens develop tolerance for one another, and that ultimately extends to confidence in governmental institutions. The second approach holds that institutions mitigate the risk inherent in a trust relationship, thereby encouraging individuals to trust one another. As Rothstein puts it, “In a civilized society, institutions of law and order have one particularly important task: to detect and punish people who are ‘traitors’, that is, those who break contracts, steal, murder and do other such non-cooperative acts and therefore should not be trusted.” (pp. 490-1). Institutions, then, serve to monitor dyadic relationships and punish non-cooperative behavior. Of course, this assumes that the institutions charged with monitoring and punishing are perceived as legitimate.

If citizen trust precedes (or is independent of) institutions, then individuals ought to be insensitive to the institutions within which they interact. If the causal relationship is such that institutions are crucial for fostering citizen trust, then the legitimacy of the institutions within which individuals make trust decisions should be directly related to the
degree of trust and trustworthiness. Laboratory experiments are especially well suited to examine these causal relations.

One set of institutional mechanisms allows for punishment by one party of the other in the trust game – *second-party punishment*. Fehr and Rockenbach (2003) allow trustees to specify an amount that should be returned. In one treatment the trustee can choose whether to punish the reciprocator, and the trustee knows both the threshold and whether the punishment option has been selected. They find that the highest level of trustworthiness is observed when sanctions are possible but not implemented by the trustee, and the lowest level is observed when sanctions are implemented. This result is replicated by Houser, et al. (2008), where the threat of punishment backfires by reducing reciprocity when the trustee asks for too much.

Another type of institution introduces an additional player whose role is to punish the behavior of the players in the trust dyad – *third-party punishment*. Bohnet et al. (2001) focus on whether an ex post mechanism that is charged with enforcing trustworthiness in fact enhances trust. In this study, punishment is implemented automatically, using a “robot”. The experiment uses a reduced form (binary) trust game and adds treatments in which the absence of reciprocity can trigger a fine against the trustee. The enforcement mechanism is imperfectly implemented with a known probability. All actors know the associated thresholds that trigger enforcement and the costs of any associated penalties. They find that trust thrives when there is the possibility of sanctions that are not actually invoked. Trust declines when institutional enforcement is imposed. Using different experimental designs, Kollock (1994) and Van Swol (2003) reach similar conclusions.

Rather than using a robot, Charness and Cobo-Reyes (2008) introduce a third party, a subject who is empowered to punish or reward trusters and trustees. Their findings vary with the treatment. Both trusters and trustees show greater pro-social behavior (more is sent and more is returned) when there is the threat of punishment. On the other hand, when punishment is exercised, trust decreases. Interestingly, even though the third party gains nothing from the exchange by the trustee and trustee, that third party is willing to bear the cost of punishing. Banuri et al. (2009), adapt the trust game to study bribery, and show that third-party sanctions can reduce the incidence of and rewards to bribery, essentially by reducing the “trustworthiness” of the bribed official.

A third institutional mechanism that has been brought into the laboratory takes groups of subjects and has them vote on whether to make a trust decision as well as decide how much to return. This mimics a political mechanism in which the representative body decides some level of trust with respect to another group or government. The trustee group decides an appropriate level of reciprocation. Kugler et al. (2007) carry out such an experiment under both individual and a group conditions. In the group treatment five subjects discuss and then decide (via majority rule) how much to send in the trust game. The trustees in the group decide how much to return using a similar format. Generally, trust is depressed in groups, as compared to the averaged decisions of individuals, but reciprocation is equivalent across conditions. Song (2008) finds a similar result, with greater control over the interaction within groups. Trust is depressed when groups make decisions. She also finds lower levels of reciprocity.
Groups, then tend to moderate the optimism of individuals in trusting, and may reduce trustworthiness. Why collective choice mechanisms depress trust is left unexplored.

Finally there has been a good deal of interest in whether information exchange, especially cheap talk, enhances trust. Does a trustee’s unenforceable promise that trust will be reciprocated have any effect on trust? In principle it should not. Charness and Dufwenberg (2006) develop an experimental design allowing trustees an opportunity to send a non-binding message to the truster. The effect of communication leads to increased amounts of trust which is then reciprocated: the promise appears to act as a formal commitment by the trustee. As Charness and Dufwenberg (2006) explain, trustees who make promises appear “guilt averse” and so carry out their action. Ben-Ner et al. (2007) allow both trusters and trustees to communicate. In a control condition no one was allowed to communicate, in one treatment subjects were restricted to numerical proposals about actions, and in the second treatment, subjects could send text plus numerical proposals. Communication enhanced trust and reciprocity, with the more complicated form of communication yielding the highest returns for subjects. Sommerfeld et al. (2008) focus on gossip, which is regarded as the intersection of communication and reputation formation. In their experiment, play is repeated within a group and gossip is allowed. The effect is to increase both trust and reciprocation, largely through reputational enhancement (see also Keser (2003) and Schotter and Sopher (2006)). Communication matters for enhancing trust and trustworthiness, a finding that is widespread in many bargaining games.

An implication of these findings is that trust is malleable. Perversely, it appears that institutional mechanisms involving monitoring and sanctions crowd out trust and trustworthiness. While the “wisdom of groups” should enhance trust, it does not. None of these studies pursue why this might be the case, but instead focus on determining the conditions that enhance or diminish trust. We posit that this work has not accounted for the legitimacy of the institutions that are designed to enhance trust and trustworthiness. This leads us back to the question of what is primary: trust or institutions? Further work needs to be done in the laboratory to figure out the causal relationship.

8. The Unanswered Questions

Experiments on trust have proliferated and added to our stock of knowledge. Despite the extensive research, there are a number of significant questions left unanswered. In this section we detail what we think are fruitful areas for future research.

First, there is still no clear cut answer to the causal relationship between trust and political institutions. Because experimental methods have an advantage in testing for causality, we believe it would be fruitful to tackle this question in the lab. As noted above, some work has examined the relationship between trust and specific institutions, but a more general understanding of the characteristics of institutional mechanisms that promote trust or determining when trust will breed “successful” institutions is paramount.

Second, we do not know enough about why monitoring and sanctioning institutions crowd out trust. If trust is a fragile substitute for institutional monitoring and sanctioning, then knowing which institutions can undermine trust is important. It appears
that punishment itself is not productive, but only the unused possibility of punishment that enhances responsible behavior.

Third, what is the long-term effect of trust for building networks? We do not understand how dyadic trust relationships build communities. Do high-trust individuals build widespread networks in which people are willing to trust, predicting that trust will be reciprocated? Or, do these networks limit the degree to which individuals initiate trust, thereby segregating communities? Our own evidence suggests that subjects use skin shade as a basis for discriminating trust. If this persists over time it is easy to see how segregation can result. In order to understand this dynamic it is important to study trust as a repeated-play game with large numbers of individuals. (In a different setting Ahn et al. (2009) offer tantalizing evidence about network formation in repeated play). A combination of internet experiments and a longer time period provides a vehicle for such studies.

Fourth, we still not know the extent to which the trust game measures behaviors that are important in natural settings. While we see that the trust game is correlated with some individual characteristics, it is unclear how trust and reciprocity is related to issues of concern to political scientists. For example, how well does the trust game predict leadership behavior? How well does it predict political efficacy? How well does it extend to cooperative behavior in other social domains?

Fifth, how much will biological and neural studies provide us insight into a complex social relationship? The multitude of experimental studies point to trust as highly conditional on the context in which the decision is made. Can these biological studies provide additional explanatory power? Their likely contribution will be through a better understanding of the neural mechanisms behind the general heuristic biases common to humans.
References


Ben-Ner, Avner, and Freyr Halldorsson. 2007 (?). "Measuring Trust: Which Measure Can Be Trusted?": University of Minnesota.


Stolle, Dietlind. 1998. "Bowling together, bowling alone: The development of
Sutter, Matthias, and Martin G. Kocher. 2007. "Trust and trustworthiness across different
age groups." Games and Economic Behavior 59 (2):364-82.
Press.
Tomlin, Damon, M. Amin Kayali, Brooks King-Casas, Cedric Anen, Colin F. Camerer,
Steven Quartz, and P. Read Montague. 2006. "Agent-Specific Responses in the
University Press.
Psychology 25 (3):221-33.
Wilson, Rick K., and Catherine C. Eckel. 2006. "Judging a book by its cover: Beauty and
Yamagishi, Toshio, Karen Cook, and Motoki Watabe. 1998. "Uncertainty, Trust, and
Commitment Formation in the United States and Japan." American Journal of
Sociology 104 (1):165-94.
Zak, Paul J., Robert Kurzban, and William T. Matzner. 2005. "Oxytocin is associated