

Appendix to “Not Just Babble: Opening the Black Box of Communication in a Voluntary Contribution Experiment,” by Olivier Bochet and Louis Putterman

1. Comparison of contributions and earnings across treatments

Table A.1. *p*-values of two-tailed Mann-Whitney tests of differences in group average contributions and group average earnings

	B	R	NCT	NCTwR	NCTwP	NCTwP&R
B	.	0.002 C(R)>C(B)	0.740	0.096 C(NCTwR)>C(B)	0.235	0.001 C(NCTwP&R)>C(B)
R	0.977	.	0.016 C(R)>C(NCT)	1	0.016 C(R)>C(NCTwP)	0.023 C(NCTwP&R)>C(R)
NCT	0.740	0.786	.	0.065 C(NCTwR)>C(NCT)	0.650	0.001 C(NCTwP&R)>C(NCT)
NCTwR	0.695	0.786	0.847	.	0.087 C(NCTwR)>C(NCTwP)	0.087 C(NCTwP&R)>C(NCTwR)
NCTwP	0.235	0.832	0.699	0.898	.	0.001 C(NCTwP&R)>C(NCTwP)
NCTwP &R	0.037 E(NCTwP&R)>E(B)	0.051 E(NCTwP&R)>E(R)	0.023 E(NCTwP &R)>E(NCT)	0.040 E(NCTwP&R)>E(NCTwR)	0.013 E(NCTwP&R)>E(NCTwP)	.

Note: Numbers to the right and above the diagonal are for tests of differences in mean contributions. Numbers to the left and below the diagonal are for tests of differences in mean earnings. A statistically significant test *p*-value is indicated by bold typeface, and the direction of the observed difference is reported below it. The number of observations per treatment equals the number of groups (see Bochet and Putterman, Table 1).

Table A.2. Average contributions and average earnings across groups

Treatment	Average Contribution	Average Earnings
B	4.752	12.851
R	6.935	12.520
NCT	4.684	12.850
NCTwR	6.722	12.485
NCTwP	4.515	12.717
NCTwP&R	8.641	14.153

2. Tests of differences in inter-group variances

Table A.3. *p*-values of two-tailed Mann-Whitney and Wilcoxon tests of differences in variance of contributions across groups

	B vs NCT	B vs. NCTwP	R vs NCTwR	R vs NCTwP&R
Mann-Whitney test	0.001 var(C(NCT))> var(C(B))	0.001 var(C(NCTwP))> var(C(B))	0.001 var(C(NCTwR))> var(C(R))	0.218
Wilcoxon test	0.005 var(C(NCT))> var(C(B))	0.005 var(C(NCTwP))> var(C(B))	0.007 var(C(NCTwR))> var(C(R))	0.139

	NCT vs NCTwP	NCTwR vs. NCTwP&R
Mann-Whitney test	0.049 var(C(NCTwP))> var(C(NCT))	0.001 var(C(NCTwR))> var(C(NCTwP&R))
Wilcoxon test	0.093 var(C(NCTwP))> var(C(NCT))	0.007 var(C(NCTwR))> var(C(NCTwP&R))

Note: The number in each cell is the *p*-value of the respective test. Each observation is a variance of contributions among groups in a given period.

3. Further indicators of dispersion of group outcomes in treatments with and those without NCT

Here, we describe further tests of the observation that the dispersion of contribution outcomes among groups is greater in those treatments that include numerical communication than in those without pre-play communication, with the exception that low performing groups in the **NCTwP&R** treatment have higher contributions than those in the other treatments.

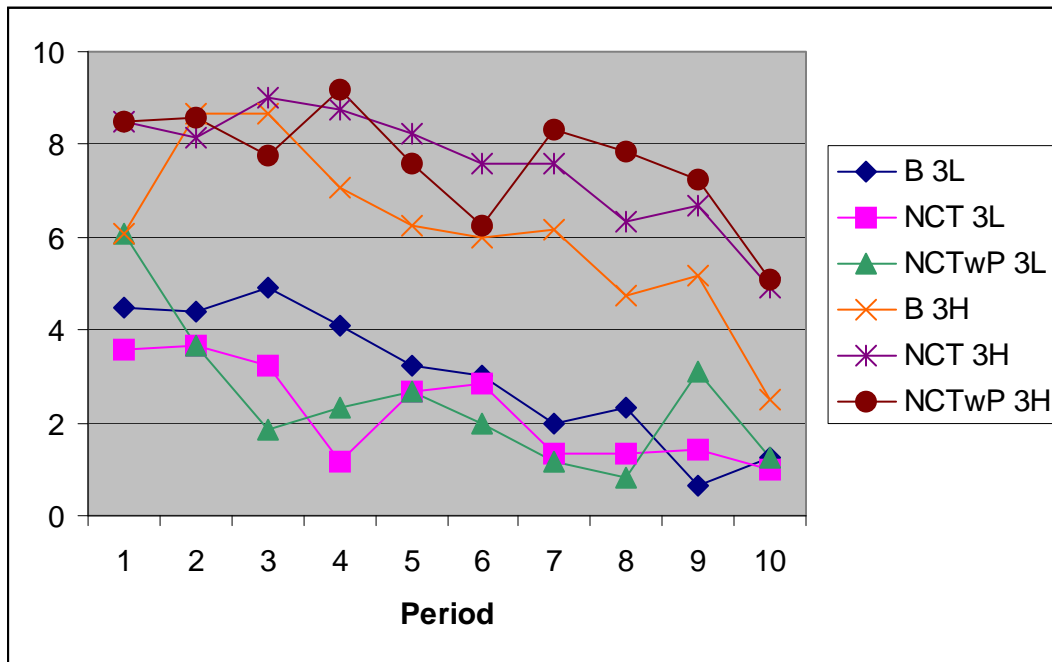
Figure A.1 plots average contribution by period for the three (of eleven) groups in the **NCT** treatment, the three (of twelve) groups in the **B** treatment, and the three (of eleven) groups in the **NCTwP** treatment that had the highest average contribution levels, and for the three groups in each of the same treatments with the lowest contribution levels.¹ For both highest and lowest groups, performance of **NCT** groups is more

¹ Each line follows the same three groups for the ten periods. That is, we identify the three groups which, on average, had the highest or lowest contributions over ten periods,

extreme than that of **B** treatment groups in all but one period (period 2, for the high groups; period 9, for the low groups). High and low groups in the **NCTwP** treatment also tend to be more extreme than their counterparts in the **B** treatment, although there are more violations for this comparison.

FIGURE A.1

Average contribution by period and by treatments, sorted by the three lowest and three highest groups



To test the significance of the observed differences, we can either look at the average contributions as shown in the figure, or calculate for the three high and three low groups of each treatment the absolute value of the difference between that group's average contribution and the average contribution of all groups in the treatment. We can run Mann-Whitney tests both of whether, for example, the three highest contributing groups in the **NCT** treatment had higher average contributions than the corresponding groups of the **B** treatment, or whether the differences from treatment mean are higher for the highest contributing groups in **NCT** than for those in the **B** treatment. Since there are

rather than graphing values for whichever three groups were highest or lowest in each individual period.

only three observations for each group (each being an average over ten periods), the power of these tests is low. Both the test for contributions themselves and the test for deviation from treatment mean find a significant difference among the three higher groups in **NCT** and those in **B**, although in both cases only at the 10% level and in a 1-tailed test. Contributions in the high groups in **NCTwP** are not significantly different from those in **B** or **NCT**, and there are no statistically significant differences for low groups in this figure.

FIGURE A.2

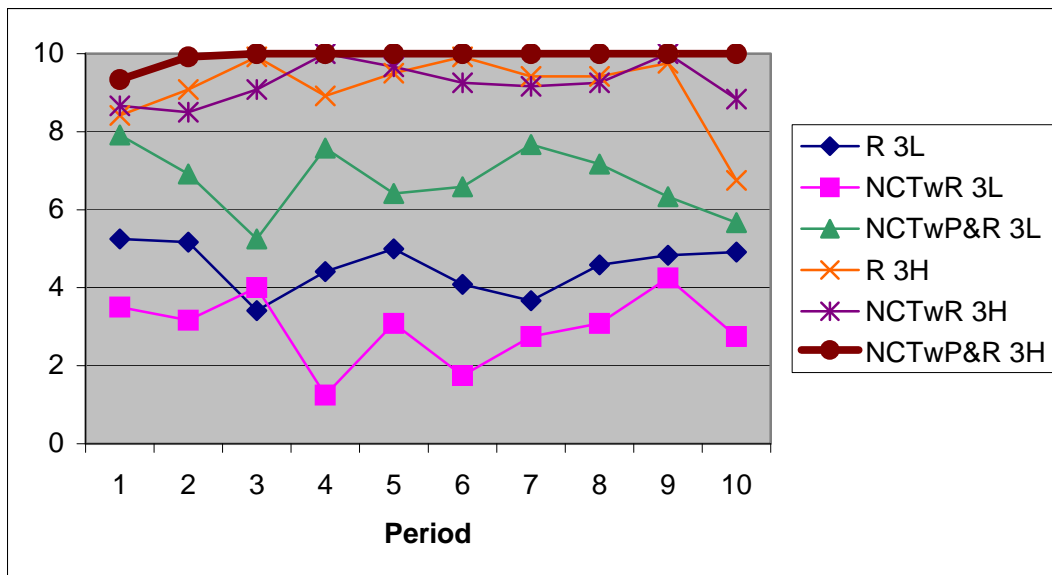


Figure A.2 provides corresponding comparisons for the 3 highest and 3 lowest contributing groups in each of the three treatments with reduction (punishment) opportunities: **R**, **NCTwR** and **NCTwP&R**. Among the low-contributing groups, the average contributions in the low-contributing groups in the **NCTwR** treatment are below those in the **R** treatment in all but one period, and this difference is significant at the 10% level in a 1-tailed test. The test for deviations from treatment mean does not find a difference in this case, however. Whereas the finding that low-contributing groups contributed less in **NCTwR** than in **R** is consistent with the general theme of contribution variances being higher among groups with cheap talk than groups without cheap talk, the lowest performing groups in the **NCTwP&R** treatment have *higher* contributions than those in either the **R** or the **NCTwR** treatment. The difference between the three lowest

NCTwP&R groups and the three lowest NCTwR groups is significant at the 10% level in 2-tailed tests, when either average contribution or average deviation from treatment mean is used. The lowest contributing groups contribute less in the NCTwR treatment than in the R treatment, significant at the 10% level in a 1-tailed test of means, but there is no difference using the test for deviations from treatment mean. Likewise, the difference between the low groups in NCTwP&R and those in the R treatment is significant only in the test of means, and only at the 10% level in a 1-tailed test.

Among the high contributing groups, the regularity that those with NCT have higher contributions than those without holds in these cases. High-contributing groups in the NCTwP&R treatment achieve 100% contribution in most periods. Both the test using means and the test using deviations from treatment means find a difference between the three high-contributing groups in the NCTwP&R treatment and those in the R treatment, significant at the 10% level in 2-tailed tests. The deviation from treatment mean contributions in the three high-contribution groups in NCTwR are higher than those in R, significant at the 10% level, although there is no difference found between those two treatments in the test for mean contributions. The difference between high-contributing groups in NCTwP&R and their counterparts in NCTwR is significant at the 10% level in 2-tailed tests both of average contributions and of deviations from treatment average contribution.

4. Dependence of contributions on difference of past contribution from group mean

Table A.4. Adjustment of contributions in response to differences from means, B and R treatments

Dependent variable: (contribution by subject i in period t)-(contribution by subject i in period $t-1$).

	B			R		
	(1)	(2)	(3)	(4)	(5)	(6)
(Contribution by i in period $t-1$) – (Average contribution by others in i 's group in period $t-1$)	-0.371 (0.067) [0.001]	-0.445 (0.077) [0.001]	-0.237 (0.072) [0.001]	-0.331 (0.059) [0.001]	-0.227 (0.287) [0.010]	-0.329 (0.058) [0.001]

Contribution by i in period $t-1$	-0.233 (0.066) [0.001]	-0.294 (0.075) [0.001]	-0.424 (0.080) [0.001]	-0.179 (0.054) [0.001]	-0.668 (0.107) [0.001]	-0.163 (0.052) [0.002]
Constant	0.703 (0.365) [0.054]	0.217 (0.805) [0.787]	3.396 (0.668) [0.001]	1.165 (0.401) [0.004]	1.216 (0.862) [0.158]	1.303 (0.494) [0.008]
Individual Fixed effects	No	Yes [0.004]	No	No	Yes [0.001]	No
Period fixed effects	No	No	Yes [0.001]	No	No	Yes [0.100]
R^2	0.303	0.370	0.359	0.246	0.433	0.261

Note: Panel corrected standard errors regression (using the `xtpcse` command in Stata). $N = 432$, all regressions. In all the regressions, numbers in parentheses are standard-errors and numbers in square brackets are p -values. In the individual and period fixed effects rows, number in square brackets are p -values of tests of the joint significance of individual fixed effects, and period fixed effects, respectively.

5. Non-parametric tests for Result 4.

Result 4 states that during communication stages, 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. Because the contributions of members of a given group cease to be fully independent variables after the first period, due to their being reported to the group members, the assumptions supporting the significance tests on regression coefficients backing this result may not be fully valid, despite the use of fixed effects and corrected standard errors. Therefore, in Tables A.5 and A.6 we present results of non-parametric Spearman correlation tests of the correlations between the key variables: changes in each subject's (announced) contribution, and difference between own and others' contribution in the previous period (announcement). The observations of a given period are tested separately, with those of period 10 left out due to possible end-game effects. All but one test for the actual contributions (Table A.5) are supportive at the 5% level of significance, while all of the tests for the announcements (Table A.6) are supportive at the 1% level.

Table A.5. Correlation between (change in *i*'s contribution between period *t* and *t-1*) and (difference between contribution of *i* in *t-1* and contribution of others in *i*'s group in *t-1*)

Period	2	3	4	5	6	7	8	9
B	$\rho = -0.562$ [0.001]	$\rho = -0.295$ [0.041]	$\rho = -0.601$ [0.001]	$\rho = -0.459$ [0.001]	$\rho = -0.526$ [0.001]	$\rho = -0.505$ [0.001]	$\rho = -0.715$ [0.001]	$\rho = -0.714$ [0.001]
R	$\rho = -0.492$ [0.001]	$\rho = -0.416$ [0.003]	$\rho = -0.680$ [0.001]	$\rho = -0.321$ [0.025]	$\rho = -0.546$ [0.001]	$\rho = -0.123$ [0.425]	$\rho = -0.548$ [0.001]	$\rho = -0.510$ [0.001]

Note: Non-parametric Spearman correlation tests. N = 48, all tests. Number ρ is Spearman's correlation coefficient. Numbers in square brackets are *p*-values. N is the number of paired observations. Subsequent correlation tables follow the same conventions.

Table A.6. Correlation between (first change of announced contribution by subject *i*) and (first announced contribution by *i* – average first announced contribution by others in *i*'s group)

Period	2	3	4	5	6	7	8	9
NCT	$\rho = -0.601$ [0.001]	$\rho = -0.370$ [0.013]	$\rho = -0.608$ [0.001]	$\rho = -0.663$ [0.001]	$\rho = -0.551$ [0.001]	$\rho = -0.511$ [0.001]	$\rho = -0.461$ [0.002]	$\rho = -0.479$ [0.001]
NCTwR	$\rho = -0.440$ [0.003]	$\rho = -0.397$ [0.008]	$\rho = -0.465$ [0.001]	$\rho = -0.572$ [0.001]	$\rho = -0.534$ [0.001]	$\rho = -0.424$ [0.004]	$\rho = -0.432$ [0.003]	$\rho = -0.595$ [0.001]

Note: Non-parametric Spearman correlation tests. N = 44, all tests.

6. Non-parametric tests for Result 5

Result 5 states that the relationship between announced punishment and announced contribution replicates that between actual punishment and actual contribution in that the less one contributes relative to others in one's group, the more one is targeted for punishment. Due to the potential problem of non-independence of observations, we conducted period by period non-parametric tests of the relationship between announced contribution and announced punishment, the results of which appear in the top row of Table A.7. These tests find a correlation between the absolute negative deviation of the announced contribution, and announced punishment, significant at the 1% level. Corresponding tests of the correlation between binding punishment and the absolute negative deviation of *i*'s binding contribution appear in the second row of the table; five of the nine correlations are significant at the 1% level and a sixth at the 5% level.

Table A.7. Correlation between the announced (respectively binding) punishment received by i and the absolute negative deviation by i from the average announced (respectively binding) contribution by others in i 's group, NCTwR treatment.

Period	1	2	3	4	5	6	7	8	9
Announced	$\rho = 0.483$ [0.001]	$\rho = 0.592$ [0.001]	$\rho = 0.490$ [0.001]	$\rho = 0.362$ [0.016]	$\rho = 0.549$ [0.001]	$\rho = 0.430$ [0.004]	$\rho = 0.738$ [0.001]	$\rho = 0.407$ [0.006]	$\rho = 0.537$ [0.001]
Binding	$\rho = 0.648$ [0.001]	$\rho = 0.592$ [0.001]	$\rho = 0.667$ [0.001]	$\rho = 0.237$ [0.122]	$\rho = 0.090$ [0.560]	$\rho = 0.757$ [0.001]	$\rho = 0.217$ [0.158]	$\rho = 0.578$ [0.001]	$\rho = 0.328$ [0.030]

Note: Non-parametric Spearman correlation tests. N = 44, all tests.

7. Non-parametric tests for Result 6

Result 6 says that in response to announced possible punishment, subjects who announced low contributions tended to increase their announced possible contributions. Again, we conducted non-parametric tests because of the possible non-independence of the observations used in our regression tests. Table A.8 shows the Spearman correlation between the first change of i 's announced contribution and the amount of punishment aimed at j if j is not the group's highest contributor, by period, along with its significance level. For the NCTwR treatment, the correlations for all but one period are statistically significant, mostly at the 1% level. For the NCTwP&R treatment, only periods 1 – 3 and 9 show significant positive correlations between the two variables; the low and in one case negative correlations in other periods may in part reflect the high contributions and low incidence of punishment in this treatment.

Table A.8. Correlation between first change of announced contribution by subject i and announced reduction threat (i not highest contributor)

Period	1	2	3	4	5	6	7	8	9
NCTwR	$\rho = 0.270$ [0.076]	$\rho = 0.459$ [0.002]	$\rho = 0.107$ [0.491]	$\rho = 0.454$ [0.002]	$\rho = 0.448$ [0.002]	$\rho = 0.458$ [0.002]	$\rho = 0.696$ [0.001]	$\rho = 0.414$ [0.005]	$\rho = 0.458$ [0.002]
NCTwP&R	$\rho = 0.493$ [0.001]	$\rho = 0.332$ [0.028]	$\rho = 0.486$ [0.001]	$\rho = 0.139$ [0.369]	$\rho = 0.223$ [0.145]	$\rho = -0.084$ [0.589]	$\rho = 0.020$ [0.900]	$\rho = 0.031$ [0.840]	$\rho = 0.381$ [0.011]

Note: Non-parametric Spearman correlation tests. N = 44, all tests.

8. Regression and non-parametric tests for Result 7.

Result 7 says that 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. We first show this by regression.

Table A.9. Actual contribution as a function of others' past and announced contributions, and own announced contribution, NCT and NCTwR treatments.

Dependent variable: period t contribution by subject i

	NCT			NCTwR		
	(1)	(2)	(3)	(4)	(5)	(6)
Average <i>binding</i> contribution of group members other than i in period $t-1$	0.434 (0.072) [0.001]	0.298 (0.087) [0.001]	0.366 (0.755) [0.001]	0.431 (0.063) [0.001]	0.140 (0.079) [0.078]	0.448 (0.065) [0.001]
Average last <i>announced</i> contribution of group members other than i in period t	0.142 (0.095) [0.135]	0.124 (0.113) [0.273]	0.168 (0.094) [0.074]	0.267 (0.077) [0.001]	0.146 (0.078) [0.063]	0.257 (0.077) [0.001]
i 's last announced contribution in period t	0.265 (0.076) [0.001]	0.298 (0.087) [0.001]	0.273 (0.073) [0.001]	0.397 (0.052) [0.001]	0.255 (0.064) [0.001]	0.397 (0.051) [0.001]
Constant	-0.684 (0.426) [0.109]	-2.588 (1.262) [0.040]	0.255 (0.658) [0.698]	-1.378 (0.496) [0.005]	4.440 (1.446) [0.001]	-1.035 (0.617) [0.093]
Individual fixed effects	No	Yes [0.001]	No	No	Yes [0.001]	No
Period fixed effects	No	No	Yes [0.135]	No	No	Yes [0.359]
R^2	0.245	0.589	0.268	0.500	0.684	0.510

Note: Panel corrected standard errors regression. $N = 352$, all regressions.

For the same reason as above, we also perform non-parametric tests for correlation between binding contribution of subject i and others' average announcements, and between binding contribution of subject i and the own announcement of subject i . Results are shown in Tables A.10 and A.11, respectively. Almost all of the correlations are statistically significant, most of them at high levels.

Table A.10. Correlation between binding contribution by subject i in period t and average last announced contribution of others in i 's group in period t

Period	2	3	4	5	6	7	8	9
NCT N=44	$\rho = 0.276$ [0.070]	$\rho = 0.466$ [0.001]	$\rho = 0.475$ [0.001]	$\rho = 0.382$ [0.011]	$\rho = 0.292$ [0.054]	$\rho = 0.508$ [0.001]	$\rho = 0.225$ [0.142]	$\rho = -0.157$ [0.308]
NCTwR N=44	$\rho = 0.483$ [0.001]	$\rho = 0.529$ [0.001]	$\rho = 0.741$ [0.001]	$\rho = 0.650$ [0.001]	$\rho = 0.649$ [0.001]	$\rho = 0.525$ [0.001]	$\rho = 0.626$ [0.001]	$\rho = 0.213$ [0.165]

Note: Non-parametric Spearman correlation tests. N = 44, all tests.

Table A.11. Correlation between binding contribution by subject i in period t and i 's last announced contribution in period t

Period	2	3	4	5	6	7	8	9
NCT N=44	$\rho = 0.343$ [0.023]	$\rho = 0.448$ [0.002]	$\rho = 0.407$ [0.006]	$\rho = 0.230$ [0.134]	$\rho = 0.468$ [0.001]	$\rho = 0.488$ [0.001]	$\rho = 0.277$ [0.069]	$\rho = 0.247$ [0.106]
NCTwR N=44	$\rho = 0.520$ [0.001]	$\rho = 0.491$ [0.001]	$\rho = 0.624$ [0.001]	$\rho = 0.704$ [0.001]	$\rho = 0.526$ [0.001]	$\rho = 0.557$ [0.001]	$\rho = 0.624$ [0.001]	$\rho = 0.485$ [0.001]

Note: Non-parametric Spearman correlation tests. N = 44, all tests.

9. Non-parametric tests for Result 8

Result 8 says that subjects received costly punishment for contributing less than their announced “possible” contribution. For the same reasons as above, we performed non-parametric tests for a correlation between the deviation between announced and actual contribution and actual punishment received, for early and later periods of the NCTwR treatment, using the individual-level observations of one period at a time. In all periods, the correlations (shown in Table A.12) are positive, and in five of the nine, they are statistically significant at the 10% level or better.

Table A.12. Correlation between the binding punishment received by i and the difference between announced and actual contribution made by i , NCTwR treatment.

Period	1	2	3	4	5	6	7	8	9
Difference between announced and actual contribution	$\rho = 0.280$ [0.066]	$\rho = 0.094$ [0.542]	$\rho = 0.490$ [0.001]	$\rho = 0.034$ [0.812]	$\rho = 0.337$ [0.025]	$\rho = 0.295$ [0.052]	$\rho = 0.292$ [0.054]	$\rho = 0.232$ [0.129]	$\rho = 0.029$ [0.851]

Note: Non-parametric Spearman correlation tests. N = 44, all tests.

10. Non-parametric tests for Result 9

Result 9 says that contributing less than the amount specified in a promise statement drew actual costly punishment in even larger amount than did contributing less than indicated as a “possible” amount. We conducted non-parametric tests of the correlations 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,² and (c) the corresponding difference for underfulfilled announced contributions. The results, shown in Table A.13, show that all correlations for all three variables are statistically significant at the 1% level.

Table A.13. Correlation between binding punishment received by i and negative deviation, broken promise and unfulfilled announcement, NCTwP&R treatment

Period	1	2	3	4	5	6	7	8	9
Negative deviation	$\rho = 0.529$ [0.001]	$\rho = 0.627$ [0.001]	$\rho = 0.579$ [0.001]	$\rho = 0.617$ [0.001]	$\rho = 0.484$ [0.001]	$\rho = 0.574$ [0.001]	$\rho = 0.683$ [0.001]	$\rho = 0.643$ [0.001]	$\rho = 0.434$ [0.003]
Broken promise	$\rho = 0.119$ [0.438]	$\rho = 0.486$ [0.001]	$\rho = 0.588$ [0.001]	$\rho = 0.645$ [0.001]	$\rho = 0.707$ [0.001]	$\rho = 0.546$ [0.001]	$\rho = 0.677$ [0.001]	$\rho = 0.637$ [0.001]	$\rho = 0.221$ [0.149]
Unfulfilled announcement	$\rho = 0.432$ [0.003]	$\rho = 0.467$ [0.001]	$\rho = 0.493$ [0.001]	$\rho = 0.469$ [0.001]	$\rho = 0.485$ [0.001]	$\rho = 0.422$ [0.004]	$\rho = 0.612$ [0.001]	$\rho = 0.437$ [0.003]	$\rho = 0.237$ [0.120]

Note: Non-parametric Spearman correlation tests. N = 44, all tests.

11. Non-parametric tests for Result 10.

Result 10 says that contributing less than announced led to lower future contributions in groups in the NCT and NCTwR treatments. For the same reasons as above we performed non-parametric tests of the correlation between average contribution in period t and average extent of “lying” in period $t - 1$, by group. The tests are shown in Table A.14. Although the correlations are negative in all periods, most are insignificant, perhaps in part due to the small sample size. Unlike most of the tables above, the tests in this table uses one observation per group rather than per subject. Observations from

² Treated as 0 if i did not select a promise statement.

different periods are not pooled because those of a given group in different periods may not be statistically independent.

Table A.14. Correlation between the average contribution of group i in period t and the average lie of group i in period $t-1$

Period	2	3	4	5	6	7	8	9
NCT	$\rho = -0.473$ [0.141]	$\rho = -0.447$ [0.168]	$\rho = -0.698$ [0.016]	$\rho = -0.391$ [0.233]	$\rho = -0.490$ [0.125]	$\rho = -0.146$ [0.668]	$\rho = -0.433$ [0.182]	$\rho = -0.255$ [0.449]
NCTwR	$\rho = -0.137$ [0.685]	$\rho = -0.346$ [0.296]	$\rho = -0.182$ [0.590]	$\rho = -0.278$ [0.407]	$\rho = -0.746$ [0.008]	$\rho = -0.127$ [0.708]	$\rho = -0.178$ [0.598]	$\rho = -0.664$ [0.025]

Note: Non-parametric Spearman correlation tests. N = 11, all tests.

12. Non-parametric tests for Result 11.

Result 11 says that the greater the degree to which members contributed less than announced or promised, the smaller were average subsequent contributions in their group. For the reasons above, non-parametric correlation tests are performed and are reported in Tables A.15 (for NCTwP treatment) and A.16 (for NCTwP&R treatment). They show negative correlation coefficients between both lie variables and next period contribution in every period of both treatments, with the same problem of small sample size and limited significance as with Table A.14, but with a larger number of significant correlations for each variable in the NCTwP&R treatment (Table A.16).

Table A.15. Correlation between the average contribution of group i in period t and the average lie (announcements and promises) of group i in period $t-1$, NCTwP treatment.

Period	2	3	4	5	6	7	8	9
Average lie on announcement, period t-1	$\rho = -0.305$ [0.361]	$\rho = -0.321$ [0.335]	$\rho = -0.150$ [0.658]	$\rho = -0.200$ [0.553]	$\rho = -0.220$ [0.514]	$\rho = -0.461$ [0.153]	$\rho = -0.247$ [0.462]	$\rho = -0.375$ [0.254]
Average lie on promises, period t-1	$\rho = -0.418$ [0.199]	$\rho = -0.746$ [0.008]	$\rho = -0.478$ [0.136]	$\rho = -0.482$ [0.132]	$\rho = -0.336$ [0.311]	$\rho = -0.393$ [0.231]	$\rho = -0.694$ [0.017]	$\rho = -0.479$ [0.135]

Note: Non-parametric Spearman correlation test. N = 11, all tests.

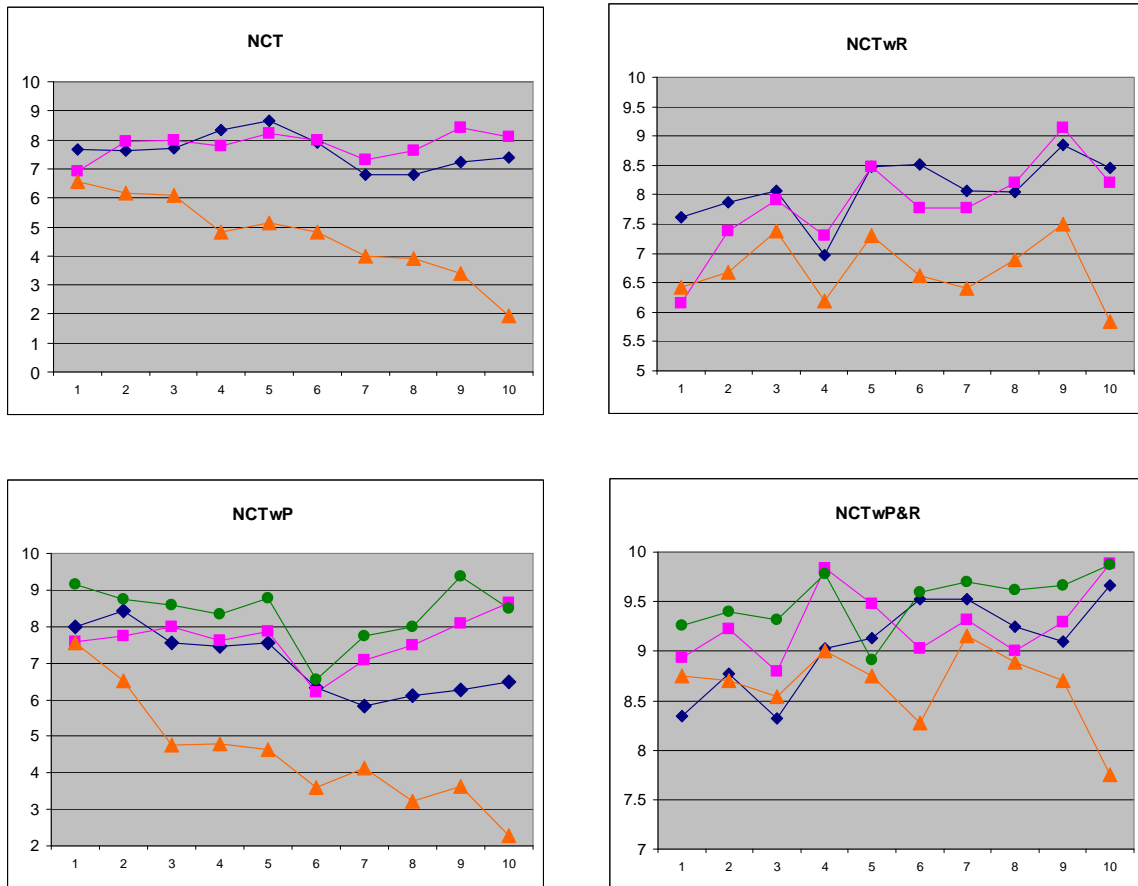
Table A.16. Correlation between the average contribution of group *i* in period *t* and the average lie (announcements and promises) of group *i* in period *t*-1, NCTwP&R treatment.

Period	2	3	4	5	6	7	8	9
Average lie on announcement, period <i>t</i> -1	$\rho = -0.220$ [0.514]	$\rho = -0.341$ [0.304]	$\rho = -0.674$ [0.022]	$\rho = -0.552$ [0.078]	$\rho = -0.538$ [0.087]	$\rho = -0.649$ [0.030]	$\rho = -0.207$ [0.540]	$\rho = -0.091$ [0.790]
Average lie on promises, period <i>t</i> -1	$\rho = -0.100$ [0.769]	$\rho = -0.583$ [0.059]	$\rho = -0.674$ [0.022]	$\rho = -0.743$ [0.008]	$\rho = -0.283$ [0.398]	$\rho = -0.828$ [0.001]	$\rho = -0.426$ [0.190]	$\rho = -0.686$ [0.019]

Note: Non-parametric Spearman correlation test. N = 11, all tests.

13. Figure A.3 and non-parametric tests of significance of difference in gap between announced and actual contribution.

Figure A.3 Average announcements, promises and binding contributions



Note: Triangles = Average contribution, Diamonds = Average first announcement, Squares = Average last announcement, Circles = Average promise. Notice that vertical scales differ as appropriate to each treatment's observations.

Using data at group level, we run MW tests to check whether the gap between average last announced and average actual contributions differs across treatments. We find that the average gaps are significantly lower for groups in **NCTwP&R** than for those in **NCT** and **NCTwP** treatments (2-tailed test p -values below 1%) but that the difference is not significant for groups in the **NCTwR** treatment. As for the gap between average promise and average actual contribution, it is significantly greater in **NCTwP** than in **NCTwP&R** (p -value in 2-tailed group level MW test equals .001). The results are summarized in Table A.17, below. Because gaps may not be entirely independent across periods, we also ran the same type tests but for each period separately. We find similar results. Interestingly, for each pair of treatments in which there is a significant difference, the associated per period tests are significant starting in Period 3 but not before.

Table A.17. p -values of two-tailed Mann-Whitney tests of differences in group average gaps between average last announced contribution and average binding contribution; and differences in group average gaps between average promised contributions and average binding contributions

	NCTwR	NCTwP	NCTwP&R
NCT	0.007 NCT>NCTwR	1	0.001 NCT>NCTwP&R
NCTwR	.	0.013 NCTwP>NCTwR	0.300
NCTwP	.	.	0.001 0.001 NCTwP>NCTwP&R

Note: The number in each cell displays the p -value of the test. The first number in the lower right cell refers to the test with average announced contribution, while the second number refers to the test with average promised contribution. The number of observations for each treatment equals the number of groups (see Table 1). The inequalities mean “average gap in first treatment exceeds average gap in second treatment.”

14. Tests of significance of differences in group average proportion of lies on announcements and lies on promises.

We conducted Mann-Whitney tests to compare the average proportion of non-fulfillment of announcements, and where relevant promises, at group level over the whole 10 periods for the relevant pairs of treatments. The tests, shown in Table A.18, show that a significantly higher proportion of announcements were not fulfilled in each treatment without punishment opportunity than in each treatment with such opportunity—i.e., there were more lies on announcement in **NCT** than in **NCTwR**, in **NCT** than in **NCTwP&R**, in **NCTwR** than in **NCTwP**, and in **NCTwP** than in **NCTwP&R**. The test for lying on

promises also shows that there were a significantly higher fraction of lies on promises in the NCTwP treatment than in the NCTwP&R treatment.

Table A.18. *p*-values of two-tailed Mann-Whitney tests of differences in group average proportion of lies on announcements of contribution and group average proportion of lies on promises.

	NCTwR	NCTwP	NCTwP&R
NCT	0.010 NCT>NCTwR	0.898	0.001 NCT>NCTwP&R
NCTwR	.	0.013 NCTwP>NCTwR	0.332
NCTwP	.	.	0.001 0.001 NCTwP>NCTwP&R

Note: The number in each cell is the *p*-value of the test. The first number in the lower right cell refers to the test with average proportion of lies on announced contributions, while the second number refers to the test with average proportion of lies on promised contribution. The number of observations for each treatment equals the number of groups in the treatment (see Table 1 of the paper). The inequalities mean “average proportion of lies in treatment _ exceeds average proportion of lies in treatment _”.

15. Tests of difference in group average earnings loss due to punishment, and Figure A.4.

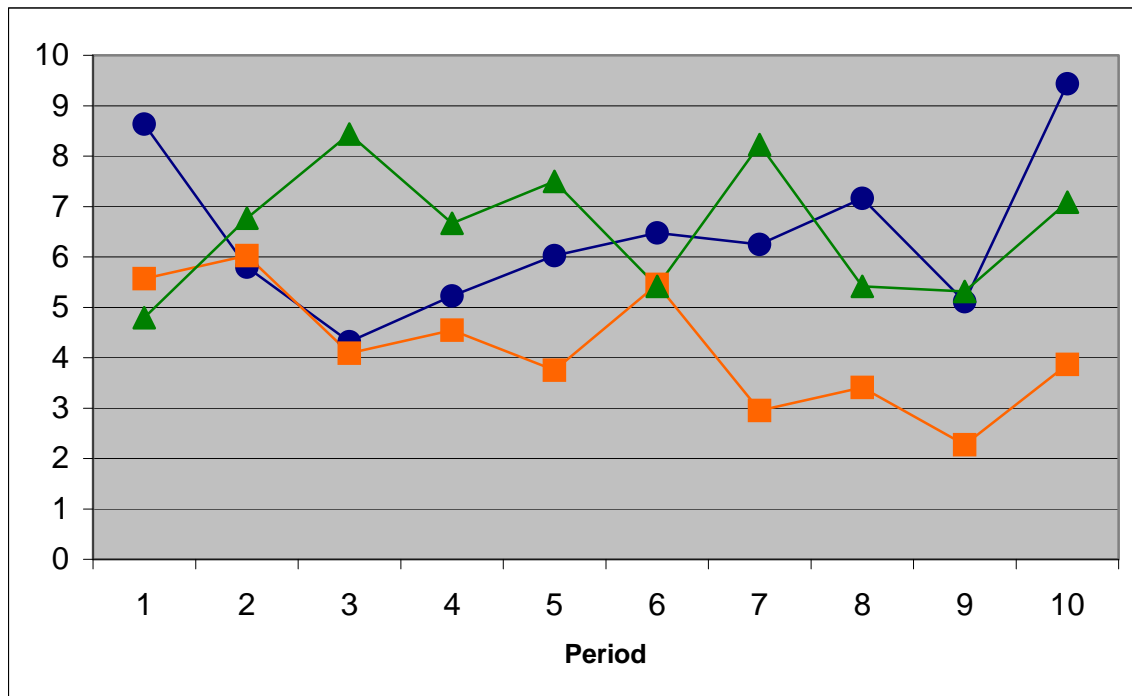
We conducted Mann-Whitney tests for differences in the levels of average earnings loss due to punishment across pairs of treatments in which punishment is possible. Results are shown in Table A.19. The difference is significant for one pair of treatments only: there is more loss of earnings in the NCTwR than in the NCTwP&R treatment, significant at the 10% level.

Table A.19. *p*-values of two-tailed Mann-Whitney tests of differences in group average earnings loss due to punishment

	NCTwR	NCTwP&R
R	0.608	0.288
NCTwR	.	0.088 NCTwR>NCTwP&R

Note: The number in each cell is the *p*-value of the test. Number of observations per treatment is same as number of groups (see Table 1 of the paper). The inequality means “average earnings loss in first treatment exceeds average earnings loss in second treatment.”

Figure A.4 Average earnings loss from punishment, NCTwR and NCTwP&R



Note: Triangles = Average loss from punishment, **R** treatment, Circles = Average loss from punishment, **NCTwR** treatment, Squares = Average loss from punishment, **NCTwP&R** treatment. Average loss is at group level; thus, loss per subject is one quarter of the amount shown.