The Journal of Socio-Economics 40 (2011) 1–13

Contents lists available at ScienceDirect

The Journal of Socio-Economics

journal homepage: www.elsevier.com/locate/soceco

Lavish returns on cheap talk: Two-way communication in trust games

Avner Ben-Ner[a,*], Louis Putterman[b], Ting Ren[c]

[a] Carlson School of Management, University of Minnesota, United States
[b] Department of Economics, Brown University, United States
[c] HSBC School of Business, Peking University, China

ARTICLE INFO

Article history:
Received 29 August 2010
Accepted 8 September 2010

JEL classification:
C72
C91
D63

Keywords:
Trust game
Trust
Trustworthiness
Reciprocity
Commitment
Communication
Cheap talk

ABSTRACT

We conduct trust game experiments in which subjects can sometimes exchange proposals either in numerical (tabular) form, or using chat messages followed by exchange of numerical proposals. Numerical communication significantly increases trusting and trustworthiness; inclusion of 1-min verbal communication in a chat room generates an even larger and more robust effect. On average, trustors send $9.21 of their $10 endowment as compared to $7.66 in the standard trust game, and trustees return 56% vs. 45%. Chat enhances the likelihood that trustors and trustees will adhere to non-binding agreements they make—an additional interpretation of trusting and trustworthiness—and increases the probability that subjects will propose, accept, and abide by equal-division agreements. Analysis of the content of subjects' verbal communication shows that what is said, and not only the fact that things are said, significantly affects outcomes.

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 et al., 1995, hereafter BDM) has become 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 absent in past trust experiments is the possibility of two-way 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 is a matter of interpretation only (Camerer, 2003). If the parties are given the opportunity to reach an understanding before the interaction, however, the status of their subsequent actions as manifestations of trusting 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.

Received economic theory assuming self-interested actors leaves room for communication to increase trustworthiness and thus trust only insofar as agreements can be backed by sanctions, including withdrawal of future cooperation. By contrast, many students of social behavior including a growing number of behavioral economists posit biologically evolved and culturally elaborated mechanisms that may help to support trust even in the absence of material sanctions. In practice, a look in the eye and a handshake seem often to provide valued assurance even when formal enforcement is unavailable, and a face-to-face meeting may be considered indispensable for sizing up the character ("type") of the other party. While it is unclear a priori whether less intimate channels of communication in conditions of anonymity and no prospect of future interaction can serve a similar purpose, our results indicate that this can indeed be the case.[1]

[1] Ben-Ner and Putterman (2009) permitted subjects in trust games to engage in pre-play communication that could end with binding contracts. Interestingly, they found that most subjects forewent the contracts despite their modest cost, and that the lion's share of mutually beneficial exchanges took place without contracts in place. We discuss other differences with the present paper in Section 2.

* Corresponding author.
E-mail address: benne001@umn.edu (A. Ben-Ner).

1053-5357/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.socec.2010.09.009

We conducted one-time-only computer-mediated interactions between subjects who remained fully anonymous to one another. Our data provide evidence not only of a high incidence of self-commitment but also of a significantly stronger impact, on both self-commitment and on its credibility to others, of communication in words than of communication in the numerical strategy space alone. Not only are there significantly higher earnings and a more equal distribution of those earnings under a condition permitting than under one not permitting the exchange of words, but content of messages also displays effects, insofar as (a) first-movers send significantly more when their counterparts and they have verbally agreed on a course of action than when they have simply exchanged matching strategies in numerical form and (b) second-movers adhere to their agreements significantly more often if they have issued an explicit promise of trustworthiness.

The paper joins a small but growing literature on two-sided communication in experiments,[2] and is one of only a handful of papers that investigate the effect of communication on trusting and trustworthiness. 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.[3] In the BDM trust game, *A* and *B* begin with 10 units of experimental currency, *A* selects $X_a \in (0, 1, \ldots, 10)$, *B* receives $3X_a$, and *B* selects $0 \leq X_b \leq 3X_a$ to return to *A*, yielding payoffs $(\pi_a, \pi_b) = (10 - X_a + X_b, 10 + 3X_a - X_b)$. Assuming that *A* and *B* interact only once, anonymously, and without the possibility of acquiring reputations, the prediction of economic theory assuming rational, own-payoff-maximizing actors who know other actors to be of this type is unambiguous. Returning $X_b > 0$ violates either rationality or strict self-interest or both. Because *B* will never return any money, *A* will never send any. In contrast, sending by *A* need not violate either rationality or self-interest if *A* has reason to believe that some *B*'s have a non-selfish preference, for instance reciprocity (Hoffman et al., 1998; Fehr and Gächter, 2000). Suppose, for example, that self-interested *A*'s believe that there exist both Type R *B*'s who prefer to return $1 > r > 1/3$ of what they receive rather than return nothing, with the others being Type N *B*'s who simply maximize own earnings and therefore return nothing. Then in the absence of further information, such *A*'s consult their estimate s^e of the share s of *B*'s who are of Type R, and (assuming no risk aversion) send their entire endowment if and only if $s^e > (1/3r)$. For example, if $r = 2/3$—the value that equalizes payoffs between the two actors—then *A* sends her full endowment if she believes that more than half of *B*'s are of Type R, since *A*'s expectation of X_b is $s^e \cdot r \cdot 3X_a = s^e \cdot 2X_a > X_a$ for $s^e > 0.5$.

How might communicating before acting alter trust game outcomes? If we revert to assuming that agents on both sides are strictly own-payoff-maximizing, it can have no effect. However, in a world in which proportion s of *B*'s are of Type R, the opportunity for pre-play communication could alter play if Type R second-movers can send signals of "R-ness" with lower cost or higher fidelity than can those of Type N. If Type N's nonetheless have some capacity to send these signals, we might see an equilibrium in which both types attempt to signal R-ness, with *A*'s attempting to screen. The possibilities are further enriched if some second-movers but not others have an aversion to breaking a commitment or lying.[4] In this case *B*'s may be of four types, RT, RL, NT and NL, where R and N are as before and T and L represent truth-telling (self-committing, lying-averse, etc.) and lying (non-self-committing, etc.) dispositions.[5] Then pre-play communication may alter play not only due to the possibilities it introduces of signaling and screening for reciprocity, but also because it allows RT and NT subjects to commit themselves to particular courses of action, and because signaling and screening for self-commitment may also be possible now. The possibility to self-commit can affect play even if having attribute T or L is strictly unobservable, since rational, self-interested *A*'s will (again abstracting from risk aversion) send money to *B*'s who promise to return some share, say r', so long as their expectation of the proportion σ of subjects who have attribute T satisfies $\sigma^e > (1/3r')$.[6]

Reality is surely more complicated, with individuals falling not into four but into a far larger number of types with respect to both reciprocity and self-commitment. There is the possibility, too, that how reciprocating or self-committed a given second-mover is in a given exchange will depend in part on her assessment of the qualities of the first-mover with whom she is paired, so that the first-mover as well has reason to engage in signaling.[7] Consider, for instance, that a first-mover who requests that the lion's share of money received be returned might be judged to be greedy and thus not worthy of reciprocity or commitment by a second-mover, although the latter may be inclined towards both when dealing with a "fairer" partner. This consideration might lead strictly self-interested first-movers to propose more equal exchanges. Ultimately, whether communication increases trust and/or trustworthiness is an empirical question, which we address in detail in this paper.

There have been few experiments with pre-play communication in trust games, and only two we know of in which both parties could exchange proposals or task-relevant oral or text 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 et al. (2000) and Buchan et al. (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 (a similar condition is studied by Andreoni (2005)). Fehr and

[2] For reviews of the literature on communication in games in general, see Ben-Ner and Putterman (2009) and Bochet and Putterman (2009). For additional references to the trust game literature, see our working paper.

[3] "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).

[4] Charness and Dufwenberg (2006) posit that *B* may return money when she promises to do so not because she suffers disutility from lying *per se*, but because she gets disutility from creating and then failing to live up to expectations that will harm *A* if not fulfilled.

[5] The utility of a subject in role *B* might be written as $U_i = (10 + 3X_a - X_b) - \rho_i R - \lambda_i L$ where R is an indicator of fulfillment of a reciprocity "obligation" taking value 0 if $X_b \geq \alpha X_a$, where $\alpha \geq 1$ is determined by prevailing norms, and value 1 otherwise, and L is similarly an indicator of fulfillment of a promise (non-lying) "obligation" taking value 0 if *B* fulfills her promise (assuming that her counterpart has also done so) and value 1 if not. ρ and λ are potential disutility terms that bind agents of one type to the social behavior of reciprocity or non-lying, respectively, but have no behavioral impact on other agents. There is a large literature on how preferences other than for the maximization of own material well-being could have evolved, as well as on gene–culture interactions; see for example Boyd and Richerson (1985), Field (2001), and Gintis et al. (2005).

[6] A smaller proportion of truth-tellers suffices if some non-truthtellers are reciprocators—i.e., the set of RL types is not empty.

[7] Theoretical models in which manifesting a social preference depends on the perceived intentions of an interaction partner include Falk and Fischbacher (2006) and Cox et al. (2007).

List (2004), Fehr and Rockenbach (2003), Houser et al. (2008) 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. Servátka et al. (2008), too, introduce a treatment in which trustees can send a free-form written message to trustors, obtaining similar results to Charness and Dufwenberg. They also demonstrate that pre-play promises by trustees induce substantially more trusting and trustworthiness than either a pecuniary deposit mechanism or an option for the trustee to send both a message and a pecuniary deposit. Charness and Dufwenberg (2007) find that there is a stronger effect when the trustee can send a free-form message than when he or she can only choose whether or not to send a "canned" promise made available by the experimenter, a result that resembles our finding about numerical vs. text messages and a treatment in Bochet and Putterman (2009) in which "canned" promises are available in a voluntary contributions game.[8]

Ben-Ner and Putterman (2009), using a similar experimental framework as the present paper, compare the effects of numerical and verbal communication to the standard BDM game and find significant returns to communication. However, they use a fixed-order design (a standard BDM-style interaction followed by a numerical interaction and then an interaction with pre-play text communication), thus being unable to fully distinguish the effect of form of communication and its absence from order (such as learning) effects.[9] Furthermore, subjects who reached agreement are, in interactions with either form of communication, given the option to enter into binding contracts should they achieve an agreement. Although interesting for illustrating the proposition that trusting and trustworthy behaviors can substitute for costly contracting, therefore, the pre-play communication in that paper can be viewed as a component of contract negotiations. In contrast, the present paper has no contract opportunities, so we capture the pure effect of pre-play communication in what is otherwise a standard form trust game. In addition, that paper (unlike the present one) does not study communication content. Another experiment that includes two-way task-relevant communication is reported by Bicchieri et al. (2010). Subjects engaged in one interaction with no pre-play communication, one with pre-play chat communication, and one with pre-play face-to-face communication, always in the same order and each time with a different partner. The game played resembles the BDM trust game except that second-movers received no endowments. One set of 32 subjects were required to discuss topics other than the game, while 32 others (hence, 16 pairs per condition) were permitted to discuss any topic. Like the present paper, Bicchieri et al. find that relevant communication significantly increases both trusting and trustworthiness. Our paper differs in having a larger subject pool, comparing numerical to text communication, using an interaction table to facilitate pre-play proposals, analyzing the content of text communication, controlling for order effects, and using the original BDM payoff structure without alteration.[10]

2. 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 a series of trust game interactions, the number of which is not pre-announced. As announced to subjects, each interaction involved a different anonymous partner, with *A*'s and *B*'s seated in different buildings. In each interaction, both *A* and *B* begin with 10 units of experimental currency labeled experimental dollars (*E*$), and *A* and *B* respectively select X_a and X_b as described above, precisely the moves and payoffs studied by BDM. The only modification of the BDM action space is that rather than selecting X_b from all integers in the relevant range, *B* selects a fraction $r_b \in (0, 1/6, 1/3, 1/2, 2/3, 5/6, 1)$, which determines *B*'s sending choice $X_b = r_b \cdot 3X_a$.[11] *A*'s options were displayed as row headings and *B*'s as column headings in a table appearing on the computer screens of both players (Fig. 1). Each cell lists the resulting value of X_b, above, and the resulting final earnings pair (π_a, π_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 min pre-laboratory sign-up survey and the 30 min laboratory experiment.[12] Earnings in the laboratory portion averaged $18.18 for a total of ten interactions per session.

We study three types of interaction or experimental conditions. 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 r_b by clicking on a column. In numerical proposal (NP) interactions, *A* first clicked-and-highlighted both a row and a column, described in the instructions as a proposal; then *B* clicked-and-highlighted a row and column, also described as a proposal, which might or might not agree with *A*'s. Subjects were told that these actions had no direct effect on payments. Then *A* chose any row and *B* chose any column, as in a simple interaction. In chat plus numerical proposal (CNP) interactions, *A* and *B* could exchange text messages in a private chat room for 1 minute, then went through the same procedures as numerical proposal. Messages could not reveal one's identity or contain threats or promises except ones regarding the permissible game moves.[13]

Every subject participated in 10 interactions, five of one condition and five of another. Subjects were not informed in advance of the number of interactions or the future conditions, only that there will be several interactions, each with a new partner. We study five session designs to investigate the effects of communication both by inter-subject and by intra-subject comparisons, controlling for order effects. In S–NP, each subject engaged in five interactions with no communication, then five interactions with only numerical proposals. In NP-S, this order is reversed, with each subject engaging

[8] By "canned" promise, we mean a promise message written by the experimenter that the subject can either select or not, but cannot alter, and where both subjects in the interaction know this to be the case.

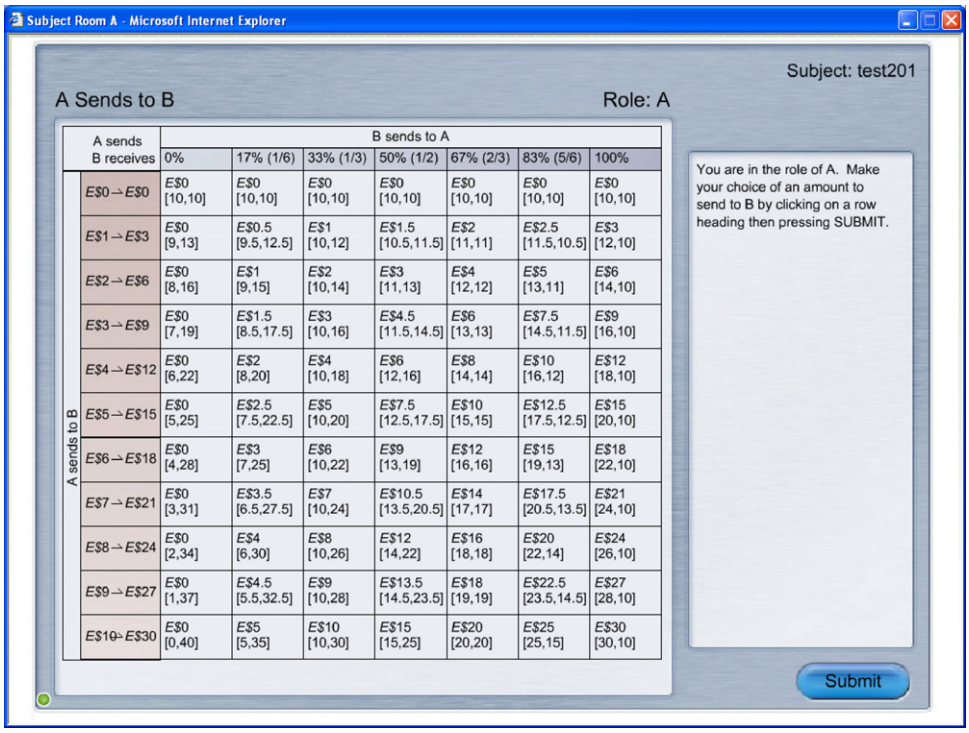
[9] Chat length in Ben-Ner and Putterman (2009) was 4 minute, whereas in the present paper is 1 minute.

[10] When the second-mover has no endowment, as in Bicchieri et al., there is a stronger possibility of confounding first-mover trusting with first-mover altruism.

[11] 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* is given no endowment, as in the example of Bicchieri et al.

[12] Subjects were told that the laboratory portion would take up to 1 hour. The survey, completed between 2 days and a week before the lab portion, produced data on personal background and characteristics that we plan to use in conjunction with the experiment decisions in other research.

[13] Subjects were told they would forfeit all earnings if they violated these rules. A review of the chat messages shows no violations.



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

Fig. 1. Interaction table.

in five NP interactions and then five S interactions. In S–CNP, five simple interactions are followed by five interactions with chat and numerical proposal. CNP–S reverses this order. We carried out also an NP–CNP treatment with five NP interactions followed by five CNP interactions, but due to resource limitations, we did not also reverse that order. Note that because subjects were not informed in advance of the exact number of interactions, nor did they know what condition would follow the initial one, there should be no "end game" effects,[14] and initial interactions under a given condition (for example, the S interactions of the S–NP and of those in the S–CNP 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 moral values, ethical concerns, or preferences such as reciprocity or lying-aversion, (3) rational, and (4) assume that their counterparts in the experiment are like them in these respects. Under these assumptions and assuming that the 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 each interaction. 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.[15]

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 games, and large numbers have done so as well in past experimental trust games[16]; (b) communication has been found 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)[17]; (d) many people have a preference for keeping their word[18]; (e) there exist individual differences that generate a distribution of types in terms of money-maximizing vs. other preferences[19]; 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:

[14] Although word of there being exactly ten interactions might have spread in a smaller subject pool had the experiment been conducted in a large number of sessions, the very large student population (over 40,000) from which our subjects were drawn by individualized e-mail invitations and the small number of sessions (nine in total) meant that this did not happen, as attested by the numerous unanswered inquiries in subjects' chat messages "do you know when this will end?" "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.

[15] Theoretically, there could be a cooperative equilibrium in an infinitely repeated version of the trust game with random matching, analogous to that in the prisoner's dilemma game analyzed by Kandori (1992). But the facts that our subjects knew they would never meet twice, that they would each engage in only a few interactions, and that behaviors in past interactions were not public, rule out such an equilibrium. Even if this were not the case, the multiplicity of decision pairs that can render both *A*'s and *B*'s better off (see Fig. 1) would also give rise to a severe coordination problem. While investments in reputation analogous to those discussed by Kreps et al. (1982) cannot be ruled out, their model belongs more properly to a world of non-standard preference types that we address under the heading of behavioral economics.

[16] See the summary in Camerer (2003) and the symposium in *Journal of Economic Behavior and Organization*, vol. 55, no. 4, 2004.

[17] Put in terms of the notation in note 5 above, many act as if maximizing a utility function in which $\rho_i > 20$, generating a large utility loss if they return too little. References to the literature on reciprocity include Hoffman et al. (1998), Fehr and Gächter (2000) and Gintis et al. (2005), among many other contributions.

[18] Gneezy (2005), Sanchez-Pages and Vorsatz (2007), and Charness and Dufwenberg (2006).

[19] Kurzban and Houser (2001), Fischbacher et al. (2001), and Page et al. (2005).

Table 1
Summary of experimental designs.

Design	Description	No. of sessions	Total participants ($A+B$)	Number of interactions
S–NP	5 simple interactions, 5 interactions with one numerical proposal per subject	2	48	232
NP-S	5 interactions with one numerical proposal per subject, 5 simple interactions	2	50	241
S–CNP	5 simple interactions, 5 interactions with chat plus one numerical proposal per subject	2	64	309
CNP–S	5 interactions with chat plus one numerical proposal per subject, 5 simple interactions	2	68	335
NP–CNP	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.

- 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 independently-crafted written messages, not only because of the way in which this aids coordination, signaling and screening of types, but also because with verbal messages 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; this will especially be so when agreements are reached in words—the chat room equivalent of a hand shake;
- many subjects will adhere to agreements, especially ones reached verbally and including explicit promises or assurances;
- many subjects will be drawn to and will tend to follow through on 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 expected. 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 even 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 thus augment the bullet point above by hypothesizing that

- rather than make the proposal which maximizes her prospective earnings assuming trustworthiness of B and subject to B receiving some benefit, even a selfish A who does not care about fairness will tend to make the (X_a, X_b) = ($E\$10$, $E\$20$) proposal in order 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 enhancement of trusting, trustworthiness and fairness by communication, especially communication of words.

3. Results

Nine sessions with 12–18 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 numbering over forty thousand. Table 1 shows the number of sessions and subjects by treatment. The total number of interactions to be analyzed should ideally be 1320 (1/2*10*264) but is slightly smaller because computer breakdowns left a few interactions uncompleted. We begin by analyzing the first five interactions, when subjects operated under their first mode of interaction without knowledge of what would follow.[20]

3.1. The effect of communication condition on sending and returning

Table 2 shows mean and median sending and return proportions, by condition. In rounds 1–5 (R1–R5) average sending is higher in NP (7.68) than in S (7.34), and higher still in CNP (8.65) where more than half of all sending decisions were of the full endowment, yielding median sending of 10. This entails a gain of 0.34 associated with numerical communication relative to the standard trust game, and 1.31 associated with verbal communication. Since all sent amounts are tripled, the gains (how much is generated by the interaction between A and B) are truly large, 3.93 in CNP vs. S. The same ordering appears for average proportion of tripled amount returned, ranging from 42.50% in S to 50.05% in NP to 56.75% in CNP.[21] The middle portion of Table 2 focuses on the second block

[20] 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 instructions 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.

[21] It may be noted that both trust and trustworthiness levels are high in our S condition relative to other trust experiments 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 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. In the first simple trust interaction in our experiment, by 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. Since the subjects in both our own experiment and BDM were

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

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

[a] Excluding the cases when *A* sent zero.

of interactions (rounds 6–10), where each type of interaction was preceded by a block of five rounds of a different type of interaction (hence learning might have taken place). The ordering and magnitude of amounts sent and proportions returned are similar to those in rounds 1–5.

To see whether these differences are statistically significant, we conducted Mann–Whitney tests in which each observation is the average, for one subject, of amount sent or proportion returned in rounds 1–5, in the first row of Table 3.[22] The two-tailed tests summarized in the first row of columns (1), (4) and (7) find first-mover sending to be significantly higher in CNP than in NP and S conditions, but not significantly different in NP than in S. For proportion returned by second-movers who received positive amounts, the tests show significant differences for all comparisons: NP > S, CNP > NP and CNP > S.

The comparisons among subjects exposed to differing conditions in their first five interactions provide our cleanest test of the effect of communication condition, since each subject faced only one condition during these periods and had neither experience nor knowledge of other conditions when making her decisions.[23] But other sets of tests also help us to gauge the effect of communication. First, we can check what happened to trusting and trustworthiness when the same subject switched from one condition to another. This has the advantage of being an intra-subject comparison and thus permitting matched-pair testing, but it also entails difficulties in interpreting order effects. Second, inter-subject comparisons can be done of behaviors under the different conditions in the second set of interactions. The latter comparisons include ones in which the first block's experience is held constant (for example, comparing behavior in the NP and CNP conditions of the S–NP and S–CNP designs), and ones in which that is not the case (for example, comparing behavior in the second conditions of the S–NP and the NP–CNP designs). Although less amenable to rigorous interpretation, comparisons of behaviors under conditions S, NP and CNP in all ten interactions, without regard to design, are also of some interest. The middle portion of Table 2 reports means and medians relevant to the second block comparisons (rounds 6–10), The bottom portion of Table 2 gives comparison among conditions for the 10 interactions as a whole, and the second through seventh rows of Table 3 summarize the Wilcoxon and Mann–Whitney test results for both intra-subject and inter-subject comparisons.

The results show that interactions with chat (CNP) generally exceed with at least marginal statistical significance those without chat, whether numerical proposals are possible (NP) or not (S), in terms both of the share of endowment sent by first movers (X_a) and the proportion of money received that is returned by second movers ($X_b/3X_a$) conditional on $X_a > 0$. The results for both intra-subject and inter-subject comparisons tell a story that is consistent with the one obtained from the initial, first block comparisons (rounds 1–5). Overall, while average sending by first-movers in the S and NP conditions is essentially indistinguishable (at about 7.7), average sending in CNP is considerably higher, at 9.2. The average proportions returned by second-movers is somewhat higher (at 49.2%) in NP than in S (46.0%) interactions, but considerably higher in CNP interactions (at 57.0%). Only small fractions (15.7% in S, 15.4% in NP, and 12.3% in CNP) of those *B*'s receiving positive amounts returned nothing, with more than three-quarters (77.7% in S, 79.2% in NP, and 86.7% in CNP) at least "making their counterpart whole" by returning 1/3 or more. The proportion of *B*'s returning enough ($\geq 2/3$) to make *A*'s earn at least as much as themselves, is a remarkable 78.3% in the CNP interactions vs. the still considerable 46.8% and 52.0% for S and NP interactions, respectively.

The comparisons of behaviors in the S and NP conditions, in contrast, are mixed and often insignificant. There is some support for the first row's finding that permitting exchange of numerical proposals raised the proportion of money returned by second-movers: the first intra-subject test shows that proportion returned was significantly higher in the second block when subjects transitioned from the S to the NP condition. But there are contrary if statistically insignificant findings from the other comparisons in column (2). The column (1) comparisons for amount sent by *A* in the S and NP conditions are even more inconclusive.[24]

Columns (3), (6) and (9) of Table 3 display test results for equality of earnings between *A*'s and *B*'s. While the earnings ratio for these

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 multiple interactions and lower stakes per interaction, and the use of the interaction table making transparent the implications of each choice.

[22] Since second-movers lacked a meaningful choice when nothing was sent to them, we average for a second-mover only over those periods in which the counterpart sent $X_a > 0$—the vast majority of cases.

[23] Using only subjects' first interactions (round 1) could be argued to provide a still cleaner test insofar as the interactions are entirely free of the potential effects of experimental experience. Considering only first interactions, we find that the difference between *A*'s sending in the CNP vs. NP and in CNP vs. S conditions are significant at the 1% level, but the difference between *A*'s sending in NP vs. S is not significant. The difference between *B*'s fraction returned in the CNP vs. S conditions is significant at the 1% level, that in the NP vs. S conditions is significant at the 5% level, but that in the CNP vs. NP conditions is not significant.

[24] Four of five of the test results show sending to be greater in S than in NP condition. In the two cases in which the difference is significant or marginally so, however, it could be argued that prior exposure to communication might have boosted sending in the later S condition.

Table 3
Tests of differences in sending, returning (proportion of amount received), relative payoffs of *A* and *B* across conditions.

	S vs. NP			NP vs. CNP			S vs. CNP		
	X_a (1)	$X_b/3X_a$ [a] (2)	π_b/π_a [a] (3)	X_a (4)	$X_b/3X_a$ [a] (5)	π_b/π_a [a] (6)	X_a (7)	$X_b/3X_a$ [a] (8)	π_b/π_a [a] (9)
Inter-subject comparison, rounds 1–5	S > NP n.s.	S < NP*	S > NP***	NP < CNP**	NP < CNP***	NP > CNP***	S < CNP***	S < CNP***	S > CNP***
Intra-subject comparison, less to more communication[b]	S < NP n.s.	S < NP**	S > NP n.s.	NP < CNP[d]	NP < CNP n.s.	NP > CNP n.s.	S < CNP***	S < CNP***	S > CNP***
Intra-subject comparison, more to less communication[c]	S > NP**	S < NP n.s.	S > NP*	n.a.	n.a.	n.a.	S < CNP [d]	S < CNP***	S > CNP n.s.
Inter-subject comparison, rounds 6-10	S in NP-S vs. NP in S–NP			NP in S–NP vs. CNP in S–CNP			S in NP-S vs. CNP in S–CNP		
	S > NP n.s.	S > NP n.s.	S > NP [d]	NP < CNP***	NP < CNP***	NP > CNP***	S < CNP***	S < CNP***	S > CNP***
	S in CNP–S vs. NP in S–NP			NP in S–NP vs. CNP in NP–CNP			S in NP-S vs. CNP in NP–CNP		
	S > NP [d]	S > NP n.s.	S < NP**	NP < CNP**	NP < CNP*	NP > CNP***	S < CNP*	S < CNP [d]	S > CNP***
							S in CNP–S vs. CNP in S–CNP		
							S < CNP**	S < CNP***	S > CNP***
							S in CNP–S vs. CNP in NP–CNP		
							S < CNP n.s.	S < CNP [d]	S > CNP [d]

For inter-subject comparisons, results of Mann–Whitney tests in which individual observations are the behaviors of individual subjects averaged over rounds 1–5 or 6–10. For intra-subject comparisons, results of Wilcoxon signed rank tests in which the individuals' behaviors are likewise averaged over the relevant sets of rounds.

[a] Tests include interactions in which $X_a > 0$, only.
[b] Comparisons of same subjects' average behaviors in first vs. second treatment in session type S–NP, NP–CNP or S–CNP.
[c] Same but for session type NP-S or CNP–S.
[d] Significance at the 10% level in a one-tailed test.
* Significance at the 10% level in two-tailed tests.
** Significance at the 5% level in two-tailed tests.
*** Significance at the 1% level in two-tailed tests.

Table 4
Comparison of reaching and keeping an agreement by *A* and *B* between trust game with numerical proposals (NP) and trust game with chat and numerical proposals (CNP).

	NP vs. CNP					
	Incidence of agreement (1)	Incidence of agreement kept by *A* (2)	Incidence of agreement kept by *B*, conditional on agreement being kept by *A*[a] (3)	Incidence of fair and efficient agreement (among all the interactions) (4)	Incidence of fair and efficient agreement kept by *A* (among interactions with fair and efficient agreement) (5)	Incidence of fair and efficient agreement kept by *B*[a] (among interactions with fair and efficient agreement) (6)
Inter-subject comparison, rounds 1–5	NP < CNP***	NP < CNP***	NP < CNP[b]	NP < CNP**	NP < CNP*	NP < CNP***
Intra-subject comparison in NP–CNP sessions	NP < CNP**	NP < CNP n.s.	NP > CNP n.s.	NP < CNP***	NP > CNP[b]	NP < CNP[b]
Inter-subject comparison, rounds 6–10, NP in S–NP vs. CNP in S–CNP	NP < CNP***	NP < CNP***	NP < CNP n.s.	NP < CNP***	NP < CNP**	NP < CNP n.s.
Inter-subject comparison, rounds 6–10, NP in S–NP vs. CNP in NP–CNP	NP < CNP***	NP < CNP***	NP < CNP n.s.	NP < CNP***	NP < CNP n.s.	NP < CNP n.s.

For inter-subject comparisons, results of Mann–Whitney tests in which observations are the behaviors of individual subjects averaged over rounds 1–5 or 6–10. For intra-subject comparisons, results of two-tailed Wilcoxon signed rank tests in which the individual's behaviors are likewise averaged over the relevant sets of rounds.

[a] Includes interactions in which *A* follows agreement, only.
[b] Significance at the 10% level in a one-tailed test.
* Significance at the 10% level in two-tailed tests.
** Significance at the 5% level in two-tailed tests.
*** Significance at the 1% level in two-tailed tests.

Table 5a
Determinants of *A*'s violation of numerical agreement, pooled NP and CNP 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)
CNP 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 χ^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. > χ^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

* Significance at the 10% level.

** Significance at the 5% level.

*** Significance at the 1% level.

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

	Rounds 1–5			Rounds 6–10				Rounds 1–10				
B'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)
CNP 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 χ^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. > χ^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.

* Significance at the 10% level.

** Significance at the 5% level.

*** Significance at the 1% level.

cases always ends up favoring *B* players on average, the outcome is equal (or on occasion favors *A*) in many individual interactions. The tests show earnings when $X_a > 0$ to be significantly less unfavorable to first-movers in NP than in S, in CNP than in S, and in CNP than in NP, with only a single exception for the S vs. NP comparison, and with the differences being at least marginally significant in twelve of fifteen cases. In Section 3.3 we focus further on the surprising number of interactions with precisely equal outcomes, and on the power of the associated proposals to attract agreement and adherence.

3.2. Words vs. numbers and the reaching and keeping of agreements

Does verbal communication increase trusting relative to numerical communication by making it easier to reach agreements (the state in which *B*'s proposed row and column matches *A*'s) or by increasing subjects' adherence to those agreements that are reached? The data indicate that it is both. Again, we check significance of differences across condition with tests conditioning on order and design, shown in Table 4. For interactions 1–5, a Mann–Whitney test using observations for individuals averaged over the five periods shows the difference in reaching agreement to be significant at the 1% level (agreement was reached in 60% of NP interactions vs. 79% of CNP interactions). According to the first test in column (2), *A*'s also kept significantly more of their agreements when CNP condition held than under NP, in interactions 1–5, suggesting that *A*'s had more confidence that their counterparts would follow through on agreements reached in words. The remaining rows of these columns show that the same ordering holds for frequency of agreements and frequency of adherence to agreements by *A*'s both for the intra-subject comparison of interactions 6–10 vs. 1–5 of the NP–CNP design, and for the two inter-subject comparisons that can be performed for pairs of treatments having NP vs. CNP condition in place during rounds 6–10. In contrast, there is less consistent and less significant evidence for adherence to agreements by *B*'s being different in the two treatments.[25]

We also checked for differences in adherence to agreements by estimating random effects probit regressions. This allows us to control for the number of the interaction in its sequence (dubbed "period"), and the proposed row or column. Table 5a shows the results for *A*'s trust in the sense of carrying out an agreement, and Table 5b the results for *B*'s trustworthiness in fulfilling their part. The results suggest that proposals were adhered to significantly more in CNP than in NP condition, with the effect being greater for first-movers in rounds 1–5 but for second-movers in rounds 6–10.

3.3. Attraction to "fair and efficient" exchanges

If *B* can commit to an agreement that makes it profitable for *A* to send *B* money, what agreements should we expect to see, with what frequencies? With *B*'s options restricted to returning sixths of the amount received, and with *A* sending *E*$10 so as to maximize payouts, the possibilities for mutually beneficial agreement are limited to $(\pi_a, \pi_b) = (15,25), (20,20)$ or $(25,15)$. With what frequency should we expect to see these agreed to and carried out? *A* might drive a hard bargain and insist on (25,15) since only if *A* sends money can either benefit. But *B*, having the last action, may insist that he can commit to returning 15 and no more. Since both begin with equal endowments and any returning by *B* is a matter of adherence to a normative commitment rather than of bargaining power, finally, the even split (20,20) might be the safest for either to propose. We call an interaction or proposal "fair and efficient" (F&E) if it involves the pair of actions send 10, return 2/3, since this leads to the most equal division of the largest sum of earnings.

Of the 59.8% of NP interactions in which row and column agreement was reached, 88.6% conform to one of the three mutually-beneficial types, and of the latter, 90.6% are F&E agreements, with 8.8% agreeing on a (15,25) split and only 0.6% agreeing on (25,15). Of the 85.6% of CNP interactions in which row and column agreement was reached, 93.3% are mutually beneficial, with 96.2% of the latter being F&E, 2.8% agreeing on (15,25) and 0.9% agreeing on (25,15). *A*'s numerical proposal was of the F&E type 62.9% of the time in NP and 78.8% of the time in CNP. Such proposals were agreed to by *B* when made by *A* 76.4% of the time in NP and 97.4% of the time in CNP.[26] Once such a proposal was agreed on, *A*'s fulfilled their part by sending *E*$10 94.2% of the time in NP and 98.7% of the time in CNP. The counterpart *B*'s of these trusting *A*'s then returned the promised amount 67.1% of the time in NP, 85.1% of the time in CNP.[27] In the case of S interactions, where prior proposals were not possible, F&E interactions nonetheless took place 33.0%[28] of the time, a rate quite similar to the 33.8% of actually completed F&E interactions in NP but far exceeded by the 68.3% of completed F&E interactions in CNP.[29]

Was agreeing on an equal split payoff-maximizing for *A*'s? In CNP interactions, *A*'s earned an average of *E*$17.39 under F&E agreements vs. *E*$15.00 under (15,25) agreements and $8.33 under agreements on (25,15). In NP interactions, *A*'s earned an average of *E*$15.33 with (15,25) agreements vs. *E*$15.65 under F&E agreements, but *E*$20.00 in the one and only case of agreement on (25,15). All in all, proposing the equal split seems to have been a wise course for a first-mover. Somewhat surprisingly, however, neither *A*'s nor *B*'s turn out to have been less likely to fulfill their obligations under (15,25) than under F&E agreements, and accordingly second-movers' earnings were higher on average with the former than with the latter agreement.[30]

Since subjects were randomly assigned to their roles and lacked any basis for claiming unequal deservingness, preferences for fairness are likely to be a major factor explaining the preponderance of F&E outcomes. Yet equal split outcomes have been less common in past trust game experiments.[31] We think a likely reason

[25] The inter-subject comparison between subjects facing the NP and CNP conditions in rounds 1–5 and the two such comparisons for rounds 6–10 both show higher shares of agreements kept in CNP than in NP but the difference only reaches significance at the 10% level in a one-tailed test in the first block of interactions, and the difference is insignificant and has the reverse direction for the intra-subject comparisons in the NP–CNP design.

[26] This analysis counts as proposals in the CNP 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 CNP and NP treatments are thus identical.

[27] These percentages include a few cases (three in NP, one in CNP) in which *B* returned *more* than the promised 67%.

[28] This includes eight cases in which *B* returned more than 67%.

[29] We also investigated whether reaching an F&E agreement first via chat communication increased the likelihood of its fulfillment, with formal test results reported in the last three columns of Table 4. As seen in column (4), significantly more interactions reached F&E agreements in rounds 1–5, as well as in rounds 6–10, if the subjects concerned were interacting under the CNP than under the NP condition, and given individuals reached significantly more F&E agreements during their five CNP interactions than during their five NP interactions in the NP–CNP design. Columns 5 and 6 report that F&E agreements were significantly more likely to be adhered to by both first and second movers when rounds 1–5 were conducted with than when they were conducted without chat. The same ordering is observed for rounds 6–10, but with the difference statistically significant only for first-movers when comparing the interactions in one of the two relevant pairs of designs. The second row reports that *A*'s were slightly more likely to observe a given F&E agreement in their NP than in their CNP interactions of the NP–CNP design, but adherence to F&E agreements by *B*'s was somewhat greater in their CNP interactions.

[30] *A* sent the *E*$10 required in all 24 interactions reaching agreement on (15,25), while *B* returned the required *E*$15 in 23 cases and a bonus *E*$20 in one case.

[31] 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. In five other cases,

is that reasonably fair-minded second-movers with a modest self-serving bias can easily convince themselves that returning half of the tripled amount (leading to the (15,25) outcome—an approach Andreoni (2005) calls "split the difference"–is the fair thing to do. We conjecture that listing final outcomes explicitly at the bottom of each cell of the interactions table (Fig. 1) played a role in increasing the frequency of (20,20) outcomes. If so, this would constitute a notable framing or presentation effect.

3.4. Effects of content in text communication: agreement and assurance

Our data provide convincing evidence that including a text component in non-binding pre-play communication significantly increased the sending and returning of money. But what about such communication led to these effects? 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.[32] On the other hand, the extent of the observed impact might also be due to specific aspects of message contents, such as whether *A* asks questions geared to learning something about *B*, whether *B* offers assurances of trustworthiness, whether the parties verbally agree on a course of action, and so forth. Keeping in mind that no set of words should affect behaviors under an assumption of strict rationality and payoff-maximizing preferences, we offer a partial investigation here of how the contents of chat communication affected outcomes.

We studied the chat log of each CNP interaction and coded the following variables (the first of which was already introduced in Section 3.2):

VA: Did *A* and *B* agree verbally on a specific course of action (row, column)?[33] 0 = no, 1 = yes

A Inquires: Did *A* try to learn something about *B* by asking questions or by initiating or attempting to extend non-task-related conversation? 0 = no, 1 = yes, but cursory, 2 = yes, and substantive[34]

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

No. of words: What is the total number of words exchanged in the chat messages that *A* and *B* sent each other?

Table 6 shows six GLS regressions, each with subject fixed effects, controls for interaction number (period), the four coded variables, and clustering on session.[35] Only interactions under CNP condition are included. In even-numbered columns, we also include as an additional explanatory variable a dummy for the presence of a numerical agreement (NA), 1 if *B*'s row and column proposal matched *A*'s, 0 otherwise. Regressions (1) and (2) study determinants of the amount sent by *A*. In regression (1), the variable VA has a positive coefficient which is significant at the 1% level, suggesting that after controlling for other factors, *A* sent about *E*$1.96 more if *A*'s and *B*'s chat messages included agreeing on a course of action. When both the numerical agreement variable NA and the verbal agreement variable VA are included, in column (2), both variables obtain significant positive coefficients, although the magnitude of the coefficient on VA is cut roughly in half and its significance level falls somewhat.[36] The estimate suggests that having a verbal agreement increased the amount *A* sent by slightly under *E*$1, while having a numerical agreement increased *A*'s sending by close to *E*$2.40. None of the other message content variables have significant coefficients.[37]

Regressions (3) and (4) investigate the effects of communication content on the proportion returned by *B* conditioned on *A* sending a positive amount. That amount is included as a control ("*A*'s sending") to correctly measure other effects. Perhaps surprisingly, the coefficients on VA and NA are positive but far from being statistically significant. Another of the message content variables is statistically significant, however: if *B* offered verbal assurance of trustworthiness, *B* returns a larger proportion of the amount received, significant at the 10% level, with the coefficients suggesting that the proportion of the received amount that *B* returns is about 6% higher for *B*'s who offered an assurance in the text exchange in question than for those who did not. *B* also tends to return a larger proportion when *A* sends more, but this tendency is not significant at conventional levels (or is significant at the 10% level only in a one-tailed test).

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*. Ortmann et al. (2000) report a set of 16 one-shot interactions identical to BDM, and find 13 senders of positive amounts, of whom only two received back twice or more than what they sent. One of the two 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 four 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 counterpart. Three of these four 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 four of their 63 undergraduate subjects and five 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 four 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 et al. (2003), of whom 11 out of 29 senders (37.9%) received back at least twice as much as they sent.

[32] Eckel and Wilson (2006) find that when anonymity is taken so far in the laboratory that subjects have only the experimenter's word as evidence that they are playing with a real counterpart, subjects' doubts about their counterparts' existence can affect trusting. Eckel and Wilson manipulate confidence in the existence of counterparts by showing, or not showing, real time video of the room in which counterparts sit. Availability of text messages from counterparts under some but not other conditions may have similar effects. Analysis of our subjects' chat contents suggests that one subject may have doubted the existence of a human counterpart, and that a few subjects doubted the instructions' statements that each interaction matched them with a different counterpart.

[33] Note again that absence of agreement does not mean that chat content was necessarily conflictual. We should also point out that VA covers a wide range of circumstances, from exchanges in which one of the partners suggests a course of action and the other simply says "sure," "o.k.," etc., to ones in which an explicit promise is made, counterproposals are considered, assurances are given, etc. The full text of the interactions is available from the authors on request.

[34] If *A* wrote only "Hi," "Hi, there," "How's it going?" "What's up?" etc., this is scored 1. Examples of more substantial message that are coded 2 are "What do you think about this weather?" "How do you like those Twins?" "This is interesting, don't you think?" "What's your major?" "How do you plan to spend the money?"

[35] With one exception, there were two sessions for each design (see Table 1); the clustering controls for any session-level peculiarities. We code "period" as 1 for the first interaction under a condition, even if that interaction occurs in the 6th round of the session (for example, the first CNP round of the S–CNP treatment). In this case, round 7 would be period 2, and so on. Differences of design (S–CNP vs. CNP–S vs. NP–CNP), which are not our focus here, are controlled by subject fixed effects since each design was populated by a unique set of subjects.

[36] We coded VA 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. VA thus implies NA. The reverse does not hold, however, because even though *A* and *B* did not reach agreement on a specific course of action in their text messages, *B* could nevertheless select the same proposal as had *A* in their numerical exchange.

[37] We checked for significance of these other three message content variables when VA and NA are omitted, and found the omissions to have no effect. We also explored the idea that VA or NA might be endogenous to these or other aspects of message content, but regressions with VA or NA as dependent variable and the last three content variables as independent variables returned only one mildly significant coefficient on the number of words, a negative one.

Table 6
Determinants of *A*'s sending, *B*'s % return and *B*'s violation of numerical agreement in CNP, including coded verbal communication measures, GLS estimations with individual fixed effects and clustering by session.

	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)
Verbal agreement (VA)	1.961*** (0.700)	0.959** (0.444)	1.624 (4.119)	0.632 (5.198)	−7.629[a] (5.459)	−0.864 (5.028)
Numerical agreement (NA)		2.394*** (0.361)		2.606 (7.490)		−17.773*** (5.489)
A Inquires	−0.045 (0.121)	−0.087 (0.136)	−1.860 (2.800)	−1.946 (2.960)	0.637 (4.113)	1.222 (3.742)
B Assures	0.461 (0.709)	0.571 (0.542)	6.152* (3.353)	6.112* (3.338)	−4.962** (2.446)	−4.685* (2.699)
No. of words	0.006 (0.012)	0.004 (0.012)	0.048 (0.107)	0.053 (0.119)	0.084 (0.164)	0.048 (0.147)
Period	−0.020 (0.512)	0.094 (0.510)	−7.877 (7.449)	−7.777 (7.559)	9.575[a] (6.883)	8.889[a] (6.782)
Period squared	−0.012 (0.083)	−0.020 (0.084)	1.172 (1.167)	1.167 (1.169)	−1.336 (1.081)	−1.302 (1.057)
A's sending			3.006[a] (1.929)	2.918[a] (1.938)	−1.730 (2.178)	−1.132 (1.942)
# observations	397	397	382	382	382	382
Wald χ^2	23.98	68.73	10.13	12.81	98.66	69.16
Prob. > χ^2	0.00	0.00	0.02	0.01	0.00	0.00

Notes: 1. Numbers in parentheses are robust standard errors. 2. Regressions (3)–(7) include only the observations in which *A* sent more than zero. 3. The explanatory variables NA, *A* Inquires, *B* Assures, and no. of words are coded from chat communication. 4. Round 6 = period 1, round 7 = period 2, etc. 5. Numerical agreement = 1 if *B* and *A* chose same row and column, otherwise 0.

[a] Significance at the 10% level in a one-tailed test.
* Significance at the 10% level.
** Significance at the 5% level.
*** Significance at the 1%.

In regressions (5) and (6), the difference between the return proportion proposed by *A* (and possibly agreed to by *B*) and the actual proportion returned by *B*, called "*B*'s return 'shortfall'," is the dependent variable. According to regression (6), *B*'s concurrence with *A*'s numerical proposal causes the shortfall to be almost 18% smaller (significant at the 1% level): nearly all *B*'s honor their commitments. The addition of a verbal agreement appears to have no extra effect in regression (6), and when NA is omitted, in regression (5), the coefficient on VA is larger but still only marginally significant in a one-tailed test. Interestingly, however, the chat content variable *B* Assures has a coefficient significant at the 5% level in regression (5) and one of roughly the same magnitude and only slightly lower significance in regression (6). This means that *B*'s who offered verbal assurance that they could be trusted complied with *A*'s proposal (or at least did so to a higher degree) significantly more than others, with the coefficient estimates implying that the average gap between promised and implemented share returned was about 5% smaller when assurance had been given. The indication is that most *B*'s did not offer assurances lightly or for purposes of entrapment.

In sum, we do find evidence that what is said in communication, and not just the fact of communication, mattered to subjects' binding decisions. Achieving agreement via chat content significantly affects trusting by first-movers, while providing assurance of trustworthiness is significantly correlated with proportion returned and degree of proposal fulfillment by second-movers.

4. Conclusions

We report 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 such proposals, after which they participated in the standard BDM (1995) trust game mediated by an on-screen interaction table. We show that both trusting and trustworthiness are greater in the condition permitting both written and numerical messages (CNP) than in the condition restricted to numerical communication only (NP) and in a no-communication baseline condition (S). The increase in trusting is substantial: trustors send on average *E*$9.21 as compared to *E*$7.66 without communication and *E*$7.71 with only numerical exchange. Trustworthiness is also greater: second-movers in the pre-play communication with numerical proposals condition return on average 56% of the money they receive as compared to 45% without communication and 49% with only numerical communication.

Each of our 264 subjects participated in two types of interaction (S and NP, S and CNP, or NP and CNP), each interaction being with a different counterpart seated in a different building. We confirmed that order had no qualitative effect for the comparisons of behaviors under the S and CNP conditions, but did have complicating effects for other comparisons. While subjects displayed the most trusting and trustworthiness in the CNP 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, first-movers earned an average of *E*$16.55 in CNP, *E*$13.40 in S, and *E*$14.11 in NP, and second-movers earned an average of *E*$22.03 in CNP, *E*$21.98 in S, and *E*$21.27 in NP.[38]

Of course, the earnings of *B*'s were higher in large part because some *B*'s were untrustworthy. Looking at those cases in the NP 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 the terms of agreements, while the remaining 90% is due to the violation of agreements. In the CNP 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.[39]

Our data suggest that the tendency to violate agreements is not random, but reflects real heterogeneity within the subject pool. What type of individuals violated their agreements? Looking at their pre-lab survey responses, we checked the scores of trustees who honored all agreements they entered vs. those who failed to honor at least one agreement, on a much-used psychological mea-

[38] The differences in trustor earnings are not statistically significant for conditions S versus NP, but are significant at the 1% level for conditions S versus CNP and NP versus CNP. Differences in trustee earnings are not significant between any pair of conditions. Differences between trustor and trustee earnings under each condition are significant at the 1% level.

[39] 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 $(y-x)/y$.

sure of opportunism, the Machiavellian scale.[40] Among *B*'s who at least once reached both a verbal and numerical agreement in the CNP condition on which the counterpart sent the agreed amount, the average score was higher, indicating more opportunism, among the ones who defected one or more times than among those who never defected.[41] Interestingly, there was no difference between the two groups in terms of performance on a test of cognitive ability, the Wonderlic Personnel Test. Moreover, there are no significant differences in Machiavellian scores when considering violation of merely numerical agreements.[42]

The most prominent difference between behavior in the CNP condition and in the S and NP 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. This percentage is 30.0% in S, 33.8% in NP, and 68.3% in CNP. 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 agreements (93.4%) are to the "fair and efficient" exchange. "Fair and efficient" agreements are Pareto optimal but not Nash equilibria. 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, as 99.2% of trustors did. Each time she honored a non-binding agreement with an anonymous partner whose identity she would never know, a trustee 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 one-shot interactions with anonymous counterparts?

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 trustees 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 economic theory, and they 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.

Why did text communication elicit so much more trusting and trustworthiness in our subjects than exchange of numerical proposals? 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."[43] 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 the not-quite-significant indications that these agreements also increased 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 their review of the economic literature on communication, 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." In footnotes, however, they note 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 comment: "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)." Our results are part of a growing body of evidence suggesting that if economists are to understand the impact of communication, these considerations must find their way out of the footnotes and into the text. We also show that there may be some value to analyzing what individuals say in order to better understand what they do.

Acknowledgements

The research reported here was funded by a grant from the Russell Sage Foundation. We are indebted to the late John Dickhaut for early advice on our design, to Freyr Halldorsson for help in planning and carrying out the experiments and to Bruno Garcia for help in coding chat content. We thank Pedro dal Bó and Gary Charness for helpful comments. A longer version of this paper (including more detailed analyses of the data) is available in our working paper.

References

Andreoni, J., 2005. Trust, Reciprocity, and Contract Enforcement: Experiments on Satisfaction Guaranteed, Working Paper. University of Wisconsin.
Ben-Ner, A., Halldorrsson, F., 2010. Trusting and trustworthiness: what are they, how to measure them, and what affects them. Journal of Economic Psychology 31, 64–79.
Ben-Ner, A., Putterman, L., 2001. Trusting and trustworthiness. Boston University Law Review 81 (3), 523–551.
Ben-Ner, A., Putterman, L., 2009. Trust, communication and contracts: an experiment. Journal of Economic Behavior and Organization 70 (1–2), 106–121.
Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity and social history. Games and Economic Behavior 10, 122–142.
Bicchieri, C., Lev-On, A., Chavez, A., 2010. The medium or the message? Communication relevance and richness in trust games. Synthese 176, 125–147.
Bochet, O., Putterman, L., 2009. Not just babble: a voluntary contribution experiment with iterative numerical messages. European Economic Review 53, 309–326.
Bohnet, I., Zeckhauser, R., 2004. Trust, risk and betrayal. Journal of Economic Behavior and Organization 55 (4), 467–484.
Boyd, R., Richerson, P., 1985. Culture and the Evolutionary Process. University of Chicago Press, Chicago.
Buchan, N., Croson, R., Johnson, E., 2006. Let's get personal: an international examination of the influence of communication, culture and social distance on

[40] The Mach-IV scale measures the subject's level of agreement with 20 statements. Christie and Geiss (1970) summarize early Machiavellianism research and describe the scale. For an example of its use in economics, see Gunnthursdottir et al. (2002). Note that there are too few violations of agreements by trustors in our experiment to support a parallel analysis for *A*'s.

[41] Significant at 10% level in a two-tailed Mann–Whitney test. If a one-tailed test of the hypothesis that defectors have higher Mach scores were used, the significance level would be 5%. There were 80 *B*'s satisfying the criterion of having at least one CNP verbal and numerical agreement carried out by their counterpart, of whom 64 never violated and 16 violated at least one such agreement.

[42] A detailed analysis of individual background variables in relationship to trusting (amount sent by *A*) and trustworthiness (proportion returned by *B*) is presented in Ben-Ner and Halldorrsson (2010).

[43] As noted earlier, textual promises do not have the same effect when subjects know them to be pre-written by the experimenter and merely selected by the sender from a short list of options, as in Charness and Dufwenberg (2007) and Bochet and Putterman (2009). This suggests that the power of messages requires that senders are able to construct them freely.

other regarding preferences. Journal of Economic Behavior and Organization 60, 373–398.
Camerer, C., 2003. Behavioral Game Theory. Princeton University Press, Princeton, NJ.
Charness, G., Dufwenberg, M., 2006. Promises and partnerships. Econometrica 74 (6), 1579–1601.
Charness, G., Dufwenberg, M., 2007. Broken Promises: An Experiment. Unpublished paper. UCSB and University of Arizona.
Christie, R., Geiss, F., 1970. Studies in Machiavellianism. Academic Press, New York.
Cox, J., 2004. How to identify trust and reciprocity. Games and Economic Behavior 46, 260–281.
Cox, J.C., Friedman, D., Gjerstad, S., 2007. A tractable model of reciprocity and fairness. Games and Economic Behavior 59 (1), 17–45.
Eckel, C., Wilson, R., 2004. Is trust a risky decision? Journal of Economic Behavior and Organization 55 (4), 447–465.
Eckel, C., Wilson, R., 2006. Internet cautions: experimental games with internet partners. Experimental Economics 9, 53–66.
Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games and Economic Behavior 54 (2), 293–315.
Farrell, J., Rabin, M., 1996. Cheap talk. Journal of Economic Perspectives 10 (3), 103–118.
Fehr, E., Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. Journal of Economic Perspectives 14 (3), 159–181.
Fehr, E., List, J., 2004. The hidden costs and returns of incentives—trust and trustworthiness among CEOs. Journal of the European Economic Association 2 (5), 743–771.
Fehr, E., Rockenbach, B., 2003. Detrimental effects of sanctions on human altruism. Nature 422, 137–140.
Field, A., 2001. Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity. University of Michigan Press, Ann Arbor.
Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters 71, 397–404.
Gintis, H., Bowles, S., Boyd, R., Fehr, E. (Eds.), 2005. Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life. MIT Press, Cambridge, MA.
Glaeser, E., Laibson, D., Scheinkman, J., Soutter, C., 2000. Measuring trust. The Quarterly Journal of Economics 65, 811–846.
Gneezy S U., 2005. Deception: the role of consequences. American Economic Review 95, 294–384.
Gunnthursdottir, A., McCabe, K., Smith, V., 2002. Using the Machiavellian instrument to predict trustworthiness in a bargaining game. Journal of Economic Psychology 23, 49–66.
Hoffman, E., McCabe, K., Smith, V., 1998. Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. Economic Inquiry 36, 335–352.
Houser, D., Xiao, E., McCabe, K., Smith, V., 2008. When punishment fails: research on sanctions, intentions and non-cooperation. Games and Economic Behavior 62, 509–532.
James, H.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.
Kandori, M., 1992. Social norms and community enforcement. Review of Economic Studies 59 (1), 63–80.
Kreps, D., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in finitely repeated prisoners' dilemma. Journal of Economic Theory 27, 245–252.
Kurzban, R., Houser, D., 2001. Individual differences in cooperation in a circular public goods game. European Journal of Personality 15 (S1), S37–S52.
Malhotra, D., Murnighan, J.K., 2002. The effects of contracts on interpersonal trust. Administrative Science Quarterly 47 (3), 534–559.
Ortmann, A., Fitzgerald, J., Boeing, C., 2000. Trust, reciprocity, and social history: a re-examination. Experimental Economics 3 (1), 81–100.
Page, T., Putterman, L., Unel, B., 2005. Voluntary association in public goods experiments: reciprocity, mimicry and efficiency. Economic Journal 115, 1032–1053.
Rigdon, M., 2005. Trust and Reciprocity in Incentive Contracting. University of Michigan.
Sanchez-Pages, S., Vorsatz, M., 2007. An experiment on truth telling in a sender-receiver game. Games and Economic Behavior 61, 86–112.
Servátka, M., Tucker, S., Vadovič, R., 2008. Words Speak Louder Than Money. Working Paper No. 18/2008. Dept. of Economics, University of Canterbury.
Willinger, M., Keser, C., Lohmann, C., Usunier, J.-C., 2003. A comparison of trust and reciprocity between France and Germany: experimental investigation based on the investment game. Journal of Economic Psychology 24 (4), 447–466.

Glossary of terms and notations

S: trust game—conventional Berg et al. (1995)
NP: one numerical proposal per subject *A* and subject *B*, followed by S condition
CNP: 1 minute chat followed by NP condition
S-NP: experimental design whereby the first five interactions are S and the next five interactions are NP
NP-CNP: experimental design whereby the first five interactions are NP and the next five interactions are CNP
NP-S: experimental design whereby the first five interactions are NP and the next five interactions are S
S-CNP: experimental design whereby the first five interactions are S and the next five interactions are CNP
CNP-S: experimental design whereby the first five interactions are CNP and the next five interactions are S
X_a*:* amount sent by subject in role *A* (trustor)
X_b*:* amount returned by subject in role *B* (trustee)
π_a*: A*'s experiment earnings in a single interaction
π_b*: B*'s experiment earnings in a single interaction
NA: Numerical agreement: *A*'s and *B*'s numerical proposals coincide
VA: A and *B* agree on a specific course of action