Not just babble: Opening the black box of communication in a voluntary contribution experiment

Olivier Bochet\textsuperscript{a,b}, Louis Putterman\textsuperscript{c,*}

\textsuperscript{a} Department of Economics, University of Bern, Bern, Switzerland
\textsuperscript{b} Department of Economics, Maastricht University, Maastricht, The Netherlands
\textsuperscript{c} Department of Economics, Brown University, Box B, Providence, RI 02912, USA

\textbf{Abstract}

We let subjects in a voluntary contribution experiment make non-binding numerical announcements about their “possible” contributions and, in some treatments, send written promises to contribute specific amounts. We find that announcements were responded to both by others’ announcements and by real play, for example announcements led to costly punishment when found to be misleading. We also find that adding pre-play announcements to treatments with punishment can increase efficiency by letting cost-free warnings substitute for costly punishment. The threat of punishing false announcements and promises helps reduce false signals, but only when promise statements can be sent is the effect sufficient for achieving greater efficiency.

\section{0. Introduction}

Many experiments have been conducted to study to what degree and under what conditions individuals free ride in the voluntary provision of a public good. The question is much studied in economic theory and is relevant to problems ranging from the making and soliciting of charitable contributions to environmental protection, provision of effort in partnerships, and work in teams. In experiments, contributions typically begin at an average level well above predicted free riding, in fact at more than 50\% of subjects’ endowments, but they decline steadily with repetition. Mechanisms that have been found to reduce free riding include taxing low contributors and rewarding high ones (Falkinger et al., 2000), allowing subjects to impose costly earnings reductions on one another (Fehr and Gächter, 2000a), and excluding free riders from playing with more cooperative subjects (Gunnthorsdottir et al., 2007; Önes and Putterman, 2007; Gächter and Thöni, 2005). However, in a review of 37 VCM, prisoners’ dilemma, and other social dilemma studies, Sally (1995) found pre-play communication to be the single most effective way to promote cooperation, and in a direct comparison under controlled conditions, Bochet et al. (hereafter BPP, 2006) found not only that pre-play face-to-face communication increased
contributions and earnings far more than did opportunities to sanction, but also that it was so effective that adding sanction opportunities to it led to no further improvement in outcomes.1

BPP also reported VCM experiments with two other kinds of communication. First, they conducted chat room treatments in which subjects could communicate with the members of their group on line only, while maintaining anonymity as to who was in one's group. Second, they carried out “numerical cheap talk” (NCT) treatments in which subjects (also anonymous to one another) could announce, by typing a number, a “possible amount” that they might contribute to the group account. Chat room treatments yielded higher levels of cooperation than baseline, despite the unavailability of facial expression, vocal intonation, and body language as means of conveying intention, emotion, or other information. Unlike the treatments that allowed the exchange of written messages, however, the numerical announcement treatments did not enhance average cooperation relative to the no communication baseline. A similar experiment using numerical announcements, by Wilson and Sell (1997), also found them to be ineffective at engendering cooperation.

This paper attempts to shed further light on the difference of outcomes between numerical and verbal communication. We ask two main questions. First, is numerical communication truly cheap talk in the sense of being discounted by both senders and receivers, thus amounting to ineffective babble? Second, are Sally (1995) and BPP correct in their conjecture that a major reason for the efficacy of verbal communication is the ability to issue promises?

We explore the first question by carrying out a microanalysis of the data from BPP’s NCT treatment. We demonstrate that the patterns of numerical signals sent by subjects are far from random, and that the indifferent average results of numerical signaling mask a dispersion of outcomes that includes both groups that achieved greater cooperation than their most successful counterparts in treatments with no communication and groups that, due to opportunistic reliance on false signaling by some members, achieved even less cooperation than their least successful no communication counterparts. Both the coordination successes and the false signaling “disasters” indicate that subjects took numerical announcements as something other than cheap talk in a sense that we elaborate below.

We explore the second question by adding treatments in which subjects can elect to send non-binding promise statements as a follow-up to their numerical signals. Our results show that adding only a promise option had no impact on the level of average contribution compared to the baseline treatment. However, the option of sending promise messages significantly improved outcomes in those groups whose members were also given the opportunity to sanction one another with costly punishment. We perform the same type of microanalysis of promise treatment behaviors as of those in the simpler NCT treatments. The data exhibit similar patterns, but we find that individuals paid more attention to promises than to announced contributions, and that when punishment was allowed, broken promises were severely punished and thus the percentage of false promises declined with repetition, allowing promises to grow in credibility.

The paper proceeds as follows. In Section 1, we discuss the literature on public goods games and communication. Section 2 lays out the design of the public goods experiments with and without numerical communication, sanction option, and promise option, and discusses theoretical predictions. Section 3 analyzes the experiments, emphasizing the impact of communication content on both others’ communication and on binding decisions. Section 4 concludes the paper.

1. Voluntary contribution experiments and communication

The voluntary contribution mechanism is an n-person linear public goods game with the following structure. In each of one or more periods (we focus on games of finite repetition), each of n ≥ 3 individuals is endowed with a certain number of dollars, $E$, and must divide this between a private account and a group or public account. Money put in the group account is multiplied by a factor $\lambda$ (where $n > \lambda > 1$) and divided equally among the n group members. The earnings of member $i$ in a given period are

$$y_i = (E - C_i) + \lambda \sum_{j \neq i} (C_j)/n$$

(1)

where $C_i$ ($0 \leq C_i \leq E$) is individual $i$’s contribution to the group account and the summation is taken over all group members, $i$ included. Eq. (1) shows that all group members are better off if all contribute their full endowments to the group account than if they contribute nothing, but each individual is better off still if the others contribute but he does not. Efficiency, defined as the sum of earnings, is also highest when all contribute their full endowments. We focus on the symmetric case in which each has an equal endowment and information about endowments and payoff functions is common knowledge.

In a finitely repeated VCM game, the only subgame perfect equilibrium for rational individuals who care only to maximize their own payoffs and who have common knowledge of one another’s preferences (including knowledge of one another’s knowledge of this) is $C_i = 0$ for each $i$. While an outcome having $C_i = E$ for all players dominates it, there is no credible way to punish deviations from an agreement to contribute $E$, so communication cannot in principle alter the outcome.2 Communication can change the outcome only if (a) payoffs can be altered in a manner external to the game, for

---

1 Ostrom et al. (1992) also provided either communication or punishment opportunities or both in common pool resource games. Their findings are similar except that the combination of communication and punishment achieves better results than communication alone, in their experiment.

2 A strategy of cooperating if others cooperate unravels because it is in no player’s interest to cooperate in the last period.

---

Please cite this article as: Bochet, O., Putterman, L., Not just babble: Opening the black box of communication in a voluntary contribution experiment. European Economic Review (2008), doi:10.1016/j.euroecorev.2008.09.005
example if an agreement reached by communication can be supported by the threat of penalties imposed by a third party, such as the state, or if (b) we drop the common knowledge and/or payoff maximization assumptions, allowing that some players' objective functions do not coincide with their material payoffs and/or that some players entertain the belief that players with such preferences might be present.

Isaac and Walker (1988) found that pre-play communication led their experimental subjects to contribute considerably more to a public good. Their study is one of 37 papers reporting 130 experimental treatments whose results Sally (1995) entered in multivariate regressions investigating which treatment variables best account for differing levels of cooperation and free riding. Sally concluded that face-to-face communication was the single most effective of the treatment variables, which also included the size of the monetary gain from cooperation, the number of repetitions, the discipline from which student participants were drawn, and the use of suggestive instructions by the experimenter.

To understand better what lies behind the effects of face-to-face communication, Brosig et al. (hereafter BOW, 2003) and BPP conducted additional VCM experiments in which other forms of communication were substituted for face-to-face discussion. BOW’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. BPP’s alternative treatments likewise included a no-communication baseline, but in addition, they conducted a treatment in which subjects could communicate text messages in a chat room and one in which subjects could relay non-binding possible choices in numerical form, with time for iterative reactions before each binding decision stage. For each of the four communication variants, BPP conducted two kinds of public goods game—a standard VCM treatment, and one like Fehr and Gächter’s having a “punishment” stage.3

While some of BOW and BPP’s treatments achieved almost equally large efficiency gains as did their face-to-face communication treatments, BPP’s NCT treatments performed no differently on average than did their no communication baseline. In this paper, we reanalyze the BPP treatments at the level of individual subject behaviors. We demonstrate that numerical announcements were taken seriously by group members, affecting both subsequent announcements and subsequent binding play. For example, in the treatment with numerical communication and punishment stages, subjects announced larger punishments of those who announced smaller contributions, the same relationship as exhibited in binding play, and those at whom such announced punishments were targeted responded by raising their announced contributions, a reaction also seen in binding play. Actual contributions are significantly correlated with both own and others’ announced contributions, and actually contributing less than the amount announced tended to elicit actual costly punishment. A micro-analysis of announcements and binding decisions thus provides evidence that subjects did not understand their messages to be “cheap talk” in the full sense suggested by a theory of rational, self-interested agents with common knowledge of one another’s type.

When discussing why pre-play communication raises cooperation, contrary to standard economic theory, but why this effect is observed only when that communication has an open-ended written or oral component, we conjectured that the ability to make promises plays a major role in raising rates of cooperation in treatments with face-to-face, audio–video, and chat room communication. To test this conjecture, we designed additional treatments identical to NCT, except that after iterative numerical communication and before each binding choice, we let subjects select, or not, a message promising to contribute a specific amount to the group account. We analyze the resulting treatments, finding that allowing promises significantly increased both contributions and earnings, but only when punishment opportunities were available to keep promise breakers in line.

2. Experimental design and predictions

We discuss decisions made in 18 experimental sessions, in each of which 16 (in four of the sessions, 12) undergraduate subjects (a total of 272 subjects) made a series of contribution decisions in randomly assigned and anonymous groups of four that stayed together for a total of 10 periods of play. Each period involved simultaneous decisions by each subject on contributing to a group account versus a personal account, described by Eq. (1) above, with $E = 10 experimental dollars (hereafter $E10) and $E16 per period if they perfectly cooperated and $E20 if they all contributed nothing.4 Subjects were drawn from the entire Brown University undergraduate population (numbering some 5800 students), sat at terminals in a large room, and were unable to read one another’s screens or to communicate except in the treatments and manners indicated below. Three sessions were devoted to each of six different treatments (see Table 1), of which the first four are also discussed in BPP. In the baseline (B) treatment, the entire session consisted of 10 such decisions, after each of which subjects learned of one another’s individual contribution decisions. In the

---

3 BPP’s chat room communication treatment without punishment resembled that of Frohlich and Oppenheimer (1998), except that those authors used e-mail messages, which do not provide a continuing record of messages to all group members. BPP’s non-binding numerical communication treatment without punishment, which we labeled “numerical cheap talk,” resembled the numerical pre-announcement treatment of Wilson and Sell (1997), except that our subjects could react to one another’s announcements with new nonbinding 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.

4 An experimental dollar exchanged for 0.13 real dollars at the end of the session, and total earnings averaged about $25 for a 90-min session, including a $5 participation fee.
punishment or reduction (R) treatment, each contribution decision was followed by a stage in which subjects learned of the contributions of each of the others in their group and had an opportunity to reduce the earnings of one or more group members at a fixed cost of £0.25 to the punisher per £1 of earnings loss to the person punished. Individuals were informed only of the reductions they themselves received, without knowing the identities of the punishers. Subjects in all four treatments were identified to one another only by letters B, C and D which were randomly reassigned each period (as in Fehr and Gächter, 2000a), to prevent tracking of individual behaviors and thus reduce the tendency to carry out vendettas.

In the “numerical cheap talk” (NCT) and “numerical cheap talk with reduction opportunities” (NCTwR) treatments, so-called to refer to because of the expectation of standard theory that announcements would amount to no more than cheap talk, each set of binding contribution and reduction decisions was preceded by a period of announcements and amended announcements. During these periods in the NCT treatment, subjects simply entered an amount in the group account assignment box of a screen identical to the binding decision interface (reproduced in BPP) but for the heading “Communication Stage” and a different background color. Once each had entered some number and the four numbers were displayed to each group’s members, they were free to alter their announced numbers for up to 90 seconds (a smaller amount of time in later periods). In the communication stages of the NCTwR treatment, subjects first entered possible contribution amounts, then, viewing the amounts entered by each group member, entered possible reduction amounts. Once a subject saw the four contribution announcements and the total reduction of her own earnings announced by others, she was free to alter either her announced contributions or her announced reductions of others’ earnings for up to 90 seconds (again, a smaller amount of time in later periods).

After finding that average contributions and earnings in the B and NCT (the R and NCTwR) treatments were not significantly different, we designed a variation on the NCT and NCTwR treatments as a partial test of our conjecture that this might be attributable to unavailability of a way to make verbal commitments or promises. The new treatments are identical to the old ones, including a period of iterative numerical communication of non-binding “possible” choices. However, at the end of each period’s numerical communication stage and before its binding contribution stage, subjects were asked to choose between two statements. The first option read: “I promise to contribute ___ to the group account this period.” Depending upon the choice of the subject, the other group members would then be shown either the statement “A promises to contribute X to the group account” (where X = 0, 1, … 10) or “A chooses not to make a promise,” and likewise for subjects B, C and D. The instructions given the subjects refer to “choosing a statement” rather than to “making a promise.” We label the promise-including analogy of the B and NCT treatments NCTwP, and the promise-including analogue of the R and NCTwR treatments NCTwP&R.

A dilemma for us in designing these treatments was what, if anything, to tell the subjects about whether a promise was binding. If there were no statement about this and if a high proportion of subjects contributed the amounts typed into their promise statements, we would be unable to rule out the explanation that they adhered to their promises because they understood the rules of the experiment to require them to do so. To rule out the possibility that promises were effective because of such a misunderstanding, we included in the instructions about entering binding decisions the statement “If you have chosen to promise a specific amount, you can type that amount at this time, but the computer will not prevent you from typing in a different amount.” This statement carried its own danger, because it may have been viewed as the granting of “permission to lie” by the experimenter; in fact, when instructions were being read aloud, there were a few chuckles or raised eyebrows among the subjects at this point in every session. Our results must be read, then, with an awareness of the downside bias against the efficacy of promises that we may have introduced by expressly stating that promises were not binding.

Standard economic theory assuming strictly payoff-maximizing agents with common knowledge of this preference predicts that subjects will contribute nothing to the group account in the B treatment. As pointed out by Fehr and Gächter

---

5 The instructions about this were: “At the beginning of each period there will be a communication stage. You will type in a possible amount for your group assignment… at any time thereafter during this stage, you can adjust your possible assignment to the group account… You are not committed to any of the numbers you type in during this stage.”

6 The full instructions given to the subjects, including practice problems, are provided in Bochet et al. (2005).
standard theory also predicts that the opportunity to engage in costly punishment will not be made use of and will have no effect on the level of contributions, which will still be uniformly zero. The addition of an opportunity to enter a possible contribution or a possible contribution and possible reduction decisions into a non-binding communication field would also have no effect according to theory, assuming common knowledge of payoff maximizing type. Under that common knowledge assumption, each agent realizes that each other agent will contribute nothing to the group account and spend nothing on punishment, regardless of what numbers are communicated, and there is therefore no reason to pay any attention either to the numbers typed by others or to the numbers that one types oneself. A pure numerical babble is always an equilibrium.7

Suppose, instead, that our subjects believe that another type of agent, whose objectives include but are not limited to earning more money, is present in the subject pool with some non-negligible probability. Although unconditional altruists and individuals who experience a “warm glow” from contributing are among the potentially interesting possibilities (see Palfrey and Prisbrey, 1997, and references therein), we focus here on three possible “non-standard” preferences: Positive and negative reciprocity, and truth-telling. A positively reciprocating agent is one who prefers to cooperate if he believes that others are cooperating. A negatively reciprocating agent is willing to incur a cost to punish someone who exploits him by free riding.8 An agent with a preference for truth-telling can be seen either as obtaining additional utility from adhering to her word, or as suffering a loss of utility if she breaks her word.9 Finally, suppose that subjects begin with some prior beliefs about the proportions of such individuals who are present and adjust their choices during the course of play as they update those beliefs. Subjects thus enter into a Bayesian game of the type analyzed by Kreps et al. (1982) and Guttmann (2003).

If a group of reciprocators with optimistic expectations of one another’s type are grouped together in a basic VCM experiment such as our B treatment, it is possible that they will contribute all or most of their endowments on the first decision and that, with their favorable beliefs thus supported, they will continue to contribute most of their endowments (Gunnthorsdottir et al., 2007; Gächter and Thöni, 2005). Probably more typical is an encounter of subjects with differing degrees of reciprocity and differing initial beliefs. Upon seeing some low contributions, the reciprocators in such a group will begin to reduce their contributions to the group account, in the B treatment, leading to the gradual downward slide that is usually seen in finitely repeated VCM experiments. Allow the reciprocators to punish the free riders while maintaining their own high contributions, however, as in our R treatment, and contributions may stabilize or rise rather than fall, as is found by Fehr and Gächter (2000a), Masclet et al. (2003), BPP, Page et al. (2005), and Sefton et al. (2007).

Consider now what communication might add to the Bayesian story. The opportunity to send numerical signals would not necessarily be viewed as useless by subjects who believe reciprocators or truth-tellers are common. Suppose, for example, that a substantial proportion of subjects are reciprocators and truth-tellers, and that all subjects know this to be so, although they do not know which individuals are and which are not of these types. Then subjects with the relevant preferences might, by entering a high number, try to signal intentions to contribute their endowments conditional on others doing so, and if others seem to signal a similar intention, they might proceed to contribute in fact and see whether the others follow through. If the game includes punishment opportunities, the reciprocators might signal intentions to punish low contributors, and some might follow through with actual punishment when they see evidence of attempts to mislead by contributing less than announced. Opportunistic subjects whose only goal is to maximize their payoffs might also signal and act cooperatively in some periods, with the intention of later exploiting the credulity of fellow players by signaling an intention to contribute but not following through.

The predictions of the Bayesian model which allows for non-standard player types and of the standard model with common knowledge of universal payoff maximizing type are clearly quite different. The standard model implies that “numerical cheap talk” will be of no consequence, and might even be a stream of random numbers. The Bayesian model suggests that we should look for signs of attempts to coordinate, on the parts of some subjects, and of attempts to mislead, on the parts of others. BPP’s analysis, which showed that outcomes in the NCT (respectively, NCTwR) treatment were on average indistinguishable from those in the counterpart B (respectively, R) treatment, is apparently consistent with the standard prediction regarding communication; but it does not rule out the Bayesian one. A more micro-analytic look at messages, and at inter-group differences in outcomes, is required in order to see whether NCT was really babble or was instead a flow of meaningful messages between subjects who viewed one another’s preferences as an open question.

---

7 As shown in Farrell-Rabin (1996), conflict of interest among agents means that messages cannot be self-signaling or self-committing. In such a case, if agents maximize their payoffs and types are common knowledge, it is always consistent to treat cheap talk as meaningless. The finding that the messages do not seem to be babble, in our experiment, implies that the subjects do not believe all to be rational payoff maximizers.

8 See Fehr and Gächter (2000b) and Hoffman et al. (1998), who treat conditional cooperation and willingness to punish noncooperation as two sides of the same trait. Önes and Putterman (2007) consider that the relative degrees of positive and of negative reciprocity may differ from one reciprocator to another.

9 Sánchez-Pagés and Vorsatz (2007) find evidence for the presence of subjects with what they deem to be two distinct social preferences, lie-aversion and preference for truth-telling, in sender–receiver games with conflicting preferences. Charness and Dufwenberg (2006) suggest that players may keep their word due to a desire to avoid harming others by falsely altering their expectations. Bicchieri (2005) discusses the activation of a norm of promise-keeping.
3. Results and analysis

3.1. Contributions and earnings trends

Fig. 1 shows the average number of dollars contributed to the public good in the six treatments, by period. The pattern of contributions in the B treatment conforms well to expectations from the literature: A substantial average initial contribution followed by a generally declining trend. In the R treatment, as in the similar treatment in Fehr and Gaechter (2000a), contributions show no pattern of decay until the end of the session, an outcome that analysis shows to be attributable, at least in part, to the tendency of many subjects to impose costly punishment on low contributors. This tendency is not significantly less in evidence in the last period, suggesting that it is indeed attributable to a taste, rather than being undertaken to raise future earnings.

Average contribution and its trend in the NCT and NCTwP treatments resemble closely those of their no communication and no promise counterpart, the B treatment, and the same holds when comparing the NCTwR treatment to its counterpart treatment, R. Mann–Whitney tests confirm that average contributions over the 10 periods as a whole do not differ significantly as between the NCT, NCTwP and B treatments, or as between the NCTwR and R treatments (see Appendix, Table A.1). It appears on first inspection, then, that giving subjects the opportunity to announce possible decisions before each set of binding decisions made no difference to contributions. However, the combination of punishment opportunities, opportunities to make numerical announcements, and opportunities to send promise statements raised contributions in the NCTwP&R above those in the R and NCTwR treatments, significant at the 5% and 10% levels respectively, hence also above those of the other three treatments.

**Result 1.** Apart from the NCTwP&R treatment, average contributions and their trend over time differ significantly only between treatments without punishment opportunities and those with such opportunities. Contributions are higher on average and are more sustained in treatments with punishment. But among treatments with punishment, contributions are significantly higher in the NCTwP&R treatment.

---

10 Average contribution and earning trends in the B, R, NCT, and NCTwR treatments are shown in BPP and are reviewed here before our more detailed analysis, and for comparison with the new treatments, NCTwP and NCTwP&R.

11 Past results are surveyed in Davis and Holt (1993) and Ledyard (1995).

12 The failure of contributions to rise as steeply in the R treatment as they do in Fehr and Gaechter’s punishment condition could be due to minor differences in design. A similar tendency for the introduction of a punishment stage to stem the usual decaying trend but without significant upward trend is also found in other replications of Fehr and Gaechter, for example Carpenter and Matthews (2002). See also Nikiforakis and Normann (2008), who find that the trend of contributions depends on the effectiveness of punishment (that is, cost to the person punished).

13 Appendix Table A.2 shows average contributions by treatment. Since group members had no knowledge of what was occurring in any of the other groups in their session or other sessions, and group behaviors are thus statistically independent, our non-parametric tests use the group average contribution or the group average earning level of a subject, averaged over all 10 periods, as observations. Table A.2 shows average contributions by treatment. Since group members had no knowledge of what was occurring in any of the other groups in their session or other sessions, and group behaviors are thus statistically independent, our non-parametric tests use the group average contribution or the group average earning level of a subject, averaged over all 10 periods, as observations. The Appendix is available at http://www.econ.brown.edu/fac/Louis%5FPutterman/working/pdfs/AppendixNotJustBabble%204-24-07.pdf.
Fig. 2 shows average earnings in the six treatments and their trend over time. Neither availability of punishment nor that of numerical communication significantly affect average earnings, nor does adding promise opportunities alone do so.\textsuperscript{14} Even though contributions are more sustained in treatments with punishment, earnings are not higher in those treatments, because the costly punishment that induces the higher contributions cancels out the associated earnings gain.\textsuperscript{15} Earnings are, however, higher in the NCTwP&R treatment, where higher contributions are attained with less punishment (as will be shown later).

\textbf{Result 2.} Apart from the NCTwP&R treatment, average earnings do not differ significantly between treatments. Average earnings are higher in NCTwP&R than in all other treatments.

3.2. Differing dispersions of group outcomes

Although there was little difference between the B and NCT (R and NCTwR) treatments with respect to average behaviors, we noticed that there was greater variation among groups in treatments with NCT. Fig. 3 shows the variance of average contribution among groups in each treatment, by period. In the left panel, containing the treatments without reductions, the variances are noticeably higher for the two treatments that include NCT, in all periods. In the right panel, containing the treatments with reductions, variances are higher for NCTwP than for R in all but one period, but the variance in contributions is lowest in NCTwP&R except in two periods. Table 2 confirms that on average, the variance among groups is higher in the NCT and NCTwP treatments than in the B treatment, and higher in the NCTwR treatment than in the R treatment. Formal statistical tests are ruled out unless one considers the variance in each period to be independent of that in other periods; if we assume that to be the case, the variances are found to be significantly different (see Appendix Table A.3).\textsuperscript{16}

\textbf{Result 3.} Except when combined with promises and reductions, the introduction of numerical announcements increased the variance of average contributions per period across groups.

\textsuperscript{14} See the lower left portion of Appendix Table A.1 for the test outcomes, and the right column of Appendix Table A.2 for average earnings by treatment.

\textsuperscript{15} Cinyabuguma et al. (2004) show that about 20\% of earnings reductions in the R treatment were aimed at high rather than low contributors, and that absent these reductions and the immediate lowering of contributions that they led to, earnings would have been higher in the R than in the B treatment. Ertan et al. (forthcoming) show that earnings rise unambiguously compared with baseline treatments if only punishment of low contributors is permitted (as is the case in their experiments when subjects vote on what types of punishment to permit). On “misdirected” or “anti-social punishment,” see also Herrmann et al. (2008).

\textsuperscript{16} We performed both Mann–Whitney tests, which treat each of the 10 variances (one for each period) for each treatment symmetrically, and Wilcoxon tests, which pair the variances of two treatments to be compared for period 1, for period 2, etc. As Appendix Table A.3 reports, both tests conclude that variances are greater in NCT compared to B, in NCTwP compared to B, and in NCTwR compared to R, significant at the 1\% level. As an additional check, we graphed the average contributions of the three highest-contributing and of the three lowest-contributing groups in each pair of treatments, and found high (low) performers contributing more (less) in NCT and NCTwP than in B and NCTwR than in R in almost every period. Using either average contribution or average deviation from treatment average contribution or both, the differences are mildly statistically significant (despite having only 3 observations per group) for high groups in NCT vs. B and NCTwP vs. R, and for low groups in NCTwR vs. R. Both the highest and lowest performing groups in NCTwP&R contributed more than (for highest groups: more than or the same as) their R and NCTwR counterparts in every period, however, with these differences also being significant. See Figs. A1 and A2 and test details in the Appendix, Section 3.
3.3. Subjects’ announcements responded to others’ announcements

What can account for the greater variance of outcomes in groups with numerical communication? One possibility is that groups having members more inclined towards truth-telling and reciprocity achieved and sustained greater cooperation, whereas groups with higher proportions of opportunistic individuals had worse outcomes when announcements were possible than when they were not. For this conjecture to hold true, subjects would have to take announcements as more than random babble. The next three results will demonstrate this, beginning by showing that subjects responded to one another’s announced plans by changing their announcements, in much the same way as they responded to one another’s binding decisions by changing their subsequent binding decisions in treatments without communication.

Table 3 reports panel corrected standard error regressions which help to establish the proposition that subjects who found their announced possible contribution to be greater (less) than their group’s average tended to reduce (increase) it from iteration to iteration in a given period’s announcement stage. The change in announced contribution is the dependent variable, and in addition to the difference between own and others’ previous announced contribution, the subject’s own previous announced contribution is included. For comparison, parallel regressions on the relationship between actual contributions and actual contributions of the previous period were estimated for the B and R treatments (see the Appendix, Table A.4). For both actual contributions and announcements, the coefficients on the contribution difference are always negative, supporting the proposition, significant at the 5% level or better in all but one specification. Adjustment of contributions toward group average may be a reflection of conditionally cooperative behavioral tendencies. Due to concern over possible interdependence among contributions, which is only partially controlled by including fixed effects,

Table 2
Variance of average contribution, by treatment

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Average variance of average contribution across groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>3.245</td>
</tr>
<tr>
<td>NCT</td>
<td>7.251</td>
</tr>
<tr>
<td>NCTwP</td>
<td>9.541</td>
</tr>
<tr>
<td>R</td>
<td>5.128</td>
</tr>
<tr>
<td>NCTwR</td>
<td>9.833</td>
</tr>
<tr>
<td>NCTwP&amp;R</td>
<td>4.213</td>
</tr>
</tbody>
</table>

3.3. Subjects’ announcements responded to others’ announcements

What can account for the greater variance of outcomes in groups with numerical communication? One possibility is that groups having members more inclined towards truth-telling and reciprocity achieved and sustained greater cooperation, whereas groups with higher proportions of opportunistic individuals had worse outcomes when announcements were possible than when they were not. For this conjecture to hold true, subjects would have to take announcements as more than random babble. The next three results will demonstrate this, beginning by showing that subjects responded to one another’s announced plans by changing their announcements, in much the same way as they responded to one another’s binding decisions by changing their subsequent binding decisions in treatments without communication.

Table 3 reports panel corrected standard error regressions which help to establish the proposition that subjects who found their announced possible contribution to be greater (less) than their group’s average tended to reduce (increase) it from iteration to iteration in a given period’s announcement stage. The change in announced contribution is the dependent variable, and in addition to the difference between own and others’ previous announced contribution, the subject’s own previous announced contribution is included. For comparison, parallel regressions on the relationship between actual contributions and actual contributions of the previous period were estimated for the B and R treatments (see the Appendix, Table A.4). For both actual contributions and announcements, the coefficients on the contribution difference are always negative, supporting the proposition, significant at the 5% level or better in all but one specification. Adjustment of contributions toward group average may be a reflection of conditionally cooperative behavioral tendencies. Due to concern over possible interdependence among contributions, which is only partially controlled by including fixed effects,

17 Subjects who entered only one contribution announcement during a given period’s communication stage are treated as having an announcement change of zero.
18 Fischbacher et al. (2001) find that 50% of their subjects contribute more when they expect others also to do so, with a further 14% also doing this so long as average contributions remain below half of the endowment. Similar results are reported by Kurzban and Houser (2001).
19 In Table 3 and others to come later, we show regressions that include period or individual (elsewhere, group) fixed effects in order to control for possible interdependence of observations (exceptions are Tables 4 and 6 where we report a series of tobit regressions that contain group and period fixed
we also conducted non-parametric Spearman correlation tests of the correlations between the key variables and found the results to be supportive of the same tendency to adjust contributions or announced contributions toward the group average, significant at the 1% level for announced contributions (and at the 5% level for actual contributions; for details see the Appendix, Part 5).

Result 4. During communication periods, subjects in the NCT and NCTwR treatments adjusted their announced contributions in the direction of the average announced contributions of other group members, paralleling period-to-period movements in binding contributions in the B and R treatments.

A second common response seen in binding play is that when a group member contributes a substantially smaller amount than others, he or she tends to be targeted for punishment. We next check whether a parallel phenomenon is found in the interactions between announcements. The first column of Table 4 reports a random effects tobit regression in which the amount of announced reductions aimed at subject \( j \) is the dependent variable, and the absolute negative and positive deviations of \( j \)'s announced contribution from the average of other group members, and that average itself, are independent variables. The second column reports an otherwise identical tobit regression with group fixed effects. The coefficient on absolute negative deviation is positive and significant at the 1% level, while that on absolute positive deviation is negative and significant at the 10% level, indicating that subjects were assigned more (less) announced reductions when the difference between \( j \)’s contribution and the average contribution of other group members is positive (negative).

We present Table 3 and Table 4 below:

**Table 3**
Adjustment of announced contributions in response to differences from means, NCT and NCTwR treatments

<table>
<thead>
<tr>
<th>Dependent variable: First change of announced contribution by subject ( i )</th>
<th>NCT</th>
<th>NCTwR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>(1st announced contribution by ( i ))−(average 1st announced contribution by others in ( i )'s group)</td>
<td>−0.366</td>
<td>−0.311</td>
</tr>
<tr>
<td>[0.073]</td>
<td>[0.089]</td>
<td>[0.073]</td>
</tr>
<tr>
<td>1st announced contribution by ( i )</td>
<td>−0.343</td>
<td>−0.498</td>
</tr>
<tr>
<td>[0.078]</td>
<td>[0.102]</td>
<td>[0.080]</td>
</tr>
<tr>
<td>Constant</td>
<td>2.348</td>
<td>1.306</td>
</tr>
<tr>
<td>[0.635]</td>
<td>[1.311]</td>
<td>[0.774]</td>
</tr>
<tr>
<td>Individual fixed effects</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>[0.001]</td>
<td>[0.001]</td>
<td>[0.001]</td>
</tr>
<tr>
<td>Period fixed effects</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[0.691]</td>
<td>[0.319]</td>
<td>[0.019]</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.339</td>
<td>0.434</td>
</tr>
</tbody>
</table>

*Note: Panel corrected standard errors regression. \( N = 440 \), all regressions. In all regressions in this and the remaining tables, numbers in parentheses are standard-errors and numbers in square brackets are \( p \)-values. In the individual and period fixed effects rows, numbers in square brackets are \( p \)-values of tests of the joint significance of individual fixed effects, and period fixed effects, respectively.

We also conducted non-parametric Spearman correlation tests of the correlations between the key variables and found the results to be supportive of the same tendency to adjust contributions or announced contributions toward the group average, significant at the 1% level for announced contributions (and at the 5% level for actual contributions; for details see the Appendix, Part 5).

Result 4. During communication periods, subjects in the NCT and NCTwR treatments adjusted their announced contributions in the direction of the average announced contributions of other group members, paralleling period-to-period movements in binding contributions in the B and R treatments.

(footnote continued)

effects simultaneously). Inclusion of both period and individual effects simultaneously tends to reduce the significance levels of some coefficients, but never changes their signs and has relatively small impacts on their magnitudes. We report the results, where different, in footnotes. With respect to the estimated coefficient of the first explanatory variable in Table 3, the estimates with both fixed effects reduce its significance to the 10% level except in the case of Table 3 regression for the NCTwR treatment, in which the coefficient estimate becomes insignificant (as is already the case in column (5)).

See, for example, Table 5 in Fehr and Gächter (2000a), Table 3 in Cinyabuguma et al. (2006), and Table 2 in Önes and Putterman (2007).

Tobit estimation is used because there are numerous cases of zero punishment which constitute potentially censored observations. In particular, 251 of the 440 observations in the announced punishment regression and 282 of the 440 observations in the actual punishment regressions have zero values of the dependent variable. Following Fehr and Gächter, the negative deviation variable is assigned a value of zero if \( j \) contributed more than the average of other group members, and likewise for the positive deviation variable if \( j \) contributed less than the others’ average.

Both regressions also contain period fixed effects.

Please cite this article as: Bochet, O., Putterman, L., Not just babble: Opening the black box of communication in a voluntary contribution experiment. European Economic Review (2008), doi:10.1016/j.euroecorev.2008.09.005
punishment the further below (above) the average was their announced contribution. For comparison, the table’s third and fourth columns show parallel regressions, also using data from the NCTwR treatment, but this time data on actual, as opposed to announced, contributions and reductions. We find that the interactions of announced decisions follow a closely similar pattern to the interactions of actual decisions. The parallelism of the two regressions defies the idea that numbers communicated amounted to random noise. Again, we also conducted non-parametric correlation tests, which strongly confirm the relationship between absolute negative deviation of announced contribution and announced punishment; see the Appendix, Part 6, for details (The remaining columns of Table 4 are discussed in connection with Result 8, below.)

Result 5. The relationship between announced punishment and announced contribution replicates that between actual punishment and actual contribution: The less one’s (announced) contribution relative to others in one’s group, the more one is targeted for (announced) punishment.

Our last demonstration that subjects adjusted their announcements in response to one another’s announcements in a fashion paralleling changes in binding decisions involves the influence of announced punishment on subsequent announced contribution. Cinyabuguma et al. (2004, Table 2) and Önes and Putterman (2007, Table 3) show that in finitely repeated partner treatments like those studied here, low (high) contributing subjects who receive punishment tend to increase (reduce) their contributions in the next period. Table 5 shows panel corrected standard errors regressions in which the dependent variable is subject $j$’s initial change of contribution announcement during each communication period. The independent variables are the number of dollars by which others announce that they might reduce $j$’s earnings, interacted with dummy variables to distinguish those announcing their group’s highest contribution from others.23

Three specifications are estimated for the NCTwR treatment, three for NCTwP&R, differing in inclusion of fixed effects for individuals and/or periods. All of the first coefficients are positive and highly significant, suggesting that those who

---

**Table 4**
Announced and actual reductions as a function of announced and actual contribution deviations and the deviation of actual from announced contribution, NCTwR treatment

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Announced pun. “received” by $j$</th>
<th>Actual punishment received by $j$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Absolute negative deviation</td>
<td>1.741 (0.161)</td>
<td>1.761 (0.150)</td>
</tr>
<tr>
<td>Absolute positive deviation</td>
<td>−0.679 (0.362)</td>
<td>−0.868 (0.332)</td>
</tr>
<tr>
<td>Average contrib. ($j$ excluded)</td>
<td>−1.152 (0.318)</td>
<td>−1.393 (0.336)</td>
</tr>
<tr>
<td>Difference between announced and actual contrib.</td>
<td>0.122 (0.064)</td>
<td>0.169 (0.083)</td>
</tr>
<tr>
<td>Constant</td>
<td>4.619 (3.056)</td>
<td>2.535 (3.447)</td>
</tr>
<tr>
<td>Random effects tobit</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Tobit with group fixed effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>LR test $p$-value</td>
<td>0.001</td>
<td>1.000</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>−770.33</td>
<td>−705.33</td>
</tr>
</tbody>
</table>

Note: Tobit regressions and random effects tobit regressions with period fixed effects. $N = 440$, all regressions.

---

23 Hence, the first (second) variable is 0 if $i$ is (not) the highest announced contributor.
announced a less-than-maximum contribution increased their announcement by an average of around 25 (16) cents per dollar of announced punishment “received” in the NCTwR (NCTwP&R) treatment, a reaction qualitatively identical and quantitatively similar to that found for actual contributions following actual punishment. The second coefficient is consistently negative, suggesting that a targeted high contributor might slightly reduce her announcement, as also found for contributions in the studies cited above, but this coefficient is less statistically significant in NCTwR, and insignificant in NCTwP&R.24 Corresponding non-parametric tests of the correlation between the first change of j’s announced contribution and the amount of punishment aimed at j if j is not the group’s highest contributor, by period, are reported in the Appendix, Part 7, and are generally significant and positive for the NCTwR treatment, but more mixed for NCTwP&R treatment.

Result 6. In response to announced possible punishment, subjects who announced low contributions tended to increase their announced possible contributions.

3.4. Subjects’ binding decisions also responded to others’ announcements

Did announcements influence other announcements only, or did they influence binding, costly decisions? In this section, we demonstrate that the latter is the case.

First, both regressions and non-parametric statistical tests show that binding contributions are significantly influenced by announced possible contributions. We estimate panel corrected standard error regressions in which subject i’s contribution in period t, t = 2, ..., 9, is the dependent variable, and the independent variables are the average contribution by the others in i’s group in period t–1, the average last announced contribution of the others in the period t communication stage, and i’s last announced contribution in that stage (see Appendix Table A.9).25 As before, three specifications are estimated for each of two treatments, varying with regard to what fixed effects are included. Own announcement is significantly related to own actual contribution for both NCT and NCTwR subjects, suggesting a tendency by most subjects to make more-or-less truthful announcements (whether out of genuine aversion to lying or to contribute to others’ beliefs—which may redound to one’s own benefit—that honest types may be present). In both treatments, actual last period average contribution by others significantly and positively affects own contribution. Finally, and most importantly, average of others’ most recent announced contribution has a positive coefficient in all of the regressions, is significant at the 10% level in one of the NCT and one of the NCTwR regressions, and is significant at the 1% level in the other

24 Coefficients on both variables change little if both individual and period fixed effects are used, although the coefficient on the second explanatory variable falls somewhat short of the 10% significance cut-off for both treatments, in this case.
25 Period 1 must be excluded to allow for the lagged average contribution term, and period 10 is left out to exclude potential end-game effects.
two NCTwR regressions.\textsuperscript{26} Corresponding non-parametric tests, almost all of which produce statistically significant results, are reported in Appendix Tables A.10 and A.11.

**Result 7.** In the NCT and NCTwR treatments, actual contributions in a period are positively related to own announced contribution and to the average of others’ announced contributions.

A still more striking indication that “cheap talk” announcements had impacts on costly binding decisions is the finding that many subjects incurred real monetary costs to punish “lying” about contribution “intentions.”\textsuperscript{27} The last two columns of Table 4 report tobit regressions with the same specification as the middle two (discussed in connection with Result 5, above), except that the difference between j’s last announced contribution and his/her actual contribution in the same period is added as an independent variable. The new variable has a positive coefficient significant at the 10% level in the random effects tobit estimate and at the 5% level in the tobit estimate with group fixed effects. These estimates imply that for every one dollar of difference between announced and actual contribution, a subject received on average about 12–17 cents of punishment. This amount of punishment may not have sufficed to induce much more truth-telling, but it is important for our purposes because it demonstrates that rather than treating one another’s announcements as noise, many subjects predicated costly decisions on them. Non-parametric tests are also broadly consistent, although differing by period in level of significance (see the Appendix, Part 9).

**Result 8.** Subjects received costly punishment for contributing less than their announced “possible” contribution.

A likely reason why the NCTwP&R treatment succeeded where the NCTwP treatment did not is that more subjects were deterred from using the promise option opportunistically in NCTwP&R since other group members had the possibility of inflicting monetary costs on them were they to contribute less than the promised amount. We’ve just seen that false announcements were punished in the NCTwR treatment, yet contributions and earnings in that treatment did not exceed those in treatment R. Punishment of false promises in NCTwP&R seems likely to have made a decisive difference only if its magnitude was greater, which could conceivably be the case if broken promise statements engendered greater anger or indignation.

Table 6 reports a series of tobit regressions\textsuperscript{28} resembling the last specification in Table 4, with the addition of a dummy variable controlling for the choice of a promise, and a few changes made necessary by unusually high correlations among certain variables.\textsuperscript{29} Columns 1 and 2 contain the basic results paralleling Fehr and Gächter’s, indicating that subjects received more punishment the further below others’ average was their contribution.\textsuperscript{30} In the NCTwP treatment, subjects chose the promise statement in 78.4% of opportunities to do so; for NCTwP&R, the corresponding figure is 86.8%. In columns 3 and 4, we add the dummy variable for contributing less than promised and find it to have highly significant positive coefficients, implying that a subject received an average of between 4.69 and 4.99 experimental dollars of punishment if she broke a promise—although the dummy for making a promise also becomes significant and of opposite sign, rendering the net negative impact of making and breaking a promise some 40% smaller. In columns 5 and 6, we use instead a dummy variable which takes the value of 1 if the subject contributed less than her last “numerical cheap talk” announcement. This term also obtains highly significant positive coefficients, although the indicated average punishment is smaller, between \$3.48 and \$3.56. In columns 7 and 8 both dummy variables are included. In the random effects specification of column 7, both variables have significant coefficients, but in the fixed effects specification of column 8, only the promise dummy is significant. In both cases, the coefficients suggest at least twice as much punishment for failing to fulfill a promise as for failing to live up to an announcement.\textsuperscript{31}

These results strongly suggest that breaking a promise was likely to attract more punishment than was failing to fulfill an announcement. The net punishment per promise broken, estimated from columns 3 and 4 as \$2.61–\$2.78, can be compared to the monetary gain from breaking a promise, which averaged only \$2.87 per episode, suggesting that a promise breaker achieved little or no net gain on average.\textsuperscript{32} We also conducted non-parametric tests of the correlations

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
\textbf{Period} & \textbf{Average Punishment} \\
\hline
1 & \$2.61–\$2.78 \\
\hline
2 & \$2.87 \\
\hline
\end{tabular}
\caption{Average Punishment by Period}
\end{table}
between punishment received by \( i \) and (a) the negative deviation of \( i \)'s contribution from the group average, (b) the dollars, if any, by which \( i \) underfulfilled the amount he/she promised, and (c) the corresponding difference for underfulfilled announcements. Almost all correlations are statistically significant at the 1% level (see the Appendix, Part 10).

Result 9. Contributing less than the amount specified in a promise statement drew costly punishment of even larger amount than did contributing less than indicated as a "possible" contribution.

Results 4–9 suggest that many subjects attempted to use non-binding numerical announcements to coordinate on a more rewarding cooperative strategy. One reason why outcomes were not on average better in the treatments with numerical communication than in their counterpart treatments may be that in addition to such cooperation-seekers, there were also subjects who intentionally used misleading signals to improve their individual returns from free riding. We test this conjecture by testing whether groups in which there was less opportunistic "lying" about intentions had better outcomes than those in which there was more "lying". Our tests show that the abuse of announcements to mislead other subjects did indeed have a detrimental effect on cooperation.

Let "lie" denote the difference between a subject's last announced contribution and her binding contribution in a given period, and calculate a group's "average lie" by dividing the sum of "lies" by the number of group members.33 In Table 7, we

---

33 Although the extent of an individual’s "lie" might be defined as being equal to zero whenever his actual contribution exceeded his last announced contribution in the communication round, we let "lie" (in the few cases of this type) take negative values.

---

Please cite this article as: Bochet, O., Putterman, L., Not just babble: Opening the black box of communication in a voluntary contribution experiment. European Economic Review (2008), doi:10.1016/j.euroecorev.2008.09.005
present a series of panel corrected standard errors regressions for the NCT and NCTwR treatments, with one observation per group and period, using average contribution in period \( t \) as dependent variable. Independent variables are the average “lie” last period (period \( t-1 \)), the average last announcement in the communication stage of the current period (\( t \)), and in some specifications group or period fixed effects. The estimates support the idea that a previous period’s “lying” reduced the current period’s contributions, in both treatments.34

Result 10. Contributing less than announced led to lower future contributions in groups in the NCT and NCTwR treatments.

A similar analysis can be done for the NCTwP and NCTwP&R treatments, except that here effects of both the average gap between announced and actual contributions, and the average gap between promised and actual contributions, can be studied. Table 8 reports panel corrected standard error regressions at group level that parallel those in Table 7 but include both lagged variables and both the average last announced contribution and the average last promised contribution.35 The estimates for the NCTwP treatment support the idea that both failure to fulfill announcements and lying on promises reduced subsequent contributions, although only the coefficient on the first variable is statistically significant when group fixed effects are included (column (2)).36 Those for the NCTwP&R treatment indicate even larger negative effects of lying on promises, but the coefficient on this variable is again insignificant when group fixed effects are used, and the coefficients for lie on announcement are in this case positive, contrary to expectation.37 The high correlation between the magnitudes of the two kinds of promise breaking, and the likely persistence of behaviors over time within groups which is suggested by the significant group fixed effect coefficients and the higher \( R^2 \)’s in regressions (2) and (5), make it particularly useful in this case to check bi-variate correlations that use only one observation per group. We find that for the

Table 7
The negative impact of “lying” on group performance in NCT and NCTwR

<table>
<thead>
<tr>
<th>Dependent variable: Average contribution in group, period ( t )</th>
<th>NCT</th>
<th>NCTwR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Average “lie” in ( t-1 )</td>
<td>-0.617</td>
<td>-0.318</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.088)</td>
</tr>
<tr>
<td></td>
<td>[0.001]</td>
<td>[0.001]</td>
</tr>
<tr>
<td>Last average announcement in ( t )</td>
<td>0.726</td>
<td>0.381</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.118)</td>
</tr>
<tr>
<td></td>
<td>[0.001]</td>
<td>[0.001]</td>
</tr>
<tr>
<td>Constant</td>
<td>0.604</td>
<td>3.286</td>
</tr>
<tr>
<td></td>
<td>(0.434)</td>
<td>(1.084)</td>
</tr>
<tr>
<td></td>
<td>[0.164]</td>
<td>[0.002]</td>
</tr>
<tr>
<td>Group fixed effects</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Period fixed effects</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>[0.001]</td>
<td>[0.001]</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.631</td>
<td>0.796</td>
</tr>
<tr>
<td></td>
<td>0.736</td>
<td>0.852</td>
</tr>
</tbody>
</table>

Note: Panel corrected standard errors regression. \( N = 88 \), all regressions.

34 Estimated coefficients on average “lie” in \( t-1 \) are negative but insignificant when both group and period fixed effects are included. Non-parametric tests yield negative correlation coefficients in all periods, but most are insignificant, perhaps in part due to the small sample size. See the Appendix, part 11.

35 The table’s row headings denote a “lie” on announcement by “lie” on \( A’ \) and a “lie” on promise by “lie” on \( P’ \); the variables being measured as contribution amount announced or promised minus actual amount contributed. “Lie” on promise is defined as 0 for subjects not choosing the promise statement, and “average last promised contribution” is the average choice of those in a group selecting the promise statement, only, and is treated as 0 when no member of a group selected that statement.

36 The same holds when both group and period fixed effects are included.

37 The same holds when both group and period fixed effects are included.
NCTwP&R treatment the bi-variate correlation between average contribution and average last period lie on announcement is negative in all pairs of periods, and is of statistically significant magnitude for half of them. More generally, correlations between both average lie on promise and average lie on announcement are negative in all periods in both treatments (see the Appendix, Part 12).

**Result 11.** The greater the degree to which members contributed less than announced or promised, the smaller were average subsequent contributions in groups.

Together, Results 10 and 11 suggest that differences in the extent of false announcements and promises help to account for differences in achieved cooperation. The “flip side” of the finding that “lies” undermined cooperation is that honesty promoted it.

### 3.5. Punishment of broken promises kept “lying” in check in the NCTwP&R treatment

The facts that false promises were heavily punished (Result 9) and that fulfilled promises led to greater cooperation (Result 11) can explain the higher contributions in the NCTwP&R treatment provided that subjects responded to the punishment of false announcements and promises by being more truthful in their communications. By making communications more credible, this could make more credible the threat that free riding would be met by actual punishment. If free-riding could thus be deterred by cost-free announcements rather than costly punishment, contributions would rise without offsetting punishment costs. And with assurance that others would contribute more, conditionally-cooperative subjects would also contribute more.
It can be shown, in fact, that the punishment of “lying” seen in Tables 4 and 6 led to a closer correspondence between announcements and binding decisions on contributions in the NCTwR and especially in the NCTwP&R treatment. We graphed, for each of the four treatments with numerical communication, the evolution during the 10 periods of play of the average contribution initially announced in a period, the average final announced contribution in the same period, the average actual contribution, and for treatments with promise option, the average contribution amount promised by those choosing the promise statement (see Appendix Fig. A.3). For the NCT and NCTwP treatments, which lack punishment opportunities, announced possible contribution (and in NCTwP, promised contribution) is either roughly constant or increases from first to last announcement (and promise) of a period’s communication phase, but average actual contribution falls further and further below the announcements and promises as the experiment progresses. In the NCTwR treatment, we see no systematic tendency for promised amounts to either increase or decrease during a given communication stage, and the gap between promised and actual contribution grows more modestly with time, except for a last period jump. In the NCTwP&R treatment, however, first and last announced and promised amounts tend to converge as the experiment progresses, and actual contributions show no tendency to diverge from promised and announced amounts until period 10. Mann–Whitney tests confirm that the average gap between promised and actual contribution is less in NCTwP&R than in the NCTwP treatment, and that the average gap between announced and actual contribution is less in NCTwP&R than in both NCTwP and NCT treatments, with a less significant difference in gap size between the NCTwR and the NCTwP and NCT treatments (see the Appendix, Part 13).

Fig. 4 graphs by period, for the NCTwP and NCTwP&R treatments, the proportion of subjects who failed to contribute as much as their last announcement and the proportion (of those choosing a promise statement) who failed to contribute as much as promised. It shows a dramatic difference between the two treatments, with the proportion of subjects not fulfilling their announcements and promises starting slightly higher and rising steeply toward about 70% in the NCTwP treatment, whereas those proportions stay in the neighborhood of around 20% in the NCTwP&R treatment, with a small last period up-tick. Mann–Whitney tests show that both the average proportion of unfulfilled announcements and the average proportion of unfulfilled promises were significantly lower in NCTwP&R treatment groups (see the Appendix, Part 14).

Finally, we graphed the average cost of punishment per period due both to cost to punishers and money lost by punishment recipients in the R, NCTwR and the NCTwP&R treatments. We find total cost of punishment to be lower in NCTwP&R than in NCTwR in all but one period, with the gap tending to widen after period 3. Punishment costs are also higher in the R than in the NCTwP&R treatment in all but two periods (see Appendix Fig. A.4). MW tests find the average cost of punishment significantly higher in NCTwR than in NCTwP&R, with a p-value of 9%, but the differences in punishment costs between the other two treatments are not significant (see the Appendix, Part 15).38

The fact that the roughly 80% honest announcement and promise rates in the treatment were achieved at modest social cost of around £$1 per subject per period, plus the effectiveness of punishment threats in boosting the contributions of less

---

38 Interestingly, the average expenditure on punishment for the three groups that spend the least on it is £$3.83 per period in the NCTwR treatment versus 0 in the NCTwP&R treatment.
cooperative subjects and that of high contribution announcements in boosting the contributions of conditionally cooperative subjects, explains why earnings were considerably higher in the NCTwP&R treatment.

4. Discussion and conclusions

Because no cooperative equilibrium is possible in a finitely repeated public goods game with rational payoff maximizing agents having common knowledge of their types, standard economic theory implies that the addition of opportunities to announce possible contributions in a non-binding fashion before costly play will have no effect. Being devoid of potential efficacy, any numerical messages sent could just as well be meaningless babble.

Our earlier experiment with non-binding numerical communication appeared to confirm the expectation that such communication has no effect on play, insofar as average binding behaviors followed approximately the same patterns with as without announcement stages. However, by disaggregating the results to the level of individual groups and to within-group interactions, we see that non-binding announcements helped some groups to cooperate, while leading to a more complete break-down of cooperation in others. The increase in dispersion of group outcomes is evidently explained by between-group differences in the extent to which subjects misled other group members with false announcements. Subjects seemed to take their messages seriously, as evidenced by the fact that mutual adjustments of announced choices display the same qualitative patterns as does real play, and by the fact that real contributions and, in treatments with punishment opportunities, costly punishments are influenced by message content. It makes sense for opportunists to try to “set up” others so as to free ride on their contributions, but only if opportunists believe that their signals may be taken seriously. To such opportunists, “talk is cheap” in the lay person’s common sense of that phrase, but not in the more demanding sense of the kind of economic theory that assumes common knowledge and payoff maximization. That theory would have talk be uniformly ignored by fully rational agents, and individuals, knowing that their talk would be ignored, would waste no effort on issuing meaningful signals.

In BPP (2006), we had speculated that one reason why NCT was less effective overall than was verbal (including text) communication is that NCT prevented subjects from framing their announcements in the moral language of explicit promises. As a partial test of that conjecture, we conducted new experiments in which, in addition to typing “possible” decisions into the message space used in the “numerical cheap talk” treatments, subjects could select, or not, a statement promising to contribute a specific amount to the public good. This test was imperfect, because we explicitly told subjects that promises were not binding (rather than risk the possibility that promises would be fulfilled due to a misunderstanding of the experiment’s rules), stirring up cynicism of a kind less likely to arise when subjects make promises in a more spontaneous fashion. Nevertheless, the outcome supported the conjecture in one treatment, in which subjects could impose costly punishments. Many subjects heavily punished “lying” on promise statements, and accordingly promises became more truthful and more credible, permitting many groups to achieve high levels of cooperation partly through cost-free threats and using less costly punishment.

The goal of our research has been to shed light on why communication aids cooperation, despite the predictions of standard economic theory. Our experiments add weight to the evidence suggesting that (a) many decision-makers behave as if they were maximizing something other than their monetary payoff alone, and that (b) most decision-makers act as if they assume this to be the case. At least three “extended” or “non-standard” preferences may underlie the results of our own and similar experiments. The efficacy of written promises to contribute even in treatments without punishment suggests that many subjects get disutility from breaking their word and/or believe this to be true of others, in which case the exchanging of promises alters expectations about one another’s behaviors. Many may also get higher subjective payoffs from cooperating provided that others cooperate, so that what are prisoners’ dilemma payoffs in pecuniary terms are assurance game payoffs in the space of utilities (Guttman, 2003; Page et al., 2005). Finally, many subjects display a willingness to incur monetary costs in order to penalize free riders and those who deliberately mislead in their announcements and promises.

In the real world, people frequently do cooperate in matters of common interest. A skeptical view is that when businessmen, partners in political coalitions, and others get together to find common ground, they simply bargain over the terms of agreements and the penalties and other mechanisms they will put in place to make those agreements self-enforcing for rational, self-interested agents, avoiding any reliance on non-material preferences or norms. A more natural interpretation, however, is that such communication also allows parties to assess one another’s trustworthiness, or in the language of economic theory, the content of their utility functions. Giving one’s word alters subsequent play in part because some individuals can be counted on to penalize themselves, psychically, should they break such a bond, and because the promiser, knowing human nature, knows that retaliation for betrayal may go beyond what is in the pecuniary interest of the betrayed party.

39 In a parallel line of research, Ben-Ner et al. (2007) investigate and find significant impacts of numerical and chat room communication on trusting and trustworthiness in the two-person sequential dilemma game called the trust or investment game. Also see again Charness and Dufwenberg (2006) and Sánchez-Pagés and Vorsatz (2007).
Acknowledgements

Funding for the experiments reported here came from a grant from the MacArthur Foundation Network on Norms and Preferences, and from National Science Foundation Grant SES-0001769. We wish to thank Toby Page for collaborating on the design of the experiments. We wish to thank Arno Riedl for helpful discussions and comments. We thank again the many students who assisted in carrying out the experiments reported in Bochet et al. (2006), as well as Dennis Zachary Schubert and Dmitri Lemmerman, who programmed the new treatments.

References