

## Lavish Returns on Cheap Talk: Non-binding Communication in a Trust Experiment<sup>1</sup>

by Avner Ben-Ner<sup>\*</sup>, Louis Putterman<sup>+</sup>, and Ting Ren<sup>\*</sup>

**Abstract.** We let subjects interact with anonymous partners in trust (investment) games with and without one of two kinds of pre-play communication: numerical (tabular) only, and verbal and numerical. We find that either kind of pre-play communication increases trusting, trustworthiness, or both, in inter-subject comparisons, but that the inclusion of verbal communication generates both a larger effect and one that is robust across both inter-subject and intra-subject comparisons. In all conditions, trustors earn more when they invest more of their endowment, trustors and trustees gravitate to “fair and efficient” interactions, and the majority of trustees adhere to their commitments, whether explicit or implicit. Finally, we study trusting and trustworthiness in the sense of adhering to agreements, and we find that both are enhanced when the parties can use words, and especially when an agreement is reached with words and not only with the exchange of numerical proposals.

**Keywords:** trust game, trust, trustworthiness, reciprocity, commitment, communication, cheap talk.

**JEL codes:** C72, C91, D63

---

<sup>1</sup>The research reported here was funded by a grant from the Russell Sage Foundation. We are indebted to Freyr Halldorsson for help in planning and carrying out the experiments, and Bruno Garcia for help in coding chat content.

<sup>\*</sup> Industrial Relations Center, University of Minnesota.

<sup>+</sup> Department of Economics, Brown University.

## Lavish Returns on Cheap Talk: Non-binding Communication in a Trust Experiment

### 1. Introduction

In much of economic life, individuals, firms, and organizations carry out transactions in ways that, because they lack simultaneity and low-cost enforceability, require the trust of one party and the trustworthiness of another. The trust or investment game (Berg, Dickhaut and McCabe, 1995, hereafter BDM) has been a popular tool for the study of these behaviors. In this game one agent, the “trustor” ( $A$ ), can send some or none of an endowment provided by the experimenter to another agent, the “trustee” ( $B$ ), who receives triple the amount sent.  $B$  can then return some or none of what he or she received to  $A$ . Returning money has been interpreted as showing trustworthiness, sending money as showing trust that  $B$  will share the gains.

One aspect of the emergence of trust and trustworthiness that has been largely absent in these experiments is the possibility of communication between the parties. Without prior communication, the status of  $A$ 's sending as an indication of trusting and of  $B$ 's returning as representing a reciprocation of trust depends upon the interpretation of the researcher (Camerer, 2003). If the parties have the opportunity to reach an understanding before the interaction, in contrast, then the status of their subsequent actions as tests of trust and trustworthiness is much clearer. When player  $B$ , for instance, proceeds to diminish her earnings by adhering to a pre-play agreement, the interpretation that  $B$  is being trustworthy is immediate.

In the real world, a look in the eye and a handshake may provide valued assurance even when formal enforcement is unavailable, and a face-to-face meeting may be considered indispensable for sizing up the character of the other party. Whether less intimate channels of communication can serve the same purpose is an interesting question that this paper will help to address.

In other settings, experimentalists have found evidence that pre-play communication, although not binding, can have large effects on behavior. While the trust game literature has grown to include dozens of papers, however, none that we are aware of include opportunities for pre-play communication by both trustor and trustee. In a companion paper (Ben-Ner and Putterman, 2007), we let subjects in some trust games

exchange pre-play messages, then gave them the opportunity to enter into contracts if they reached agreements. In those experiments, we found that pre-play communication increased both trusting (sending by *A*) and trustworthiness (the proportion of money returned by *B*), and that the large majority of interactions were carried out without contracts. Because we can't rule out the possibility that the opportunity to enter a contract influenced those results, however, we conduct and report here otherwise similar experiments in which contracts make no appearance. Our new experiments also allow both between-subject and within-subject comparisons, controlling possible order effects.

On a practical level, our results have bearing on matters of considerable economic importance. As is often remarked and as is demonstrated experimentally in Ben-Ner and Putterman (2007), trust and trustworthiness can serve as cost-saving alternatives to written contracts and enforcement mechanisms. Businesses, independent contractors, government decision-makers, and participants in a wide range of private exchanges invest substantial amounts of time and resources to gauge one another's commitment and trustworthiness through face-to-face, written, telephone, and other forms of communication. The roles played by different aspects of these interactions, such as body language, tone of voice, vocabulary, etc., remain considerably under-explored. Our results suggest that verbal communication increases the likelihood of reaching agreement, raises trusting, and increases trustworthiness, and that the exchange of a few written words, with no visual or auditory contact, can have a substantial effect. Content analysis shows that *A*'s sent more and that *B*'s returned a proportion more closely approximating *A*'s suggestion when the two had verbally agreed on a course of action than when they had only a numerical agreement or no agreement at all. In addition, *B*'s who provided explicit verbal assurances to their partners were on average more trustworthy, implying real self-commitment.

While contributing a new dimension to the literature on communication, our paper also adds to insights about trust and social preferences. Our finding of a high frequency of trustee returning of money to trustors adds to the evidence of preferences and beliefs associated with *reciprocity* (Fehr and Gächter, 2000b). Our subjects' marked gravitation to fair exchanges adds to evidence of fairness preferences. Our results add to the evidence that agents are capable of committing themselves to actions that are payoff-

reducing, and that such abilities to commit are widely believed by others to be common (Charness and Dufwenberg, 2006, Gneezy, 2005). And our findings are suggestive of a considerable framing or presentation effect associated with the decision grid which we introduce to facilitate pre-play communication and decision-making.

The paper proceeds as follows. In Section 2, we briefly discuss trust and communication with references to the recent literature. In Section 3, we give details about our experimental design. Section 4 discusses our findings, and Section 5 concludes the paper.

## 2. Trust and Communication

We say that  $A$  trusts  $B$  when  $A$  chooses to engage with  $B$  in an interaction that has the potential to benefit  $A$ , but that would end up harming  $A$  were  $B$  to respond in a purely self-interested fashion.  $A$  manifests trust by making himself vulnerable to  $B$ 's response in the hope or expectation that  $B$  will act at least in part with  $A$ 's interest in mind.<sup>2</sup>

In a trust or investment game in which partners interact only once, are anonymous, and cannot acquire reputations, the prediction of economic theory assuming rational, payoff-maximizing actors who know other actors to be of this type is unambiguous.  $B$  has no incentive to give money to  $A$ , no matter how generous  $A$  may have been. Because  $B$  will never return any money,  $A$  will never send any. Returning by  $B$ 's violates either rationality or strict self-interest or both. Sending by  $A$ 's need not violate either rationality or self-interest if  $A$ 's have reason to believe that some  $B$ 's are irrational or have a non-selfish preference, for instance reciprocity (Hoffman, McCabe and Smith, 1998; Fehr and Gächter, 2000a; Ben-Ner and Putterman, 2000). In the trust game, we call  $A$  "more trusting" when  $A$  sends a larger part of her available endowment to  $B$ .

We note that to view  $A$ 's sending of money to  $B$  as trusting, in the simple trust game with standard instructions, is to add an interpretive gloss that may or may not be justified. Cox (2004) has pointed out that  $A$ 's trust, in the sense just described, could well

---

<sup>2</sup> "To say ' $A$  trusts  $B$ ' means that  $A$  expects  $B$  will not exploit a vulnerability  $A$  has created for himself by taking the action (James, 2002)." If  $A$  entered the relationship with the sole aim of aiding  $B$ ,  $A$ 's act would be one of altruism, not trust. If  $A$  "trusts"  $B$  because  $A$  knows that  $B$  has (selfish) incentives to do what is in the interest of  $A$ , this also fails to satisfy our definition. For similar definitions, see Bohnet and Zeckhauser, 2004, Eckel and Wilson, 2004, and Ben-Ner and Putterman, 2001.

be confounded with altruism toward  $B$ .<sup>3</sup> How  $B$  views  $A$ 's action is also not unambiguously revealed by  $B$ 's decision. What, if anything,  $B$  returns to  $A$  could, for instance, reflect only  $B$ 's distributive preferences. We use the terms trusting and trustworthy to describe the acts of sending and returning, respectively, but bear these *caveats* in mind. We also note that a pre-play exchange of messages can make the application of the terms trust and trustworthiness more direct, if it results in an explicit framing of the interaction in those terms by the communicators and/or in agreement on a course of action. Thus, in addition to studying how communication influences the amount of money sent by  $A$  and the proportion returned by  $B$ , we also study how communication influences trusting and trustworthiness in the sense of  $A$ 's and  $B$ 's adherence to agreements.

But how, if at all, can communicating before acting be expected to alter trust game outcomes, as opposed to our interpretation of them? In standard economic theory, which assumes strictly self-interested agents, pre-play communication can have no effect. If all agents are rational payoff maximizers and have common knowledge of this, they have no incentive to even send meaningful messages, since it's clear that no message can change  $B$ 's incentive to return nothing, hence  $A$ 's incentive to send nothing. If  $A$ 's are self-interested but believe that their counterparts may have other preferences, however, they might use communication to gauge the type of a specific counterpart, then rationally send positive amounts upon receiving certain kinds of response. Payoff-maximizing  $B$ 's will never return money in one-shot interactions without reputational implications, but a payoff-maximizing  $B$  who believes that his counterpart believes that some  $B$ 's are reciprocators or inequality-averse may attempt to use communication to signal that he is of the relevant type. If such communication raises  $A$ 's expectations that she faces a partner with such preferences, it can affect  $A$ 's sending behavior. Pre-play communication might also raise both sending and returning if partners actually have preferences for adhering to their agreements (disutility from promise-breaking), suffer

---

<sup>3</sup> This is especially the case, in our view, in variants of the trust game in which only  $A$  is provided with an endowment. Examples include Glaeser *et al.* (2000), Malhotra and Murnighan (2002) and Schotter and Sopher (2006).

guilt from letting a partner down,<sup>4</sup> or have dispositions towards reciprocity that are activated by certain social cues.

Ultimately, whether communication increases trust and/or trustworthiness is an empirical question. Because few experiments have permitted communication in the trust game, we briefly consider communication in other experimental settings. A survey focusing mainly on games of coordination is provided by Crawford (1998), who finds that communication sometimes increases efficiency when players' interests are partially aligned. Crawford writes from the standpoint of traditional theory, which assumes that the message "Sender does not find lying personally costly" and "cannot commit himself to tell the truth" (p. 287).<sup>5</sup>

Because our experimental results will be difficult to explain without violating the latter assumption, the growing body of evidence of communication effects in positive-sum games is of particular interest to us. In these games, the conflicting interests of the players cannot be overcome without non-standard preferences such as reciprocity, inequity aversion, or aversion to lying. Public goods experiments, which highlight the potential for conflict between individual and collective interests, are probably the best represented games of this kind in the literature. In such experiments, face-to-face communication has been found to be one of the most powerful ways of increasing cooperation. Isaac and Walker (1988) found that pre-play communication led their experimental subjects to contribute considerably more to a public good, a result strongly reconfirmed by Brosig *et al.* (2003) and Bochet *et al.* (2006). Ostrom, Walker and Gardner (1992) found similar results in common pool resource experiments. Isaac and Walker's is one of 37 papers with 130 experimental treatments whose results Sally (1995) entered in multivariate regressions to study which treatment variables best account for differing levels of cooperation and free riding in public goods games. Sally concluded

---

<sup>4</sup> Charness and Dufwenberg (2006) posit that *B*'s may return money when they promise to do so not because they suffer disutility from lying, but because they anticipate guilt from failing their partners' expectations of them and believe that their protestations of good faith are likely to have raised those expectations.

<sup>5</sup> An accessible survey of theoretical literature on communication is Farrell and Rabin (1996). The authors argue in favor of the common sense assumption that people take their own and others' messages seriously when there is no particular reason not to, judging that "babbling" equilibria are in these circumstances formally possible but "weird." Apart from some footnotes from which we will quote in our conclusion, however, they nevertheless limit their discussion to the assumption of actors incapable of self-commitment.

that face-to-face communication was the single strongest of the treatment variables studied.

A problem with interpreting the impact of face-to-face communication is that it adds a number of potentially separable elements to treatments involving subject anonymity and absence of communication, making it difficult to know what accounts for its influence. When subjects communicate face to face, anonymity is lost, which introduces the possible influence of identity (one learns the counterpart's gender, race, height, etc.). Concerns about possible post-interaction reward or punishment can also arise. Psychological costs may arise as well: a subject may feel greater obligation to trust, or to be trustworthy towards, a concrete other. Face-to-face meetings also make possible verbal communication, in which promises can be delivered. Commitment, sympathy, and other emotions can be conveyed by vocal intonation, facial expression, and body language. Interpersonal attraction, aversion, or bonding may result from physical proximity.

To understand better what lies behind the effects of face-to-face communication, Brosig *et al.* and Bochet *et al.* conducted voluntary contribution mechanism (VCM) experiments in which other forms of communication were in some conditions substituted for face-to-face discussion. Brosig *et al.*'s comparison treatments included a no-communication baseline, a treatment with audio and visual communication from separated compartments, a treatment with only audio communication from separated compartments, and a treatment in which subjects could view one another on video terminals but could not communicate, prior to making their decisions. Bochet *et al.*'s comparison treatments also included one without communication, with the other two being a numerical communication and a chat room communication treatment.<sup>6</sup> Of these treatments, Brosig *et al.*'s audio + video treatment approached the effectiveness of face-

---

<sup>6</sup> Bochet *et al.*'s chat room treatment resembles Frohlich and Oppenheimer's (1998) VCM in which group members could communicate with e-mail messages, while their non-binding numerical communication treatment, which they labeled "numerical cheap talk," resembles the numerical pre-announcement treatment of Wilson and Sell (1997), except that subjects in Bochet *et al.* could react to one another's announcements with new non-binding announcements for a period of a minute or longer before making binding decisions, whereas Wilson and Sell's subjects could send only one announcement before each binding decision. Bochet *et al.* also included comparison treatments in which group members could impose costly monetary punishment on one another, as in Fehr and Gächter (2000a). We ignore these in our discussion.

to-face communication most closely, followed by Bochet *et al.*'s chat room treatment. Bochet *et al.*'s numerical communication treatment, in which subjects sent members of their group non-binding possible contribution choices but no verbal messages, produced roughly the same average behavior as the no-communication baseline but led to a higher dispersion among groups, indicating that it enhanced cooperation in some groups and worsened outcomes in others. Bochet and Putterman (forthcoming) show that most subjects in the treatment responded to one another's signals in a manner consistent with attempting to signal true intentions.

Other experiments in which reciprocity or cooperative intentions seem to interact with inclinations toward truth-telling include Ellingsen, Johannesson, Lilja and Zetterqvist (2006), and Lundquist, Ellingsen, Gribbe and Johannesson (2007). Lunquist *et al.* Study the effect of cheap talk in a bargaining game with one-sided asymmetric information. Comparing a no information treatment with three types of communication, they find that allowing freely formulated messages leads to the fewest lies and the most efficient outcomes. They interpret their results as suggesting "that lying about private information is [subjectively] costly and that the cost of lying increases with the size of the lie and the strength of the promise. Ellingsen *et al.* Study the correlation between subject type and truth-telling by having subjects play in the same pairs first a prisoners' dilemma game and then private values bargaining game with communication. Communication increases efficiency among subjects who cooperated in the prisoners' dilemma, but has no effect for other subjects. In another experiment finding a correlation between cooperativeness and truth-telling, Sanchez-Pages and Vorsatz (2007) study two sender-receiver games in which a sender can tell the truth, lie, or remain silent, and in one of which, the receiver can engage in costly punishment of the sender after learning whether the information was truthful. A correlation between negative reciprocity and truth-telling is found by the fact that "those subjects who punish the sender with a high probability after being deceived are precisely those who send fewer but more truthful messages."

There have been fewer experiments with pre-play communication in trust games, and none in which both parties could exchange proposals or verbal messages. Malhotra and Murnighan (2002) had a computer program posing as trustee sometimes propose a non-binding contract for mutual cooperation, and found that trustors who agreed to such a



contract sent more to their “counterpart.” Glaeser, Laibson, Sheinkman and Soutter (2000) and Buchan, Croson and Johnson (2006) allowed subjects to meet and engage in communication before playing trust games, but these were manipulations of social distance prior to informing subjects of their decision task. Buchan *et al.* permitted no task-relevant communication, while Glaeser *et al.* let some trustees choose to send, or not, a message promising to return at least as much as their partner had sent. Fehr and List (2004), Fehr and Rockenbach (2003), Houser, Xiao, McCabe and Smith (forthcoming) and Rigdon (2005) conducted games in which trustors could suggest amounts to be returned by their trustee counterparts and in some conditions threaten punishment should they not do so. Rigdon’s subjects could reject or accept proposals. In none of the treatments listed is communication fully two-sided, and the papers describing experiments with pre-play suggestions focus mainly on the effects of threatening or not threatening punishment, rather than on the effects of different proposal terms. Charness and Dufwenberg (2006) permit either trustor or trustee, but not both, to send a single message in a binary trust game. They find that the amounts of both trusting and trustworthiness increase significantly when trustees can send messages, but are not influenced by letting trustors send them.

### 3. Experimental Design and Predictions

We conducted an experiment in each session of which a subject, randomly assigned either role *A* or role *B* for all interactions, engages in 10 trust game interactions, each with a different anonymous partner seated in a different building. In each interaction, both *A* and *B* begin with 10 units of experimental currency labeled experimental dollars (*E*\$), *A* selects  $X_A \in (0, 1, \dots, 10)$ , *B* receives  $3X_A$ , and *B* selects  $0 \leq X_B \leq 3X_A$  to return to *A*, yielding payoffs  $(10 - X_A + X_B, 10 + 3X_A - X_B)$ —precisely the moves and payoffs studied by BDM. A slight modification of the BDM procedure is that rather than selecting  $X_B$  from all integers in the relevant range, *B* selects a fraction  $F_B \in (0, 1/6, 1/3, 1/2, 2/3, 5/6, 1)$ , and  $X_B = F_B \cdot 3X_A$ .<sup>7</sup> *A*’s options were displayed as row

---

<sup>7</sup> Thus, as in BDM, the range of *B*’s potential earnings, from 10 to 40, differs from that of *A*, which is from 0 to 30. This differs from some experiments that depart from BDM by permitting *B* to send all or part of his endowment, as well as money received from *A*, and from experiments in which *B* receives no endowment.

headings and  $B$ 's as column headings in a table appearing on the computer screens of both players (Figure 1). Each cell lists the resulting value of  $X_B$ , above, and the resulting final earnings pair  $(Y_A, Y_B)$ , below. Final decisions were made by  $A$  first clicking on a row, which became highlighted on both screens, then  $B$  clicking on a column, which was likewise highlighted. At the end of the session, subjects were paid 10 cents per  $E\$1$  plus a flat \$10 for completing a 30 – 45 minute pre-laboratory sign-up survey and the 30 minute laboratory experiment.<sup>8</sup> Earnings in the laboratory portion averaged \$20.74.

Our focus is on the effect of opportunities to engage in non-binding communication of two possible types: one in which messages can contain numerical (i.e., row and column) information about possible courses of action but no verbal information, such as suggestions, pledges, inquiries, or small talk; and the other in which both verbal and numerical information can be exchanged. Interactions with no pre-play communication were also conducted for purposes of comparison.

More specifically, what we call **simple (S)** interactions have only the two steps (1)  $A$  selects  $X_A$  by clicking on a row of the table, which is thereby highlighted on both subjects' screens, and (2)  $B$  selects  $F_B$  by clicking on a column. In **single proposal (SP)** interactions,  $A$  first clicked-and-highlighted both a row and a column, described in the instructions as a proposal; then  $B$  clicked-and-highlighted either the same or a different row and column, also described as making a proposal, which might or might not agree with  $A$ 's. Subjects were told that these actions had no direct effect on payments.<sup>9</sup> Then  $A$  chose any row and  $B$  chose any column, as in a simple interaction. In **chat plus single proposal (CSP)** interactions,  $A$  and  $B$  could exchange text messages in a private chat room for one minute, then went through the same procedures as single proposal. Messages could not reveal one's identity or contain threats or promises except regarding permissible game moves.<sup>10</sup>

---

<sup>8</sup> Subjects were told that the laboratory portion would take up to one hour. The survey produced data on personal background and characteristics that we plan to use in conjunction with the experiment decisions in other research. The pre-laboratory portion was completed between a week and two days before the lab portion, and included a set of lottery choices that determined a small portion of earnings and are used by us elsewhere as a measure of risk-aversion.

<sup>9</sup> A variant of single proposal with up to three proposals and counter-proposals was used in a related experiment; we retain the single proposal terminology here even though the distinction isn't needed for purposes of this paper.

<sup>10</sup> Subjects were told they would forfeit all earnings if they violated these rules. A review of the chat messages shows no violations.

We study five session designs to investigate the effects of communication both by inter-subject and by intra-subject comparisons, while for purposes of the latter controlling for order effects. In **S-SP**, each subject engaged in five interactions (each with a different partner) with no communication, then five interactions with only numerical proposals. In **SP-S**, this order is reversed, with each subject engaging in five **SP** interactions and then five **S** interactions. In **S-CSP**, five simple interactions are followed by five interactions with chat and single proposal. **CSP-S** reverses this order. We carried out also an **SP-CSP** treatment with five **SP** interactions followed by five **CSP** interactions, but due to resource limitations, we did not also reverse that order. Note that subjects were not informed in advance of the exact number of interactions, nor did they know what condition would follow the initial one, so there should be no “end game” effects,<sup>11</sup> and initial interactions under a given condition (for example, **S** interactions of the **S-SP** and of the **S-CSP** designs) are experimentally identical.

With respect to theoretical predictions, consider first the standard assumptions that subjects are 1) concerned only with their own payoffs, 2) not guided by values and ethical concerns, 3) rational, and 4) assume that their counterparts in the experiment are like them in these respects, serve as a useful analytical benchmark. Under these assumptions and assuming interactions of given pairs are not repeated and that informational conditions make it impossible to invest in reputation, *B* should send no money to *A* regardless of how much he receives, and understanding this *A* should send nothing to *B*. *A* and *B* thus each keep their initial endowments and earn *E*\$10 from an interaction without a contract. Under these assumptions, it also follows that (a) ability to make proposals, no matter how many, (b) concurring on the same proposal, and (c) exchanging text messages, would make no difference.

An alternative to conventional economic theory is provided by behavioral economics. Although neither theoretical nor experimental behavioral economics provide as specific a set of predictions as does conventional theory, they do supply observations that permit characterization of trusting and trustworthiness. These may be listed as: (a) people often display trust and trustworthiness in real-life interactions resembling trust

---

<sup>11</sup> “End game” effects should in any case be ruled out since each interaction has a different partner, but subjects nevertheless react differently in a known last interaction, in some experiments.

games, and large numbers have done so as well in past experimental trust games;<sup>12</sup> (b) communication is helpful in engendering cooperation (as discussed above); (c) reciprocity is often observed (that is, many people act as if obligated to return kindness to someone who acts in a kind or trusting fashion toward them);<sup>13</sup> (d) many people have a preference for keeping their word;<sup>14</sup> (e) there exist individual differences that generate a distribution of types in terms of money-maximizing versus other preferences;<sup>15</sup> and (f) individuals choose their actions to maximize their utility (inclusive of any non-money preferences) subject to beliefs about partners' types, which may be influenced by information about partner characteristics or past actions. Based on these observations, we expect that:

- as in past trust game experiments, there will be considerable amounts of sending and returning even in interactions without pre-play communication;
- the amounts sent and returned will be greater when partners can engage in pre-play communication, and more so if that communication includes written messages, not only because of the way in which this aids coordination, signaling and screening of types, but also because with text messages especially individuals can sometimes build a sense of trust in and responsibility toward their partners, and because the exchange of promises will make many feel bound to keep their word;
- trustors (*A*) will send more when trustees (*B*) have concurred on a proposal, because they will interpret this as an agreement to which *B* may feel some degree of commitment;
- subjects will tend to adhere to agreements, especially ones reached verbally and with explicit promises or assurances;

---

<sup>12</sup> See the summary in Camerer (2003) and the symposium in *Journal of Economic Behavior and Organization*, vol. 55, no. 4, 2004.

<sup>13</sup> References to the literature on reciprocity include Hoffman, McCabe and Smith (1998), Fehr and Gächter, (2000b), Ben-Ner and Putterman (2000) and Gintis, Bowles, Boyd and Fehr, eds. (2005), among many other contributions.

<sup>14</sup> Gneezy (2005), Sánchez-Pagés and Vorsatz (2004), Charness and Dufwenberg (2006).

<sup>15</sup> Kurzban and Houser (2001), Fischbacher, Gächter and Fehr (2001), Page, Putterman and Unel (2005).

- subjects will show special interest in proposals that achieve both efficiency and equity, i.e.,  $A$  sends  $E\$10$  and  $B$  returns  $2/3$  of the amount received (or  $(X_a, X_b) = (E\$10, E\$20)$ ) for payoffs  $(E\$20, E\$20)$ .

Behavioral approaches do not assume that all individuals are equally trusting and trustworthy. Strategic behaviors by more opportunistic individuals attempting to exploit more trusting or trustworthy individuals are entirely possible. One interesting point worth making, in this regard, is that if  $A$  believes that  $B$  is trustworthy and  $A$ 's aim is to maximize her payoff,  $A$  might consider proposing  $X_a = E\$10$ ,  $X_b = E\$25$ , which gives payoffs  $(E\$25, E\$15)$ , making the exchange worthwhile for  $B$  but still moreso for  $A$ . However, even an opportunistic  $A$  might hesitate to suggest this if concerned that, by undermining  $B$ 's "good will" or sense of obligation to reciprocate, the clearly self-interested maneuver would significantly reduce the likelihood that  $B$  will follow through. We therefore conjecture that

- rather than make the proposal which maximizes her prospective earnings assuming trustworthiness of  $B$  and subject to  $B$  receiving some benefit,  $A$  will tend to make the  $(X_a, X_b) = (E\$10, E\$20)$  proposal, to increase the likelihood of her counterpart agreeing and implementing it.

The hypotheses generated by standard and behavioral economics thus disagree on most predictions. Whereas standard theory predicts  $(X_a, X_b) = (0, 0)$  in all interactions, with no effect of communication, behavioral reasoning predicts some positive sending and returning in simple interactions, the presence of efficient and fair  $(E\$10, E\$20)$  proposals and interactions, and positive effects of communication on sending and returning.

#### 4. Results

Nine sessions with twelve to eighteen subjects in each role (the number always matching in a given session) were conducted in computer classrooms at the University of Minnesota, with a total of 264 subjects drawn from the general undergraduate and graduate student body of the university. Table 1 shows the number of sessions and

subjects by treatment. The total number of interactions to be analyzed should ideally be 1320 ( $\frac{1}{2} * 10 * 264$ ) but is slightly smaller due to the fact that computer break-downs left a few interactions uncompleted. We begin by analyzing the first five interactions, when subjects operated under their first and for that portion of the experiment only known mode of interaction.<sup>16</sup>

#### *4.a The pure effect of communication on sending and returning: inter-subject comparison in rounds 1 - 5*

Figure 2 gives a picture of trusting behaviors during the first five rounds of play, when subjects had experienced one condition (**S** or **SP** or **CSP**) only. For each round, Figure 2 has a left bar showing the distribution of amounts sent by subjects in role *A* in the **S** condition, a middle bar showing the corresponding information for *A*'s in **SP** condition, and a right bar showing that distribution for *A*'s in **CSP** condition. The fact that the proportion sending 10 (the lower portion of the bar) is always increasing from **S** to **SP** and from **SP** to **CSP** suggests that trusting was indeed greater with more pre-play communication.

Figure 3 parallels Figure 2 but shows the distribution of the proportion returned by subjects in the *B* role who were sent a positive amount by the person they were paired with. As with Figure 2, the availability of pre-play communication increases from left to right of each set of three bars. It's immediately apparent that the proportion of subjects who returned 67% or more of the money they received—that is, enough to make their counterpart at least as well off as themselves—increases in every period as one goes from **S** to **SP** and from **SP** to **CSP** condition. The prediction that trustworthiness is enhanced by communication thus appears to be supported.

Before verifying these conclusions formally, it's worth remarking that both trust and trustworthiness levels are relatively high in our experiment as compared to many in the literature. For example, in BDM's no history treatment, *A*'s sent an average of \$5.16, with 15.6% sending their entire endowment, and *B*'s who received positive amounts

---

<sup>16</sup> Instructions and sample screens are available for viewing at <https://netfiles.umn.edu/users/benne001/www/papers/Instructions&ScreensAprJun06.pdf>. This document begins with the written instruction sheet that subjects saw first, then moves to on-screen general instructions and instructions for specific types of interactions, and includes illustrations of the screens the subjects in each role saw during the course of an interaction.

returned an average of 29.8% of what they received, with 20% returning nothing, 10% returning half to  $2/3$ , and 16.7% returning  $2/3$  (the amount necessary to give  $A$  and  $B$  equal payoffs) or more. In a more-or-less identical treatment reported by Andreoni (2005), the numbers are similar: average sending of 45% of endowment, and average returning of 26.7% of the amount received, with only 13% of senders receiving back even half of the tripled amount.<sup>17</sup> In the first simple trust interaction in our experiment, in contrast,  $A$ 's sent an average of  $E\$6.36$  or 63.6% of their endowment, with 33.9% sending their entire endowment, and  $B$ 's who received positive amounts returned an average of 43.4%, with 14.8% returning nothing, the same number returning half, and 40.7% returning  $2/3$  or more.<sup>18</sup> Since the subjects in both our own experiment and BDM were students at the University of Minnesota, differences are unlikely to be attributable to the subject pool (unless the different decades matter) and must be due to other factors, such as the use of computers rather than physical passing of dollars in our experiment, the lower stakes per interaction, and the use of the interaction table making transparent the implications of each choice. We consider the high levels of sending and returning in our subject population relative to others again, below, when we consider the high number of “fair and efficient” interactions.

Returning to the differences by condition, consider now the summary information in Table 2 and the statistical tests in Table 3. For the tests, we average the behavior of a given subject  $A$  or  $B$  over the five periods in question and treat it as a single observation so as to avoid the problem of interdependence of observations.<sup>19</sup> In the five rounds considered, subjects had knowledge and experience under one condition only—either **S**, or **SP**, or **CSP**. Since the subjects were not told what condition would follow, the first five interactions in treatments **S-SP** and **S-CSP** occurred under identical conditions and

---

<sup>17</sup> The proportion receiving back  $2/3$  is not reported.

<sup>18</sup> Subjects in Glaeser *et al.*'s trust game send a slightly higher amount and also have high return rates, but their experiment is not easily compared with others because (a) there is no anonymity between trustors and trustees, (b) trustees receive no endowment, increasing the likelihood of confounding trust with altruism, (c) money sent by trustors is doubled, not tripled, and (d) some trustees can make non-binding promises to return at least half. Interestingly, selecting the promise statement didn't change the average proportion returned, but it concentrated the distribution of proportions returned on the one-half share, thus reducing the fraction of trustees leaving their partner worse off.

<sup>19</sup> For subjects in the  $B$  role, we investigate here only decisions made in those cases in which their counterpart sent a positive amount, since  $B$  can return nothing if  $A$  sends nothing. If, for example, a given  $B$  received a positive amount in four of five interactions, the fraction returned is averaged over those four interactions only. As Figure 2 shows, the percentage of interactions in which  $A$  sent 0 is small.

so are pooled for analysis, as are the first five interactions in treatments **SP-S** and **SP-CSP**. The shaded upper portion of Table 2 and the shaded upper left portions of Tables 3a and 3b show that (a) the average and median amount sent by *A*'s was higher in **CSP** than in **SP** interactions and higher in **SP** than in **S** interactions during those five initial interactions, (b) the average and median proportion of received money returned by *B*'s was larger in **CSP** than in **SP** and larger in **SP** than in **S** interactions during the same five interactions, and (c) all of these differences, except for the difference of *A*'s sending in **SP** vs. **S**, are statistically significant in one-tailed Mann-Whitney tests.<sup>20</sup> These inter-subject comparisons of decisions under conditions “uncontaminated” by other forms of interaction represent our first important finding.

We are also interested in the frequency with which subjects reached and kept agreements, to see whether this differs by treatment. Recall that efficiency is greater when *A*'s send more to their partners, regardless of the equality or otherwise of the final outcome, which is determined by their partners' responses. An interesting question is whether verbal communication increases trusting in part because it makes it easier to reach agreement. The data suggest that this was so. When the first interaction was under the **CSP** condition, agreement on a course of action was reached 82% of the time, whereas in the **SP** condition, agreement was reached in the first interaction only 55% of the time, a difference significant at the 1% level. The proportion of communications ending in agreement is higher under **CSP** than under **SP** in periods 2, 3, 4 and 5 as well.<sup>21</sup> For rounds 1 to 5 as a whole, agreement was reached in 60% of **SP** interactions versus 79% of **CSP** interactions, and the difference is significant at the 1% level.

Finally, the benefits of pre-play communication can be viewed in terms of two dimensions. On the one hand, the efficiency of an interaction, increasing as the combined earnings of *A* and *B* go from *E*\$20 to *E*\$40, depends only on *A*'s trusting. On the other hand, the equity of the interaction's outcome, which can be measured by the ratio of earnings of *B* to earnings of *A*, depends mainly on the share that *B* returns, or *B*'s

---

<sup>20</sup> Differences between both *A*'s sending and *B*'s fraction returned in the **CSP** vs. **SP** conditions are significant at the 5% level. The difference between *B*'s fraction returned in the **SP** vs. **S** condition is significant at the 10% level. The differences between both *A*'s sending and *B*'s fraction returned in the **CSP** vs. **S** conditions are significant at the 1% level.

<sup>21</sup> The comparison is between **SP** interactions in the pooled **SP-S** and **SP-CSP** treatments, and **CSP** interactions in the **CSP-S** treatment. The difference is significant at the 1% level in round 3 and at the 10% level in rounds 4 and 5, but is not significant in round 2.



trustworthiness.<sup>22</sup> In Figures 4 and 5, we display for rounds 1 – 5 the average total earnings and the average ratio of earnings, by condition.

Beginning from the left side of Figure 4, we see that already in the **S** condition, combined earnings are over  $E\$35$ , thus achieving more than 75% of the potential efficiency gain from trust. The next two points show average combined earnings in the **SP** condition, first for interactions without numerical agreement on a proposal (denoted w/o **NA**), then for those with agreement (i.e., w/ **NA**)—where **NA** means that the row and column selected by *A* in pre-play communication is likewise selected by *B*. Earnings are substantially higher in **SP** than in **S** if agreement is reached, but are lower if it is not, the *A*'s then being even less confident about sending money than in a game with no communication. In the fourth and fifth points, the same pattern is repeated by **CSP** interactions without and with numerical agreement, but efficiency is somewhat higher for **CSP** than for **SP** interactions when there is a numerical agreement. The final point shows average earnings in **CSP** interactions in which agreement was reached both verbally and numerically (with verbal agreement being denoted by **VA**).<sup>23</sup> Here, efficiency is remarkably close to the 100% level achieved when all trustors send their full endowments.

Turning to the ratio of *B* to *A* earnings, which reflects mainly the return rate or trustworthiness of *B*'s, Figure 5 shows a steady decline as we move toward conditions allowing greater pre-play communication. On average, *B*'s return more of what *A*'s send them when they have exchanged numerical proposals, in the **SP** condition, and more still when they've also exchanged text messages, in the **CSP** condition. Although the share returned by *B*'s is higher in interactions with agreement than in those without, the ratio of *B* to *A* earnings doesn't decline between the **SP** without agreement and the **SP** with

---

<sup>22</sup> More precisely, if *A* sends 0, *B* has no decision to make and the earnings ratio is 1. If *A* sends 1, the ratio can range from 13/9, if *B* returns nothing, to 10/12, if *B* returns everything. And if *A* sends 10, the ratio can range from 40/0  $\rightarrow \infty$ , if *B* returns nothing, to 10/30 = 1/3, if *B* returns everything. Thus, the more *A* sends, the steeper is the curve relating *B*'s to *A*'s earnings, with *B*'s choice set then corresponding to a broader set of earnings ratios.

<sup>23</sup> Note that whether **VA** occurs is judged by analyzing the text of chat exchanges. We say **VA** occurs if there's an identifiable verbal agreement the terms of which match the row and column selections of both *A* and *B* in the numerical communication stage of the same round. Row and column numbers per se need not have been mentioned in the chat exchange; for example, *A* writing "I send 10, you return 20" and *B* writing "o.k." is treated as an agreement on the send 10 row and the return 67% column, if confirmed by matching numerical choices.

agreement points, because observations in which *A* sends nothing are associated with an equal earnings ratio.<sup>24</sup> In contrast to the two points for **SP**, the higher proportion of *A*'s who send all or most of their endowment allows greater trustworthiness by *B*'s to lower earnings disparity for **CSP** interactions with agreement relative to those without agreement. Even those **CSP** interactions in which no agreement was reached have lower average inequality than **S** and **SP** interactions. Finally, the **CSP** interactions with both verbal and numerical agreement display the highest equality of earnings. The general trend of Figures 4 and 5, looked at together, is therefore that both efficiency and equality are increased by pre-play communication, especially when agreement is achieved (as it is more often the better the quality of communication that is available).

#### *4.b More communication effects: inter- and intra-subject comparisons, all rounds*

Figures 6 and 7 parallel Figures 2 and 3 but show the shares of each behavior displayed during rounds 6 – 10 rather than 1 – 5. These figures, along with Table 2, suggest that in rounds 6 – 10 *A*'s again sent more and *B*'s again returned a higher fraction of what they received under the **CSP** condition than under either **SP** or **S**. Tables 3a and 3b confirm that the differences between **CSP** and **SP** and between **CSP** and **S** in rounds 6 – 10 are significant at the 1% level.

In contrast with these first results, the relationships between sending and returning in **S** and **SP** are clearly not the same in rounds 6 – 10 as in rounds 1 – 5. Unlike the earlier rounds, sending and returning are actually a little higher, on average, in **S** than in **SP**, during rounds 6 – 10. But these differences are not significant.

Part of the explanation for why the comparison in rounds 6 – 10 no longer favors **SP** over **S** may lie in the fact that the **S** interactions of rounds 6 – 10 followed five rounds with either **SP** or **CSP** communication, and that these experiences appear to have had lasting effects on trust and trustworthiness. Sending by *A*'s in rounds 6 – 10 of the **SP-S** treatment (average 7.79) and in the same rounds of the **CSP-S** treatment (average 8.09) exceeds that in **S** interactions of the **S-SP** and **S-CSP** treatments (average 7.34), and Table 3 indicates that the difference in the case of **CSP-S** versus **S-SP** and **S-CSP** is

---

<sup>24</sup> This fact also constrains the height of the point for the **S** interactions, in the figure, making the steep decline displayed all the more remarkable.

significant at the 1% level.<sup>25</sup> Much the same is true of proportion returned by *B*, which averages 47.3% in the *S* interactions of the **SP-S** treatment and 48.4% in those of the **CSP-S** treatment versus 42.5% in initial interactions of the **S-SP** and **S-CSP** treatments. According to Table 3b, the difference in proportion returned in *S* interactions of **CSP-S** and in those of **S-SP** and **S-CSP** is significant at the 10% level. It thus appears that subjects interacting without communication were more trusting and trustworthy after experiencing interactions with communication.

However, increased trusting and trustworthiness in simple interactions that follow after interactions with communication is not by itself enough to explain the unexpected performance of *S* versus **SP** in rounds 6 – 10; in particular, it leaves unexplained why subjects in the **SP** condition of the **S-SP** treatment were not themselves more trusting and trustworthy. It appears that those subjects' five rounds of experience under the *S* condition were not as favorable to trusting and trustworthiness as were initial rounds under **SP** and **CSP**, but a complete explanation is impossible on the basis of these treatments alone.<sup>26</sup>

A major reason for including the second block of interactions under different conditions was to permit intra-subject comparisons to complement the inter-subject comparisons of section 4.a. As in the inter-subject comparisons just reported, the data provide the most straightforward conclusions in the case of the comparisons involving *S* and **CSP**. Regardless of order (**S-CSP** or **CSP-S**), *A*'s sent more and *B*'s returned a higher fraction under **CSP** interactions than in ones under *S*. These differences are significant at the 1% level in one-tailed Wilcoxon matched pair (signed-rank) tests, except that when *S* follows **CSP**, the fall in trusting by *A* is not significant—an exception that follows from the “lasting effect” observation of the previous paragraphs.<sup>27</sup> In the treatment permitting intra-subject comparison between **SP** and **CSP**, run in one order only, sending is significantly higher ( $p < .05$ ) in **CSP** but there is no significant difference in proportion returned. In the two treatments involving **SP** and *S*, there is mild support for the proposition that trustworthiness is higher with pre-play communication, in that the

---

<sup>25</sup> See the fourth and fifth cells of the first row of Table 3.

<sup>26</sup> We plan to explore in detail the influence of experience in prior interactions on the behavior of our subjects in another paper.

<sup>27</sup> The tests use individual observations averaged over 5 periods.

proportion returned is higher under **SP** regardless of order; but the difference is significant (at 5%) only when **SP** comes second. As to the behavior of *A*'s, sending is greater in **SP** if it follows **S**, but this difference is not significant. Moreover, the opposite relationship holds when **S** follows **SP**, and in that case the difference is significant at the 5% level. The conclusions based on inter-subject comparison in rounds 1 – 5—that trust and trustworthiness rise as one passes from **S** to **SP** and from **SP** to **CSP**—are thus supported unambiguously in the intra-subject comparisons for **S** versus **CSP**, only.

Our data permit some controlled *inter*-subject comparisons in rounds 6 – 10, and these produce more evidence that trusting and trustworthiness are higher in **CSP** than in the other two treatments. Specifically, behaviors in rounds 6 – 10 of the **S-SP** and **S-CSP** treatments let us compare choices under **SP** versus **CSP** communication following the same prior experience of five interactions without communication. *A*'s send an average of  $E\$9.85$  in **CSP** of **S-CSP** versus  $E\$7.76$  in **SP** of **S-SP**. *B*'s who receive positive amounts return an average of 59.2% in **CSP** of **S-CSP** and an average of 47.6% in **SP** of **S-SP**. Hence, in rounds 6 – 10, as in rounds 1 – 5, both trusting and trustworthiness are greater in **CSP** than in **SP** after identical experience—here, the experience of 5 interactions without communication. Both differences are significant at the 1% level (not shown in Table 3a or 3b).

A second such comparison, this time between **S** and **CSP** behaviors, is offered by the **SP-S** and **SP-CSP** treatments. *A*'s send an average of  $E\$9.19$  in rounds 6 – 10 of **SP-CSP** versus  $E\$7.79$  in rounds 6 – 10 of **SP-S**, and *B*'s return an average of 49.2% in **CSP** interactions of **SP-CSP** versus 47.3% in **S** interactions of **SP-S**. Thus, after five interactions under **SP**, subjects are more trusting and more trustworthy when their next interactions allow chat and numerical proposals (**CSP**) than when they allow no communication (**S**), a difference that is again consistent with the result for interactions 1 – 5 discussed in section 4.a. The difference of *A*'s sending is significant at the 5% level, but the difference in *B*'s returning is not significant (not shown in Table 3a or 3b).

The overall effect of communication in our experiment is summarized visually in Figure 8, which shows the proportion of round 1 – 10 interactions overall, by condition, in which *A*'s sent each possible amount (Figure 8a) and in which *B*'s who received something returned each possible proportion (figure 8b). Mann-Whitney tests find that

sending and proportion returned are significantly higher overall in **CSP** than in **S** and **SP** ( $p < 0.01$  in both cases). Sending is slightly higher but proportion returned is slightly lower overall in **S** than in **SP**, and the differences between these two treatments are not statistically significant.

#### *4.c The effect of communication on fulfillment of agreements*

As we've argued, parting with promised money when an agreement has been made is a clearer demonstration of trust and trustworthiness than sending and receiving in the traditional trust game, which are subject to varying interpretations. In the **SP** and **CSP** conditions of our experiment, in which subjects had the opportunity to reach non-binding pre-play agreements either by exchange of numerical proposals or by exchanging words and numerical proposals, the large majority of subjects lived up to those agreements. In rounds 1 – 5 (6 – 10), *A*'s sent what they agreed to 89% (91%) of the time in **SP** condition and 99% (98%) of the time in **CSP** condition; and given such compliance by their counterpart, *B*'s sent the proportion they had agreed to send 74% (70%) of the time in **SP** and 84% (85%) of the time in **CSP**. For all interactions combined, the relevant figures are 82% in **SP** and 99% in **CSP**, for *A*'s, and 72% in **SP** and 85% in **CSP**, for *B*'s. Our prediction that subjects would adhere to non-binding agreements more often when these were reached by exchange of words is thus supported, with the differences between adherence rates in **SP** versus **CSP** for *A*'s being significant at the 1% level for all rounds, and the difference for *B*'s being significant at the 5% rate for rounds 1 – 5 and at the 1% rate for periods 6 – 10 and 1 – 10.

A major reason why agreements reached with words are more likely to be adhered to than those reached with numbers only is that our protocol for exchanging numerical proposals does not allow subjects to express intention or commitment. Among **CSP** interactions in which agreement on a proposal is indicated in the numerical exchange part of the communication are 272 cases in which the terms were first agreed to verbally then numerically, and 68 cases in which there was no explicit verbal agreement before the numerical one. We find that in 4.4% of cases *A*'s violated an agreement made numerically only, in **CSP**, whereas there was a 0.7% violation rate for agreements that were first made verbally, a difference significant at the 1% level in a 1-tailed Mann-

Whitney test. Given fulfillment by *A*, *B*'s violated an agreement made numerically only in 21.5% of cases, in **CSP**, whereas there was a 13.3% violation rate for agreements that were first made verbally. Thus both trusting and trustworthiness in the sense of adhering to agreements are significantly higher when the agreement was first made verbally, with the difference for *B*'s being significant at the 5% level in a 1-tailed Mann-Whitney test.

Figure 9 displays visually the pattern of violation of agreements in the **SP** and **CSP** interactions. The upper line, graphing the percent of agreements violated by *B*'s, shows the substantially higher violation rate in **SP** interactions with numerical agreement than in **CSP** interactions in which only numerical agreement is achieved, and the still lower violation rate in **CSP** interactions with both verbal and numerical agreement. The lower line shows the corresponding pattern for *A*'s, which differs slightly in that there are slightly more violations by *A*'s in the **CSP** interactions with both types of agreement. But the fraction of violations is in both cases very small, between one half and one percent of agreements only, so the slight reversal is inconsequential.

We defer to the end of the next section a regression analysis of the determinants of violating agreements.

#### *4.d Effects of agreement and assurances in text communication*

Although there is overwhelming evidence in our data that non-binding pre-play communication that included a text component significantly increased the sending and returning of money, determining what about such communication led to these effects is difficult. Some of the effect could be due to a reduction of social distance or to the “rendering real” of the unknown counterpart, effects that could be relatively independent of communication’s specific contents. On the other hand, such outcomes might require a certain amount of “chit chat” or exposure of some facet of personality, and outcomes might also depend on quite specific message contents, such as whether *B* explicitly promises to uphold an agreement, or whether *A* asks *B* for assurances. In this section, we offer a partial investigation of how the specific contents of verbal communication and the action of agreeing to a numerical proposal affected outcomes, keeping in mind that no set of words exchanged should affect behaviors under an assumption of strict rationality,

payoff-maximizing preferences, and common knowledge. We studied the chat log of each **CSP** interaction and coded the following variables:

*A, B* Agree: Did *A* and *B* agree verbally on a specific course of action (row, column)?<sup>28</sup>  
0 = no, 1 = yes

*A* Inquires: Did *A* try to learn something about *B* by asking questions or initiating or attempting to extend non-task-related conversation  
0 = no, 1 = yes, but cursory, 2 = yes, and substantive<sup>29</sup>

*B* Assures: Did *B* offer assurances about his trustworthiness, such as “you have my word,” “you can count on me,” etc.  
0 = no, 1 = yes

Non-Task Talk: Did *A* and *B* communicate about matters unrelated to their decision  
0 = no, 1 = yes

Table 4 shows six GLS regressions, each with subject fixed effects, period and period squared, the four coded variables, and condition/treatment dummy variables.<sup>30</sup> Only interactions under **CSP** condition are included. Regressions (1) and (2), which attempt to explain the amount sent by *A*, differ only by inclusion of a Numerical Agreement (**NA**) control in (2). This dummy variable is 1 if *B*'s *numerical* proposal matched *A*'s, 0 otherwise (recall that *A* knew *B*'s proposal choice before making his binding decision). In column (1), only *A, B* Agree has a significant coefficient, suggesting that after controlling for other factors, *A* sent about E\$1.87 more if *A* and *B* had reached an agreement by way of chat. To see whether agreeing on the course of action verbally has effects over and above those of agreeing numerically, we include both the numerical agreement variable **NA** and the verbal agreement variable *A, B* Agree in

---

<sup>28</sup> The condition *A, B* Agree = 1 is identical to that of presence of a **VA** or verbal agreement in section 4.a and Figures 4 and 5. Note that *A, B* Agree = 0 does not imply that there was *disagreement* between the pair of subjects. They may simply not have discussed specific terms, or they may have agreed on terms that don't match the exchange of numerical messages of that round.

<sup>29</sup> Although this item is inclusive of whether *A* asked *B* for assurances, a matter we mention above, we decided not to limit it because there were very few such explicit requests, but a variety of inquiries about *B* might be interpreted as attempts to probe *B*'s type or “character.”

<sup>30</sup> We code “period” as 1 for the first interaction under a condition, even if that interaction occurs in the 6<sup>th</sup> round of the session (for example, the first CSP round of the S-CSP treatment). In this case, round 7 would be period 2, and so on.

regression (2).<sup>31</sup> The result indicates that each of the two, numerical and verbal agreement, significantly increased *A*'s sending. None of the other message content variables have significant coefficients.

Regressions (3) and (4) investigate the effects of communication content on the proportion returned by *B* assuming *A* sent a positive amount. That amount is included as a control (*A*'s sending), anticipating its strong effect and thus the need to control for it in order to correctly capture other effects. Perhaps surprisingly, *A, B Agree* shows no effect here, the coefficients being negative and insignificant. *NA* is also insignificant, in regression (4), although positive. In both regressions, the amount sent by *A* is a highly significant determinant of not only the amount but also the proportion *B* returns, suggesting that displaying more trust brings out more trustworthiness in the counterpart.<sup>32</sup> In both regressions, there is a positive coefficient, significant at the 10% level, on the variable *B Assures*, indicating that those subjects *B* who offered explicit verbal assurances to their counterparts tended to return about 7% more of the amount they received back to their partners, all else being equal. None of the other message content variables have significant coefficients. The coefficients of the trend terms are significant, implying a small decline in the proportion returned from the first to the third interaction followed by a small recovery in the fourth and fifth interactions. There are significant condition/treatment dummies, indicating lower proportions returned in the two treatments in which *CSP* came in the second block of rounds, after controlling for amount sent.

It is also informative to look at what proportion *B* returned in light of what *A* proposed, especially when *B* may or may not have concurred with the proposal. We do this in regressions (5) and (6), in which the difference between the proportion proposed by *A* (and possibly agreed to by *B*) and the actual proportion returned by *B*, "*B* 'shortfall'," is the dependent variable. In this case, *A, B Agree* has a significant negative

---

<sup>31</sup> We coded "*A, B Agree*" as 1 only if we could verify that what the two subjects appeared to agree on in their verbal messages matched the exchange of numerical proposals that immediately followed. "*A, B Agree*" thus implies *NA*. The reverse doesn't hold, however, because even though *A* and *B* did not reach agreement on a specific course of action in their verbal messages, *B* could nevertheless select the same proposal as had *A* in their numerical exchange.

<sup>32</sup> Being treated as if one possessed an esteemed trait, trustworthiness, may induce an individual to attempt to act accordingly, whereas being treated with suspicion by being sent a small amount only may have the opposite effect. Although Schotter and Sopher (2006) conclude that trust does not lead to trustworthiness, their conclusion is based on inter-temporal comparisons rather than cross-sectional analysis.



coefficient when there is no control for numerical agreement, **NA**, but in regression (6), to which that variable is added, the effect of verbal agreement becomes insignificant and it is **NA** that instead seems to cause *B*'s closer adherence to the proposal.<sup>33</sup> In these regressions, *B* Assures is no longer significant. No other coded message content variable is significant, either, and the significant coefficients on the trend variables indicate that *B* increasingly strayed from the proposed return proportion during periods 1 to 4 before coming slightly closer to the proposal again.

Regressions (5) and (6) suggest that there is something about *B*'s giving assent to *A*'s proposal, whether in words or by the clicking of rows and columns, that is associated with greater adherence to the proposal's terms by *B*. One possibility is that *B*'s preferred choice simply happened to correlate with the proposal and that *B* honestly revealed this in words or clicks. Another is that, in contemplating what to do, a majority of *B*'s took into account that once they had given their assent, they would feel bound to abide by the agreement. On this interpretation, most subjects, when placed in the role of *B*, were capable of self-committing to an agreement, at least one involving stakes of not more than \$2 in real money per interaction. The strong effects of "A, *B* agree" and **NA** in regressions (1) and (2) can also, by this interpretation, be read as signs that most subjects placed in the *A* role tended to believe most *B*'s to be capable of such self-commitment.

We also estimated sets of regressions to investigate the determinants of agreements being violated by *A* or *B* players. Table 5a displays a set of random effects probit regressions for **SP** and **CSP** interactions in which a numerical agreement was reached.<sup>34</sup> The dependent variable is 1 if *A* fails to send the amount specified, otherwise 0. The left block includes only interactions of rounds 1 – 5, the middle block those of rounds 6 -10, and the right block those of all ten rounds. All specifications include trend terms (period and period squared<sup>35</sup>) which obtain no significant coefficients. When the amount that *A* proposed to send is included among the explanatory variables, in the 2<sup>nd</sup> and 4<sup>th</sup> specifications of each block, it obtains a significant negative coefficient with an

---

<sup>33</sup> We also estimated versions of (5) and (6) that include the *A*'s sending variable among the regressors. Results are qualitatively quite similar, with *A*'s sending negative but significant only at the 10% level in the equation corresponding to (5) and insignificant in the version corresponding to (6). The significant coefficient on **CSP** in **SP-CSP** becomes negative and insignificant.

<sup>34</sup> The attempt to estimate these regressions using subject fixed effect terms failed, evidently due to the near absence of variation with respect to violating or not violating agreements for a given subject.

<sup>35</sup> As with previous regressions, round 6 is numbered period 1, and so on.

interesting implication: *A*'s who proposed to send less were more, not less, likely to default on an agreement. Evidently, a low proposal is usually made by a less trusting *A*, and individuals in such a state of mind were more likely than others to “chicken out.” When a dummy variable for **CSP** interactions is added, as in the 3<sup>rd</sup> and 4<sup>th</sup> specifications of each block, it obtains a negative coefficient, usually significant, indicating that pre-play chat made trustors much less likely to default on any agreement reached.

The pattern of violations by *B*'s, which were much more common, is analyzed in Table 5b with essentially the same set of specifications, except that *B*'s proposed column is substituted for *A*'s proposed row. As before, the trend terms do not obtain significant coefficients. The coefficients on column proposal are positive and, for rounds 1 – 10 as a whole, significant: the higher the proportion *B* was to return, the more likely was *B* to default. Finally, the dummy variable for **CSP** interactions is negative and is significant in rounds 6 – 10 and 1 – 10 as a whole, indicating that agreements were significantly more likely to be fulfilled by *B*'s when they were preceded by pre-play chat.<sup>36</sup>

#### *4.e. An attraction to “fair and efficient” exchanges*

We call an interaction or proposal “fair and efficient” (F&E) if it involves the pair of actions  $(X_a, X_b) = (E\$10, E\$20)$ , since this leads to the most equal division of the largest sum of earnings. In those exchanges in which proposals were possible (**SP** and **CSP**), *A*'s proposal was fair and efficient 62.9% of the time in **SP** and 78.8% of the time in **CSP**. Such proposals were agreed to by *B* when made by *A* 76.4% of the time in **SP** and 97.4% of the time in **CSP** (the latter often reflecting verbal agreement in the immediately preceding exchange of numerical messages).<sup>37</sup> In both comparisons, the difference between **SP** and **CSP** is significant at the 1% level in a one-tailed Mann-Whitney test. Once such a proposal was agreed on, *A*'s fulfilled their part by sending *E*\$10 94.2% of the time in **SP** and 98.7% of the time in **CSP**. The counterpart *B*'s of these trusting *A*'s then returned the promised amount 67.1% of the time in **SP**, 85.1% of

---

<sup>36</sup> We also tried to estimate regressions for **CSP** interactions only to see whether pre-play verbal agreement made a difference to the violation rate. But we were unable to obtain meaningful estimates given the regression method and the small number of observations and especially of violations.

<sup>37</sup> This analysis counts as proposals in the **CSP** treatment the row, column pair clicked on by *A* and that clicked on by *B*, not contents of the prior text messages. The method of counting for the **CSP** and **SP** treatments are thus identical.

the time in **CSP**.<sup>38</sup> Both differences are significant at the 1% level in one-tailed Mann-Whitney tests. In the case of **S** interactions, where prior proposals were not possible, F&E interactions were nonetheless carried out 33.0%<sup>39</sup> of the time, versus a similar 33.8% of completed F&E interactions in **SP** and a much larger 68.3% in **CSP**. The difference between the **S** and **SP** conditions is small and insignificant, but the corresponding differences between the **CSP** and **S** and between the **CSP** and **SP** conditions are both significant at the 1% level.

The fact that twice as large a share of interactions meet the F&E criterion when pre-play communication included words as when verbal exchange was unavailable is yet another indication of the power of verbal communication. Also noteworthy is the fact that 85% of those *B*'s who agreed to do so proceeded to voluntarily forfeited the equivalent of \$2.00 in fulfillment of each verbal agreement with an anonymous counterpart to an F&E interaction, which suggests substantial ability to self-commit in the presence of a fairness norm. Finally, the fact that nearly 99% of the *A*'s concerned trusted their partners to do this and in most cases successfully doubled the value of the dollar they sent suggests a widespread belief that such behavior is common, at least among university students in Minnesota.

The proportion of F&E outcomes in our experiment differs dramatically from that seen in other trust experiments that adhere to the BDM design (i.e., two agents begin with equal endowments, *A* can send *B* an amount that is tripled, and *B* can send back any portion of that amount). Of the 32 paired subjects in the original BDM treatment without history, only one interaction (3.1%) involved *A* sending 10 and *B* returning 20.<sup>40</sup> Ortmann, Fitzgerald and Boeing (2000) report one set of 16 one-shot interactions identical to BDM, and find 13 senders of positive amounts, of whom only 2 received back twice or more than what they sent. 1 of the 2 sent 10 and got back 20, for a 6.3% rate of F&E interactions. Cox (2004) reports 32 one-shot BDM-type interactions, in which just 4 of the 26 *A*'s who sent positive amounts (12.5% of all interactions, 15.4% of cases with positive sending) received back enough to earn as much or more than their

---

<sup>38</sup> These percentages include a few cases (three in **SP**, one in **CSP**) in which *B* returned *more* than the promised 67%.

<sup>39</sup> This includes eight cases when *B* returned more than 67%.

<sup>40</sup> In five other cases, in which *A* sent less than 10, *B* returned 2/3 or more of the amount received—thus in only 20% of the cases of positive sending did *A* end up earning as much as or more than *B*.

counterparts. 3 of these 4 were *A*'s who sent 10, so there was a 9.4% rate of F&E interactions, overall. Fehr and List (2004) allowed their trustors to make a non-binding suggestion of the amount to be returned at the time they sent money to a trustee, but only 4 of their 63 undergraduate subjects and 5 of their 38 CEO subjects sent their entire endowment, and only one F&E exchange was completed, involving one of the CEO pairs. Despite permitting trustors and trustees to get acquainted before interacting, Buchan *et al.* found that only one of 88 subjects in 4 different countries received back twice what they sent. One of the few reported cases in which more than a third of *B*'s sent back enough to equalize earnings or make *A*'s better off is that of German subjects in Willinger, Keser, Lohmann and Usunier (2003), of whom 11 out of 29 senders (37.9%) received back at least twice as much as they sent.<sup>41</sup> It seems likely that a higher proportion of subjects in the *B* role sent back 2/3 or more of what they received, in our experiment, because of the clarity regarding this choice's fairness imparted by the information display in our interaction table. Trustees in other experiments who were informed of the same rules but who interacted without a table may have been more likely to think that returning half (rather than two thirds) is fair, failing to consider their endowment. The multiple-interaction set-up of our experiment may also have encouraged experimentation with trusting by *A*'s, since no one decision determined a subject's full earnings. But it remains significant that what they learned through such experimentation was that trusting typically paid off given the behaviors of their counterparts, and that there was no fall off of either trusting or trustworthiness over time.

The pull of the fair and efficient exchange is interesting from another standpoint. As mentioned in section 2, first mover *A* could propose a more advantageous exchange that would still benefit her counterpart,  $(X_a, X_b) = (E\$10, E\$25)$ . However, without an enforceable contract *A* might well worry that *B* would breach the agreement due to its lack of fairness. This may help to explain why such proposals were observed so rarely in our experiment (4.3% of proposals in **SP**, 2.2% of proposals in **CSP**).<sup>42</sup> Further, in only one of the four cases in which *B* agreed to such a proposal and *A* sent the required E\$10

---

<sup>41</sup> The number of F&E interactions in our sense is not indicated, though we're told that 13 of 29 senders of positive amounts sent their whole endowments. Among Willinger *et al.*'s French subjects in the same treatment, 6 out of 29 positive senders received two or more times what they sent.

<sup>42</sup> Fehr and List's (2004) CEO trustors often sent suggestions of returning more than 2/3, but theirs were only suggestions, with no agreement stage in the design.

did *B* return  $E\$25$ , a 25% completion rate that contrasts sharply with the 67% to 85% completion rates for F&E agreements, reported above. F&E proposals also dominated in an otherwise similar experiment (Ben-Ner and Putterman, 2007) in which subjects were given the opportunity to enter into binding contracts at a small cost. In that setting, standard theory clearly predicted that *A* would make and *B* would accept the ( $E\$10$ ,  $E\$25$ ) proposal, which would then be implemented in a binding contract. Yet it was almost never observed, while ( $E\$10$ ,  $E\$20$ ) was carried out in 33.8% of interactions with numerical communication and in 68.3% of those with chat plus numerical communication, proportions quite similar to those of the present experiment. Moreover, both *A* and *B* followed through on ( $E\$10$ ,  $E\$20$ ) proposals significantly more often than they did on others.

## 5. Conclusions

In this paper we report the results of a series of trust experiments in which, in two of three conditions, subjects were able to exchange anonymously either non-binding numerical (tabular) proposals or a minute's worth of written messages followed by numerical proposals, after which they participated in the standard BDM (1995) trust game. We show that in the condition permitting both written and numerical messages (**CSP**), both trusting and trustworthiness are greater than in the experimental condition restricted to numerical communication only (**SP**) and in a no-communication baseline condition (**S**, played via computer using the same interaction table as in the other two conditions). The increase in trusting is substantial: trustors send on average  $E\$9.21$  as compared to  $E\$7.66$  without communication, the standard trust game, and to  $E\$7.71$  with one numerical exchange. Trustworthiness is also greater: trustees in the pre-play communication with numerical exchanges return on average 67% of the money they receive as compared to 50% without communication and 56% with only numerical communication.

Each of our 264 subjects participated in two types of interaction (**S** and **SP**, **S** and **CSP**, or **SP** and **CSP**), each one with a different counterpart seated in a different building, and we confirmed that order had no qualitative effect for the **S/CSP** comparison. While subjects displayed the most trusting and trustworthiness in the **CSP**

condition, both behaviors were also present in the other conditions. Thus, rather than ending up with the *E*\$10 available to each party in the sub-game perfect Nash equilibrium for rational payoff-maximizing subjects, trustors earned an average of *E*\$16.55 in **CSP**, *E*\$13.40 in **S**, and *E*\$14.11 in **SP**, and trustees earned *E*\$23.42 on average in **CSP**, *E*\$26.51 in **S**, and *E*\$25.56 in **SP**.<sup>43</sup>

Of course, the earnings of *B*'s were higher in large part because some *B*'s were untrustworthy. Looking at those cases in the **SP** condition in which pre-play agreements were reached, we find that an average of 10% of the excess of trustee over trustor earnings can be attributed to the slight favoring of *B*'s in contract terms, while the remaining 90% is due to the violation of agreements. In the **CSP** condition, the corresponding proportions are 12.5% due to unequal agreement and 87.5% due to violation, when agreement was numerical only, and 11% due to unequal agreement and 89% due to violation, when the agreement was both verbal and numerical.<sup>44</sup>

Although we reserve most analysis of individual differences for another paper, we make one exception here, using the background data that our subjects provided in the prior sign-up survey to ask which subjects in the *B* role were likely to renege on their agreements?<sup>45</sup> We checked the scores on a much-used psychological measure of opportunism, the Machiavelli scale, of individuals who honored all agreements they entered versus those who failed to honor at least one agreement.<sup>46</sup> Consistent with expectations, the average score was higher, indicating more opportunism, among the defectors, significant at the 5% level.<sup>47</sup> Interestingly, there was no difference between the

---

<sup>43</sup> The differences in both trustor and trustee earnings are not statistically significant for conditions **S** versus **SP**, but are significant at the 1% level for conditions **S** versus **CSP** and **SP** versus **CSP**. Differences between trustor and trustee earnings under each condition are significant at the 1% level.

<sup>44</sup> These figures are derived by first calculating the final ratio of *B* to *A* earnings in those interactions with agreements in the condition in question, then calculating the corresponding ratio of earnings if all agreements had been implemented. Let the first ratio be  $1 + y$  and the second be  $1 + x$ . Then the share of the excess earnings of *B*'s attributable to agreement terms is  $x/y$ , and the share due to violation of agreements is  $(y - x)/y$ .

<sup>45</sup> There are very few violations of agreements by *A* (it is in *A*'s self-interest to keep the agreement if *B* can be trusted, and *A* cannot trap *B* by making an agreement that she will later violate), so we cannot analyze them statistically.

<sup>46</sup> The Mach-IV scale measures the subject's level of agreement with 20 statements. Christie & Geiss (1970) summarize early Machiavellianism research and describe the scale. For an example of its use in economics, see Gunnthorsdottir, McCabe and Smith (2002).

<sup>47</sup> 54 *B*'s made at least two agreements and never reneged, while 15 *B*'s reneged on at least one agreement. If we expand the first group by including *B*'s who never reneged but who made one or more agreements, an additional 9 subjects are added and the significance level of the difference falls to 10%.

two groups in terms of performance on a test of cognitive ability, the Wonderlic Personnel Test.

The most prominent difference between behavior in the **CSP** condition and in the **S** and **SP** conditions is the much higher percentage of interactions in which the trustor sent all his or her endowment to the trustee, and the trustee returned at least two-thirds of the amount he or she received (triple of what the trustor sent). This percentage is 30.0% in **S**, 33.8% in **SP**, and 68.3% in **CSP**. These interactions are “fair and efficient” – trustor and trustee end up with the same payoffs, and they maximize their joint take. From an examination of the record of verbal exchanges it turns out that 68.5% of trustor-trustee pairs made a verbal agreement about how to act in their binding interaction, and also confirmed that agreement in the non-binding numerical exchange stage. The vast majority of these specific agreements (93.38%) are to the “fair and efficient” exchange. “Fair and efficient” agreements are Pareto optimal but do not represent Nash equilibrium. Nonetheless, a remarkable 86.1% of the time subjects in the role of trustee carried out the formally non-binding agreement when their counterpart followed up their agreement by sending *E*\$10 (99.21% of trustors did so). Each time she honored a non-binding agreement with an anonymous partner whose identity she would never know, a trustor was giving up \$2 of a maximum of \$4 she could earn in that one of ten exchanges. Why did she do so, even in the last interaction with yet another randomly chosen counterpart?

An obvious possibility is that most subjects are inclined to keep their word, once given, at least when amounts of the sorts at stake here are on the table. While sending by trustors is consistent with self-interest given observed trustee behaviors, amounts returned by the latter imply that most have either preferences that permit self-commitment (for instance, disutility from lying, breaking their word, or letting down others’ expectations), a tendency towards reciprocity, fairness preferences, or more than one of these. Equally noteworthy is that trustor sending, even if self-interested, implies belief that many people have such preferences. Put differently, subjects in the trustor role work from assumptions about human nature that are at odds with conventional, non-behavioral game theory and earn an extra 34 to 66% in the average transaction, as a result. Because trustors find them trustworthy, trustees also earn more than in Nash equilibrium.

A definitive explanation as to why text but not numerical communication elicits greater trusting and trustworthiness in our subjects is impossible to obtain without further research, but a reasonable conjecture is possible here as well. Our numerical proposal treatment gave subjects no chance to declare an interpretation of the numbers they sent – i.e., whether they were merely a proposal, a possibility being considered, or a behavior to which they would commit themselves. Text communication let subjects declare commitment, which was generally convincing to their counterparts and most often “self-committing.” The exchange may also have reassured subjects that their counterpart was a real person, not a programmed computer, and it may have imparted not just intellectual but also psycho-social “realness” to the counterpart, adding force to self-commitment. Our regressions showing that an express verbal pre-play agreement increased sending by trustors and adherence to the proposal by trustees, and that explicit expressions of assurance by trustees are associated with their returning a larger proportion, are consistent with these interpretations.

In a review of the literature on communication that the authors intentionally limited to the conventional approach, Farrell and Rabin (1996) wrote that “[f]or the purposes of this work, we’ll model ruthless economic [A] who tells the truth only whenever she finds it pays, and we’ll suppose that [B] expects that.” They admit, however, that “[p]eople in reality do not seem to lie as much, or question each other’s statements as much, as game theory suggests they should.” Referring to an experiment in which “subjects outperformed the ... theoretical upper bound on gains from trade with private information” they commented “We interpret this, as well as the fact that people typically say what they want to have been believed even when the incentives clearly imply that cheap talk should not be believed, as *suggesting that some people tell the truth despite incentives to lie* (emphasis ours).” It may have been an indication of the times that those statements appeared only in footnotes. We believe that our results justify at least sometimes placing them in the text itself.



## References

- Andreoni, James, 2005, "Trust, Reciprocity, and Contract Enforcement: Experiments on Satisfaction Guaranteed," Working Paper, University of Wisconsin.
- Ben-Ner, Avner and Louis Putterman, 2000, "On Some Implications of Evolutionary Psychology for the Study of Preferences and Institutions," *Journal of Economic Behavior and Organization* 43: 91-99, 2000.
- Ben-Ner, Avner and Louis Putterman, 2001, "Trusting and Trustworthiness," *Boston University Law Review* 81 (3): 523-551.
- Ben-Ner, Avner and Louis Putterman, 2007, "Trust, Communication and Contracts: An Experiment," Working Paper.
- Berg, J., John Dickhaut, and Kevin McCabe, 1995, "Trust, Reciprocity and Social History," *Games and Economic Behavior* 10: 122-42.
- Bochet, Olivier, Talbot Page and Louis Putterman, 2006, "Communication and Punishment in Voluntary Contribution Experiments," *Journal of Economic Behavior and Organization* 60: 11-26.
- Bochet, Olivier and Louis Putterman, 2007, "Not Just Babble: A Voluntary Contribution Experiment with Iterative Numerical Messages," unpublished paper, Brown University and University of Maastricht.
- Bohnet, Iris, Richard Zeckhauser, 2004, "Trust, Risk and Betrayal," *Journal of Economic Behavior and Organization* 55 (4): 467-84.
- Brosig, J., Ockenfels, A., Weimann, J., 2003. .The effect of communication media on cooperation. *German Economic Review* 4, 217-42.
- Buchan, Nancy, Croson, Rachel, Johnson, Eric, 2006, "Let's get personal: An international examination of the influence of communication, culture and social distance on other regarding preferences." *Journal of Economic Behavior and Organization* 60: 373-98.
- Camerer, Colin, 2003, *Behavioral Game Theory*. Princeton, New Jersey: Princeton University Press.
- Charness, Gary and Martin Dufwenberg, 2006, "Promises and Partnerships," *Econometrica* 74 (6): 1579-1601.
- Christie, Richard and Florence Geiss, 1970, *Studies in Machiavellianism*, New York: Academic Press.

- Cox, James, 2004, "How to identify trust and reciprocity," *Games and Economic Behavior* 46: 260-81.
- Crawford, Vincent, 1998, "A Survey of Experiments on Communication via Cheap Talk," *Journal of Economic Theory* 78: 286-98.
- Eckel, Catherine and Rick Wilson, 2004, "Is Trust a Risky Decision?" *Journal of Economic Behavior and Organization* 55 (4): 447-65.
- Ellingsen, Tore, Magnus Johannesson, Jannie Lilja and Henrik Zetterqvist, 2006, "Trust and Truth," Department of Economics, Stockholm School of Economics.
- Farrell, Joseph and Matthew Rabin, 1996, "Cheap Talk," *Journal of Economic Perspectives* 10 (3): 103-18.
- Fehr, Ernst and Simon Gächter, 2000a, "Cooperation and Punishment," *American Economic Review* 90: 980-94.
- Fehr, Ernst and Simon Gächter, 2000b, "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives* 14 (3): 159-81.
- Fehr, Ernst and John List, 2004, "The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among CEOs," *Journal of the European Economic Association* 2(5): 743-71.
- Fehr, E. and Bettina Rockenbach, 2003, "Detrimental Effects of Sanctions on Human Altruism," *Nature* 422: 137-40.
- Fischbacher, Urs, Simon Gächter and Ernst Fehr, 2001, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment," *Economics Letters* 71: 397-404.
- Frohlich, Norman and Joe Oppenheimer, 1998, "Some Consequences of e-mail vs. Face-to-Face Communication in Experiment," *Journal of Economic Behavior and Organization* 35 (3): 389-403.
- Gintis, Herbert, Samuel Bowles, Robert Boyd and Ernst Fehr, eds., 2005, *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge, MA: MIT Press.
- Glaeser, Edward, David Laibson, José Scheinkman and Christine Soutter, 2000. "Measuring Trust." *The Quarterly Journal of Economics* 65, 811-846.
- Gneezy, Uri, 2005, "Deception: The Role of Consequences," *American Economic Review* 95: 384-294.

- Gunnthursdottir, Anna, Kevin McCabe and Vernon Smith, 2002, "Using the Machiavellian Instrument to Predict Trustworthiness in a Bargaining Game," *Journal of Economic Psychology* 23: 49-66.
- Hoffman, Elizabeth, Kevin McCabe and Vernon Smith, 1998, "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," *Economic Inquiry* 36: 335-52.
- Houser, Daniel, Erte Xiao, Kevin McCabe and Vernon Smith, (forthcoming) "When Punishment Fails: Research on Sanctions, Intentions and Non-Cooperation," *Games and Economic Behavior* (in press).
- Isaac, R. Mark and James M. Walker, 1988, "Communication and Free-Riding Behavior: The Voluntary Contributions Mechanism," *Economic Inquiry* 26: 585-608.
- James, Harvey S., 2002, "The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness," *Journal of Economic Behavior and Organization* 47: 291-307.
- Kurzban, Robert and Daniel Houser, 2001, "Individual Differences in Cooperation in a Circular Public Goods Game," *European Journal of Personality* 15 (S1): S37-S52.
- Lundquist, Tobias, Tore Ellingsen, Erik Gribbe and Magnus Johannesson, 2007, "The Cost of Lying," Department of Economics, Stockholm School of Economics.
- Malhotra, Deepak and J. Keith Murnighan, 2002, "The Effects of Contracts on Interpersonal Trust," *Administrative Science Quarterly* 47 (3): 534-59.
- Ortmann, Andreas, John Fitzgerald and Carl Boeing, "Trust, Reciprocity, and Social History: A Re-examination," *Experimental Economics* 3 (1): 81-100.
- Ostrom, Elinor, James Walker and Roy Gardner, 1992, "Covenants with and without a Sword: Self Governance is Possible." *American Political Science Review*. 86 (2): 404-416.
- Page, Talbot, Louis Putterman and Bulent Unel, "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency," *Economic Journal* 115: 1032-1053.
- Rigdon, Mary, "Trust and Reciprocity in Incentive Contracting," U. of Michigan, 2005.
- Sally, David, 1995, "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992," *Rationality and Society* 7 (1): 58-92.
- Sanchez-Pages, Santiago and Marc Vorsatz, 2007, "Enjoy the Silence: An Experiment on Truth Telling," Edinburgh School of Economics Discussion Paper 155, University of Edinburgh.

Schotter, Andrew and Barry Sopher, 2006, "Trust and Trustworthiness in Games: An Experimental Study of Intergenerational Advice," *Experimental Economics* 9: 123-45.

Willinger, Marc, Claudia Keser, Christopher Lohmann, Jean-Claude Usunier, 2003, "A Comparison of Trust and Reciprocity between France and Germany: Experimental Investigation Based on the Investment Game," *Journal of Economic Psychology* 24 (4): 447-66.

Wilson, Rick and Jane Sell, 1997, "'Liar, Liar...': Cheap Talk and Reputation in Repeated Public Goods Settings," *Journal of Conflict Resolution* 41 (5): 695-717.

Figure 1. Interaction table

Subject Room A - Microsoft Internet Explorer

Subject: test201

A Sends to B Role: A

A sends B receives	B sends to A						
	0%	17% (1/6)	33% (1/3)	50% (1/2)	67% (2/3)	83% (5/6)	100%
<i>E\$0</i> → <i>E\$0</i>	<i>E\$0</i> [10,10]	<i>E\$0</i> [10,10]	<i>E\$0</i> [10,10]	<i>E\$0</i> [10,10]	<i>E\$0</i> [10,10]	<i>E\$0</i> [10,10]	<i>E\$0</i> [10,10]
<i>E\$1</i> → <i>E\$3</i>	<i>E\$0</i> [9,13]	<i>E\$0.5</i> [9.5,12.5]	<i>E\$1</i> [10,12]	<i>E\$1.5</i> [10.5,11.5]	<i>E\$2</i> [11,11]	<i>E\$2.5</i> [11.5,10.5]	<i>E\$3</i> [12,10]
<i>E\$2</i> → <i>E\$6</i>	<i>E\$0</i> [8,16]	<i>E\$1</i> [9,15]	<i>E\$2</i> [10,14]	<i>E\$3</i> [11,13]	<i>E\$4</i> [12,12]	<i>E\$5</i> [13,11]	<i>E\$6</i> [14,10]
<i>E\$3</i> → <i>E\$9</i>	<i>E\$0</i> [7,19]	<i>E\$1.5</i> [8.5,17.5]	<i>E\$3</i> [10,16]	<i>E\$4.5</i> [11.5,14.5]	<i>E\$6</i> [13,13]	<i>E\$7.5</i> [14.5,11.5]	<i>E\$9</i> [16,10]
<i>E\$4</i> → <i>E\$12</i>	<i>E\$0</i> [6,22]	<i>E\$2</i> [8,20]	<i>E\$4</i> [10,18]	<i>E\$6</i> [12,16]	<i>E\$8</i> [14,14]	<i>E\$10</i> [16,12]	<i>E\$12</i> [18,10]
<i>E\$5</i> → <i>E\$15</i>	<i>E\$0</i> [5,25]	<i>E\$2.5</i> [7.5,22.5]	<i>E\$5</i> [10,20]	<i>E\$7.5</i> [12.5,17.5]	<i>E\$10</i> [15,15]	<i>E\$12.5</i> [17.5,12.5]	<i>E\$15</i> [20,10]
<i>E\$6</i> → <i>E\$18</i>	<i>E\$0</i> [4,28]	<i>E\$3</i> [7,25]	<i>E\$6</i> [10,22]	<i>E\$9</i> [13,19]	<i>E\$12</i> [16,16]	<i>E\$15</i> [19,13]	<i>E\$18</i> [22,10]
<i>E\$7</i> → <i>E\$21</i>	<i>E\$0</i> [3,31]	<i>E\$3.5</i> [6.5,27.5]	<i>E\$7</i> [10,24]	<i>E\$10.5</i> [13.5,20.5]	<i>E\$14</i> [17,17]	<i>E\$17.5</i> [20.5,13.5]	<i>E\$21</i> [24,10]
<i>E\$8</i> → <i>E\$24</i>	<i>E\$0</i> [2,34]	<i>E\$4</i> [6,30]	<i>E\$8</i> [10,26]	<i>E\$12</i> [14,22]	<i>E\$16</i> [18,18]	<i>E\$20</i> [22,14]	<i>E\$24</i> [26,10]
<i>E\$9</i> → <i>E\$27</i>	<i>E\$0</i> [1,37]	<i>E\$4.5</i> [5.5,32.5]	<i>E\$9</i> [10,28]	<i>E\$13.5</i> [14.5,23.5]	<i>E\$18</i> [19,19]	<i>E\$22.5</i> [23.5,14.5]	<i>E\$27</i> [28,10]
<i>E\$10</i> → <i>E\$30</i>	<i>E\$0</i> [0,40]	<i>E\$5</i> [5,35]	<i>E\$10</i> [10,30]	<i>E\$15</i> [15,25]	<i>E\$20</i> [20,20]	<i>E\$25</i> [25,15]	<i>E\$30</i> [30,10]

You are in the role of A. Make your choice of an amount to send to B by clicking on a row heading then pressing SUBMIT.

Figure 2. A's Sending, by round and condition, rounds 1-5

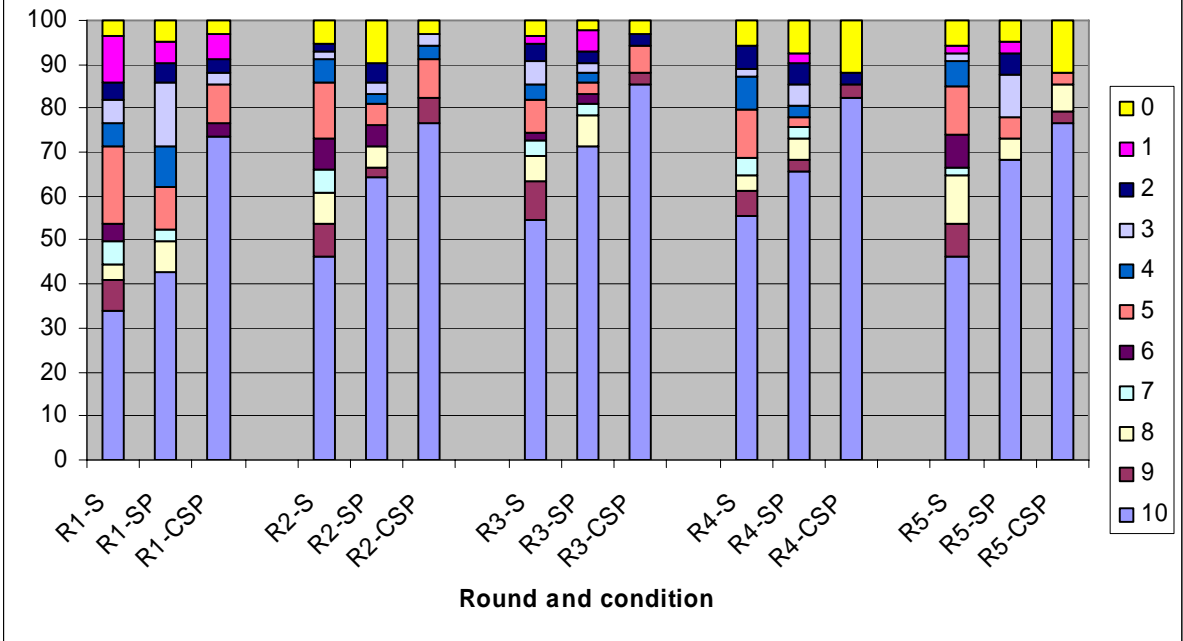
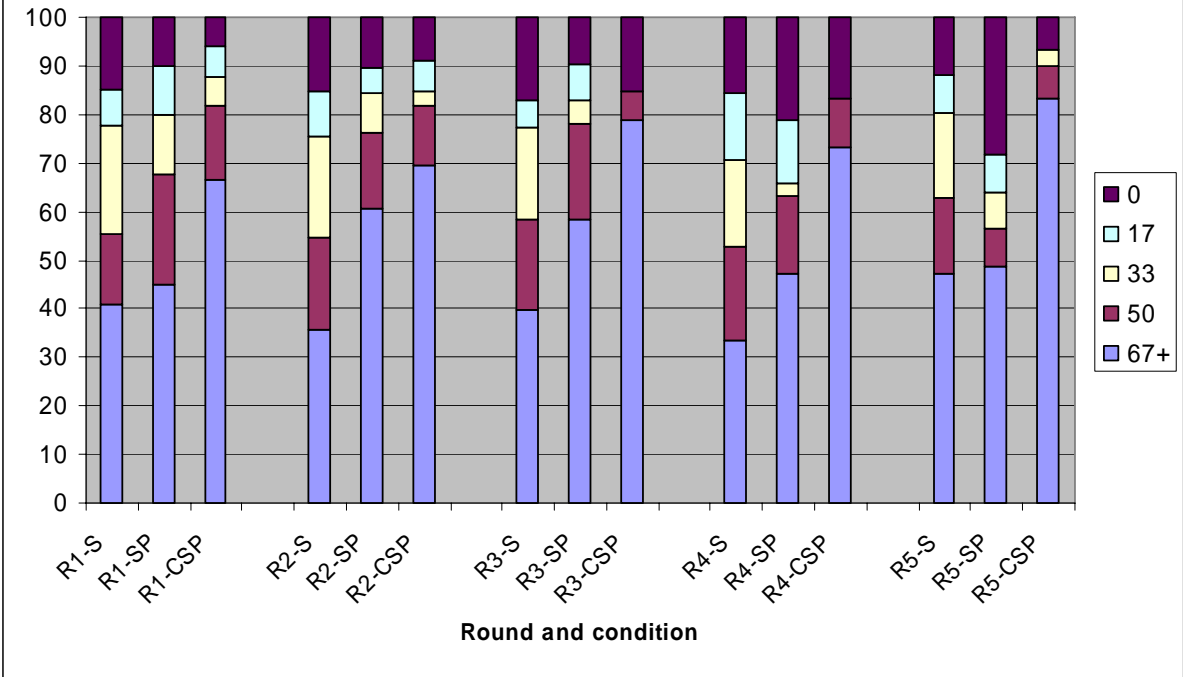
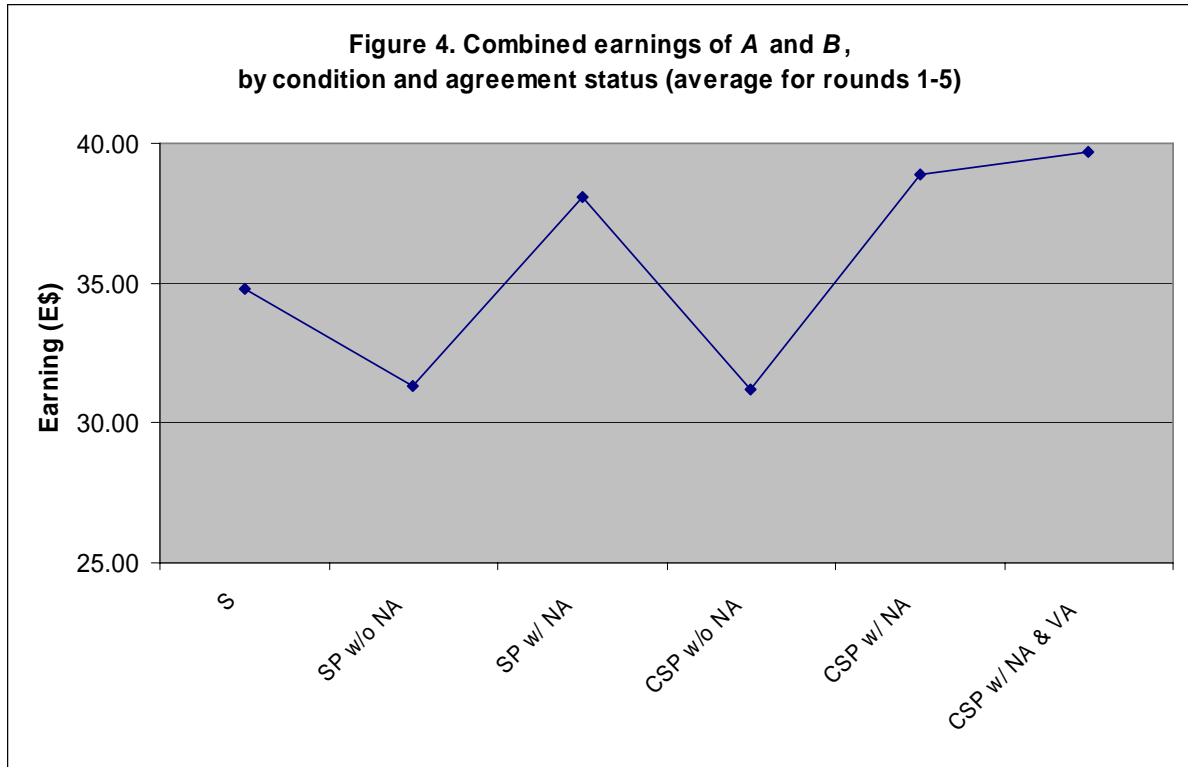


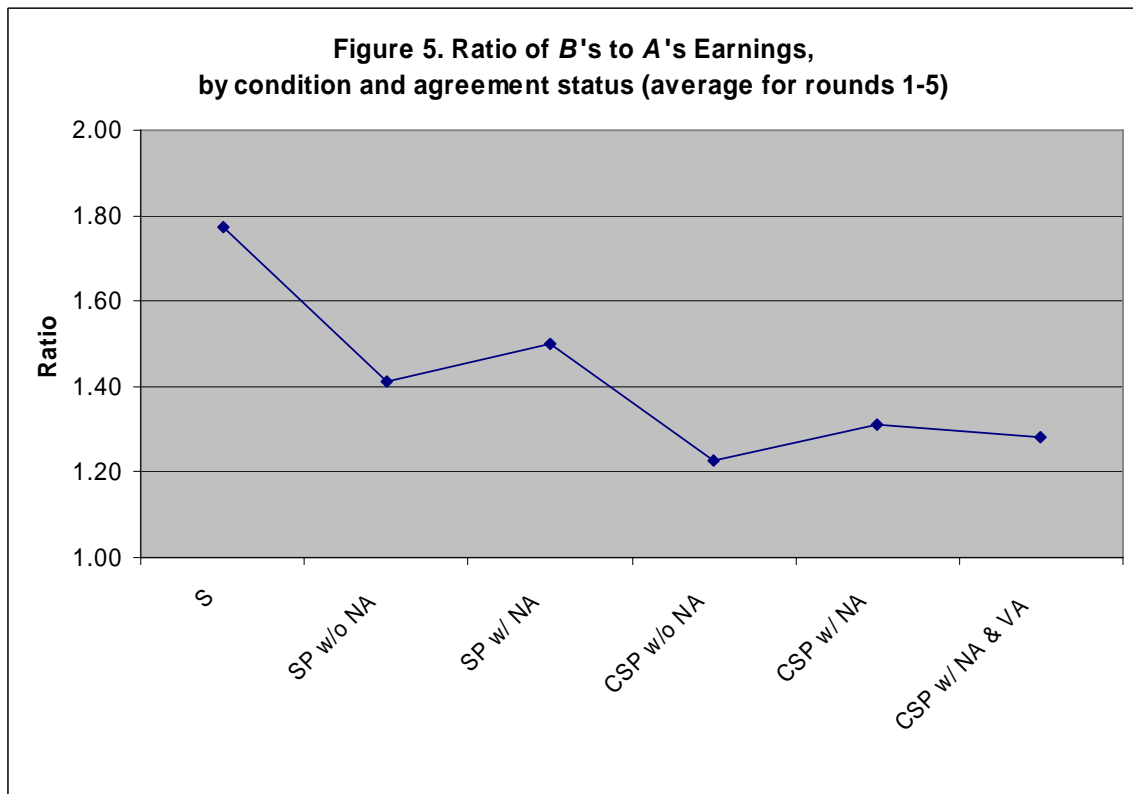
Figure 3. B's Return %, by round and condition, rounds 1-5



Note: B's Return % reflects only instances when A sent more than zero.



Note: NA = numerical agreement, VA = verbal agreement.



Note: NA = numerical agreement, VA = verbal agreement.

Figure 6. A's Sending, by round and condition, rounds 6-10

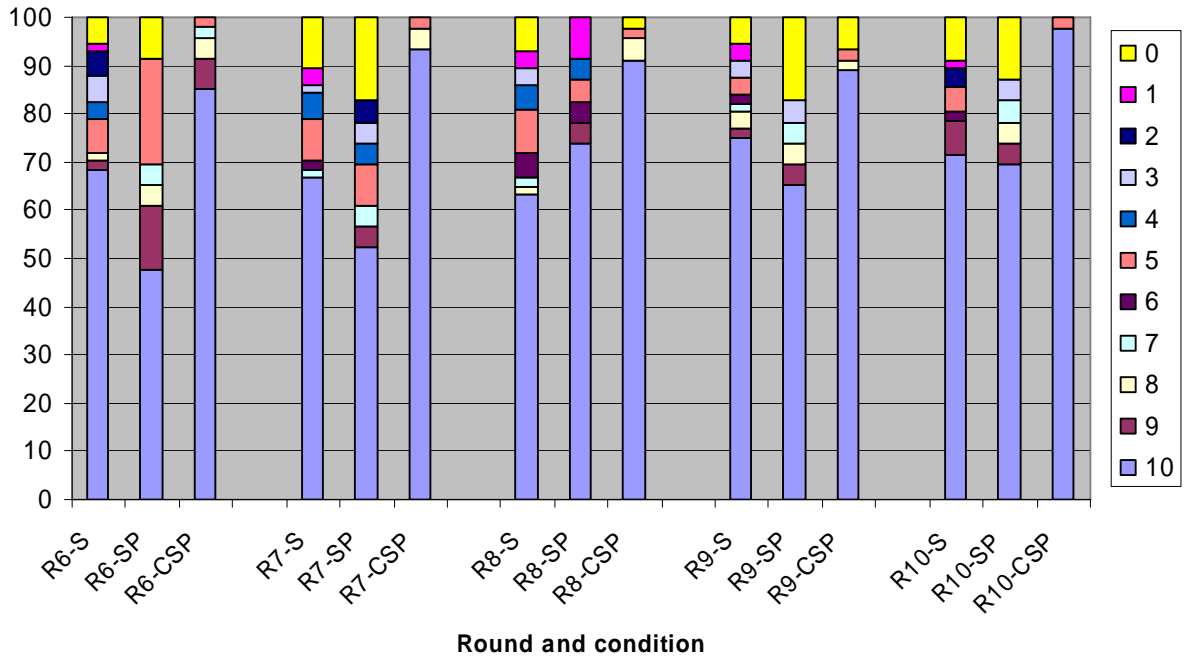
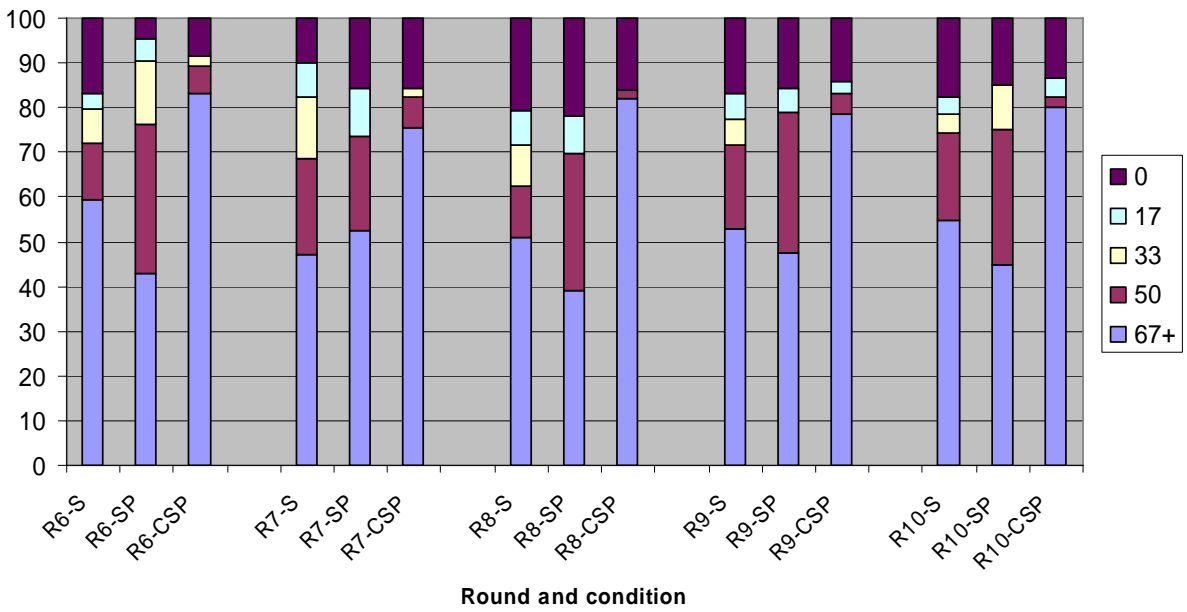
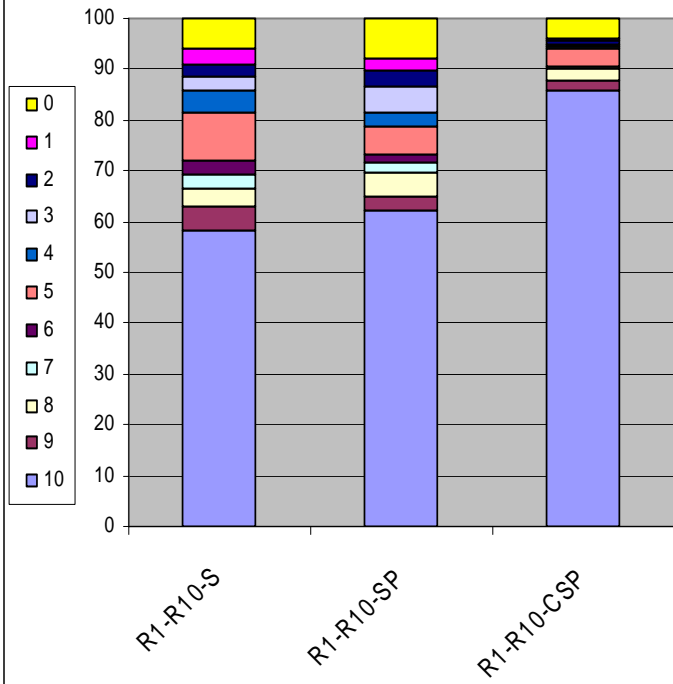


Figure 7. B's Return %, by round and condition, rounds 6-10

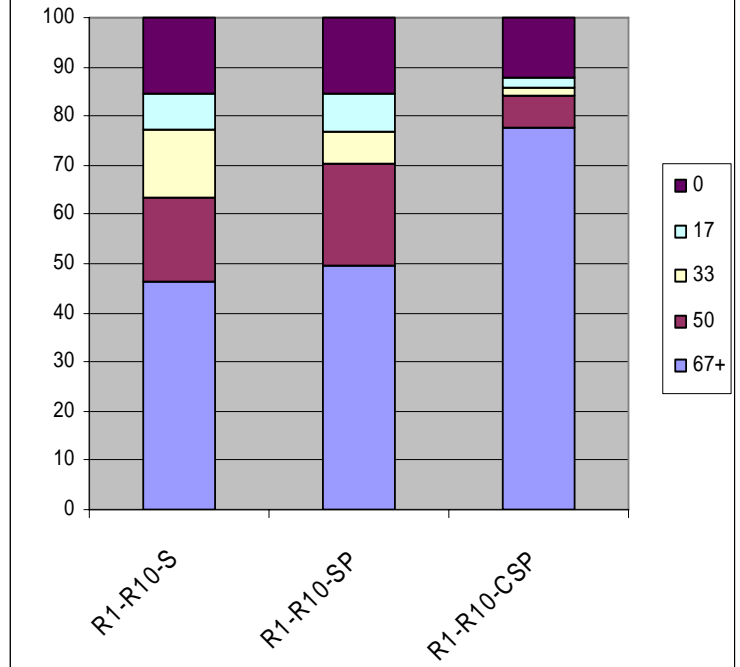




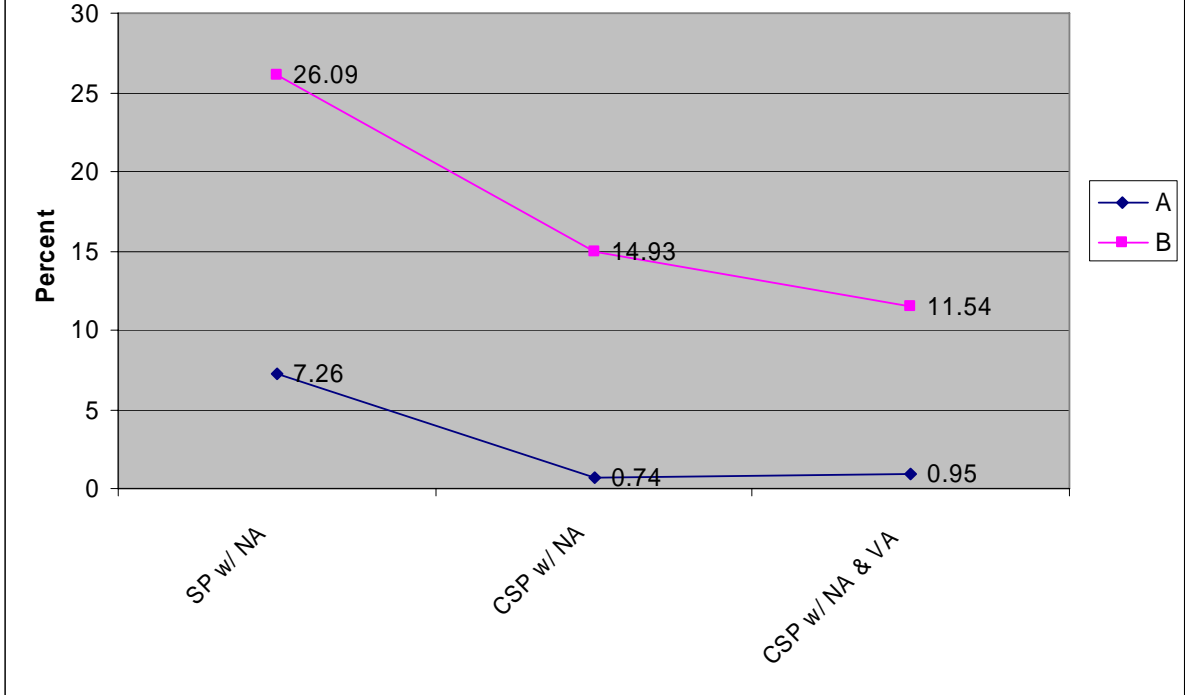
**Figure 8a. A's Sending by Condition, all rounds**



**Figure 8b. B's Return % by Condition, all rounds**



**Figure 9. Percentage of agreement violated, by condition and agreement type (average for round 1-5)**



Note: NA = numerical agreement, VA = verbal agreement.

**Table 1. Summary of treatments, sessions, and interactions**

<b>Design</b>	<b>Description</b>	<b>No. of Sessions</b>	<b>Total Participants (A &amp; B)</b>	<b>Number of Interactions</b>
<b>S-SP</b>	5 simple interactions, 5 interactions with one numerical proposal per subject	2	48	232
<b>SP-S</b>	5 interactions with one numerical proposal per subject, 5 simple interactions	2	50	241
<b>S-CSP</b>	5 simple interactions, 5 interactions with chat plus one numerical proposal per subject	2	64	309
<b>CSP-S</b>	5 interactions with chat plus one numerical proposal per subject, 5 simple interactions	2	68	335
<b>SP-CSP</b>	5 numerical proposal interactions, 5 with chat plus one numerical proposal per subject	1	34	161

Notes: 1. Anonymous interactions. 2. No two subjects interacted twice.

**Table 2. Descriptive statistics of amount sent and % returned, by round and condition**

		<b>A's Sending</b>		<b>B Return %*</b>		
		Mean (Std. Dev.)	Median	Mean (Std. Dev.)	Median	
<b>R1-R5</b>	<b>S</b>	7.34 (2.53)	7.9	42.50 (21.14)	46.6	
	<b>SP</b>	7.68 (2.70)	8.27	50.05 (20.30)	55.2	
	<b>CSP</b>	8.65 (2.12)	10	56.75 (19.88)	63.6	
<b>R6-R10</b>	<b>S</b>	<b>All</b>	7.96 (2.83)	10	47.96 (21.18)	54.63
		<b>SP-S</b>	7.79 (2.27)	8	47.34 (22.72)	50
		<b>CSP-S</b>	8.09 (3.21)	10	48.38 (20.42)	56.23
	<b>SP</b>	<b>S-SP</b>	7.76 (2.02)	7	47.58 (20.26)	55.67
	<b>CSP</b>	<b>All</b>	9.62 (0.87)	10	55.81 (19.07)	67
		<b>S-CSP</b>	9.85 (0.42)	10	59.2 (15.38)	67
		<b>SP-CSP</b>	9.19 (1.30)	10	49.24 (23.93)	57.75
	<b>R1-R10</b>	<b>S</b>	7.66 (2.69)	8.6	45.23 (21.24)	50
<b>SP</b>		7.71 (2.47)	8	49.20 (20.15)	55.67	
<b>CSP</b>		9.21 (1.59)	10	56.20 (19.29)	67	

\* Excluding the cases when A sent zero.

**Table 3a. Mann-Whitney Tests of Amount Sent by A**

		R1-R5			R6-R10				
		S (S-SP & S-CSP)	SP (SP-S & SP-CSP)	CSP (CSP-S)	S (SP-S)	S (CSP-S)	SP (S-SP)	CSP (S-CSP)	CSP (SP-CSP)
<b>R1-R5</b>	<b>S (S-SP &amp; S-CSP)</b>	--	S < SP (NS)	S < CSP (0.01)	S < S (NS)	S < S (0.01)	S < SP (NS)	S < CSP (0.01)	S < CSP (0.01)
	<b>SP (SP-S &amp; SP-CSP)</b>	--	--	SP < CSP (0.05)	SP < S (NS)	SP < S (0.05)	SP > SP (NS)	SP < CSP (0.01)	SP < CSP (0.05)
	<b>CSP (CSP-S)</b>	--	--	--	--	CSP > S (NS)	CSP > S (NS)	CSP < CSP (0.01)	CSP < CSP (NS)
<b>R6-R10</b>	<b>S (SP-S)</b>	--	--	--	--	S(SP) < S(CSP) (NS)	S > SP (NS)	S < CSP (0.01)	
	<b>S (CSP-S)</b>	--	--	--	--	--			
	<b>SP (S-SP)</b>	--	--	--	--	--	--	SP < CSP (0.01)	
	<b>CSP (S-CSP)</b>	--	--	--	--	--	--	--	CSP(S) > CSP(SP) (0.05)
	<b>CSP (SP-CSP)</b>	--	--	--	--	--	--	--	--

Notes: 1. **SP > S** means sending by A in the observations of condition **SP** is greater than sending by A in the observations of condition **S**, etc.  
 2. The number in parenthesis is the significance level in a one-tailed Mann-Whitney test. 3. The average amount sent by each subject A during the first five interactions is treated as one observation.

**Table 3b. Mann-Whitney Tests of Percent Returned by *B***

		R1-R5			R6-R10				
		S (S-SP & S-CSP)	SP (SP-S & SP-CSP)	CSP (CSP-S)	S (SP-S)	S (CSP-S)	SP (S-SP)	CSP (S-CSP)	CSP (SP-CSP)
R1-R5	S (S-SP & S-CSP)	--	S < SP (0.10)	S < CSP (0.01)	S < S (NS)	S < S (0.10)	S < SP (NS)	S < CSP (0.01)	S < CSP (0.10)
	SP (SP-S & SP-CSP)	--	--	SP < CSP (0.05)	SP > S (NS)	SP > S (NS)	SP > SP (NS)	SP < CSP (0.01)	SP < CSP (NS)
	CSP (CSP-S)	--	--	--	--	CSP > S (0.01)	CSP > S (0.01)	CSP < CSP (NS)	CSP > CSP (NS)
R6-R10	S (SP-S)	--	--	--	--	S(SP) < S(CSP) (NS)	S > SP (NS)	S < CSP (0.01)	
	S (CSP-S)	--	--	--	--	--			
	SP (S-SP)	--	--	--	--	--	--	SP < CSP (0.01)	
	CSP (S-CSP)	--	--	--	--	--	--	--	CSP(S) > CSP(SP) (0.10)
	CSP (SP-CSP)	--	--	--	--	--	--	--	--

Notes: 1. **SP > S** means the proportion returned by *B* in the observations of condition **SP** is greater than the proportion returned by *B* in the observations of condition **S**, etc. 2. The number in parentheses is the significance level in a one-tailed Mann-Whitney test. 3. The average amount returned by each subject *B* during those of the first five interactions in which counterparts *A* sent a positive amount is treated as one observation.

**Table 4. Determinants of A's sending and B's % return in CSP interactions, including coded verbal communication measures (GLS regressions)**

	A's sending (1)	A's sending (2)	B's return % (3)	B's return % (4)	B's return 'shortfall' (5)	B's return 'shortfall' (6)
A,B Agree (VA)	1.866*** (0.318)	0.917*** (0.274)	-0.010 (0.026)	-0.026 (0.029)	-0.076** (0.032)	0.010 (0.030)
A Inquires	-0.126 (0.207)	-0.171 (0.199)	-0.018 (0.014)	-0.020 (0.014)	0.014 (0.018)	0.022 (0.017)
B Assures	-0.219 (0.489)	-0.043 (0.454)	0.072* (0.040)	0.073* (0.040)	-0.032 (0.045)	-0.036 (0.044)
Non-Task	0.238 (0.248)	0.226 (0.237)	0.006 (0.020)	0.008 (0.021)	-0.017 (0.023)	-0.025 (0.022)
CSP in S-CSP	0.348 (0.414)	0.149 (0.244)	-0.741*** (0.178)	-0.745*** (0.182)	-0.611*** (0.031)	0.099 (0.096)
CSP in SP-CSP	-1.034 (2.167)	-1.666 (2.381)	-0.183* (0.105)	-0.189* (0.109)	-0.719*** (0.044)	-0.130 (0.152)
Period	-0.062 (0.335)	0.075 (0.312)	-0.077** (0.031)	-0.075** (0.031)	0.095*** (0.036)	0.086** (0.034)
Period squared	-0.003 (0.052)	-0.015 (0.049)	0.012** (0.005)	0.011** (0.005)	-0.013** (0.006)	-0.013** (0.005)
A's sending			0.033*** (0.009)	0.031*** (0.009)		
Numerical Agreement (NA)		2.358*** (0.615)		0.043 (0.040)		-0.209*** (0.049)
# observations	397	397	382	382	382	382
Wald $\chi^2$	216.32	426.49	109200.23	109343.53	110130.87	94935.78
Prob. > $\chi^2$	0.00	0.00	0.00	0.00	0.00	0.00

Notes: 1. Numbers in parentheses are robust standard errors. 2. All estimates include individual fixed effects.

3. regressions (3) – (6) include only the observations in which A sent more than zero. 4. The first four explanatory variables are coded from chat communication. 5. CSP in S-CSP and CSP in SP-CSP are dummy variables; CSP in CSP-S is the omitted category. 6. Round 6 = period 1, round 7 = period 2, etc. 7. Numerical Agreement = 1 if B and A chose same row and column, otherwise 0. 8. \*, \*\* and \*\*\* indicate significance at the 10%, 5%, and 1% levels, respectively.

**Table 5a. Determinants of A's violation of numerical agreement, pooled SP and CSP observations, by round (random effects probit regressions)**

	Rounds 1-5				Rounds 6-10				Rounds 1-10			
A's proposed row		-0.141* (0.073)		-0.142* (0.078)		-0.286** (0.121)		-0.241* (0.125)		-0.186*** (0.059)		-0.166*** (0.060)
Period	0.015 (0.691)	0.261 (0.644)	0.044 (0.733)	0.315 (0.682)	1.053 (0.800)	1.120 (0.778)	0.995 (0.805)	1.056 (0.777)	0.546 (0.496)	0.667 (0.470)	0.531 (0.473)	0.682 (0.480)
Period squared	-0.011 (0.113)	-0.046 (0.106)	-0.017 (0.119)	-0.055 (0.111)	-0.143 (0.124)	-0.152 (0.119)	-0.137 (0.125)	-0.146 (0.119)	-0.081 (0.079)	-0.097 (0.075)	-0.081 (0.076)	-0.101 (0.077)
CSP dummy			-1.437** (0.647)	-1.179** (0.524)			-0.903* (0.515)	-0.583 (0.427)			-0.853*** (0.281)	-0.750*** (0.278)
# observations	259	259	259	259	274	274	274	274	533	533	533	533
Wald $\chi^2$	0.16	3.89	5.06	8.24	2.20	7.23	5.06	8.82	1.30	10.49	10.19	16.77
Prob. > $\chi^2$	0.92	0.27	0.17	0.08	0.33	0.06	0.17	0.07	0.52	0.01	0.02	0.00

Notes: 1. Dependent variable equals 1 if A sent less than agreed, 0 otherwise. 2. Only interactions with numerical agreement are included. 3. Round 6 = period 1, round 7 = period 2, etc. 4. Numbers in parentheses are standard errors. 5. \*, \*\* and \*\*\* indicate significance at the 10%, 5%, and 1% levels, respectively.



**Table 5b. Determinants of *B*'s violation of numerical agreement, pooled SP and CSP observations, by round (random effects probit regressions)**

	Rounds 1-5				Rounds 6-10				Rounds 1-10			
<i>B</i> 's proposed column		0.532 (50.584)		0.568 (110.470)		0.018 (0.022)		0.020 (0.021)		0.037* (0.019)		0.038** (0.018)
Period	0.020 (0.568)	-0.432 (0.618)	0.030 (0.570)	-0.420 (0.620)	0.411 (0.524)	0.370 (0.527)	0.382 (0.527)	0.338 (0.529)	0.271 (0.383)	0.156 (0.390)	0.281 (0.386)	0.158 (0.393)
Period squared	0.020 (0.093)	0.086 (0.101)	0.019 (0.093)	0.085 (0.101)	-0.040 (0.083)	-0.033 (0.083)	-0.037 (0.083)	-0.030 (0.083)	-0.018 (0.061)	-0.001 (0.062)	-0.021 (0.062)	-0.003 (0.063)
CSP dummy			-0.828 (0.716)	-0.887 (0.729)			-1.515** (0.713)	-1.540** (0.711)			-0.995*** (0.370)	-1.026*** (0.370)
# observations	249	249	249	249	264	264	264	264	513	513	513	513
Wald $\chi^2$	1.65	1.30	2.94	2.72	2.83	3.44	6.78	7.61	4.70	8.08	11.41	15.30
Prob. > $\chi^2$	0.44	0.73	0.40	0.61	0.24	0.33	0.08	0.11	0.10	0.04	0.01	0.00

Notes: 1. Dependent variable equals 1 if *B* sent less than agreed, 0 otherwise. 2. Only interactions with numerical agreement and in which *A* sent the amount agreed are included. 3. Round 6 = period 1, round 7 = period 2, etc. 4. Numbers in parentheses are standard errors. \*, \*\* and \*\*\* indicate significance at the 10%, 5%, and 1% levels, respectively.