Supplementary Information:

Hierarchy, Power, and Strategies to Promote Cooperation in Social Dilemmas

Catherine Molho

Daniel Balliet

Junhui Wu

Contents

Study 1

Participant Recruitment

In Study 1, we used two methods to ensure that participants logged in the study at approximately the same time. First, participants completed a sign-up survey in which they could choose the date and time of their participation. We sent three notification emails: (1) one day before the study, (2) one hour before the study, and (3) when we posted the study online. Second, we ran sessions without requiring participants to sign-up in advance. The study was posted periodically with limited participant slots and a time frame of approximately five minutes for participants to log in. One hundred and sixty-four participants completed a sign-up survey prior to participating in the study; among them, 54 eventually took part in the study. The remaining sample for Study 1 (N = 192) did not sign up prior to participating in the study.

Cooperation and Punishment in the Public Goods Game across Time

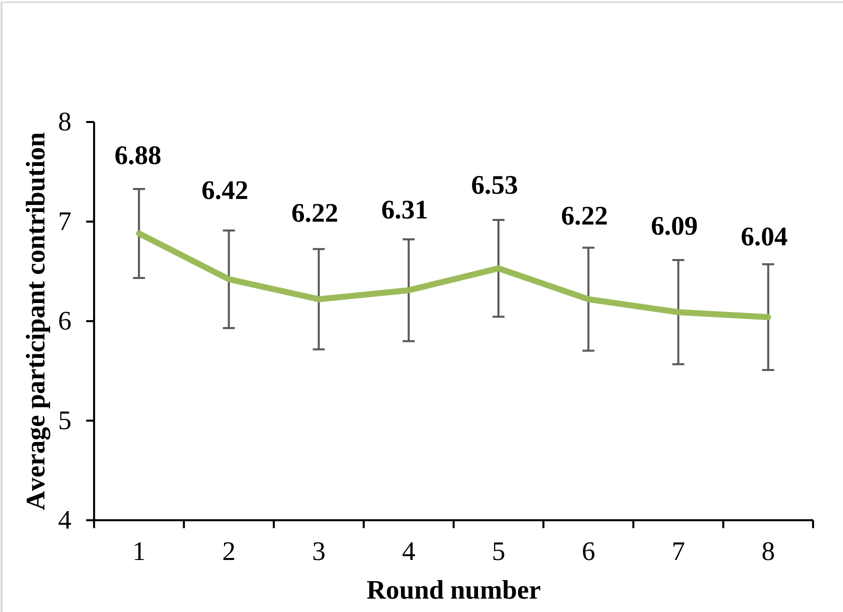

Figure 1. Average contributions (0-10 MUs) across PGG rounds.

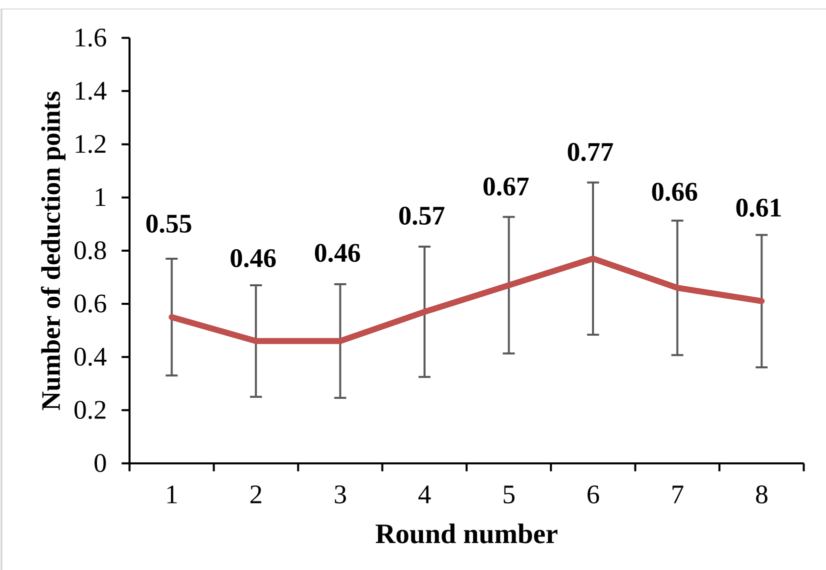

Figure 2. Average punishment (0-5 deduction points assigned) across PGG rounds.

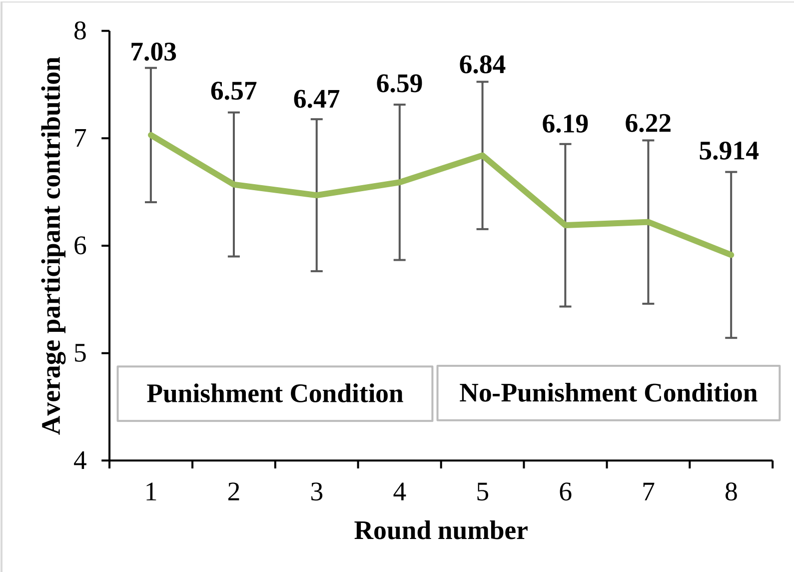

Figure 3. Average contributions (0-10 MUs) across PGG rounds among participants who first had a punishment option and then did not (Punishment/No-Punishment Treatment).

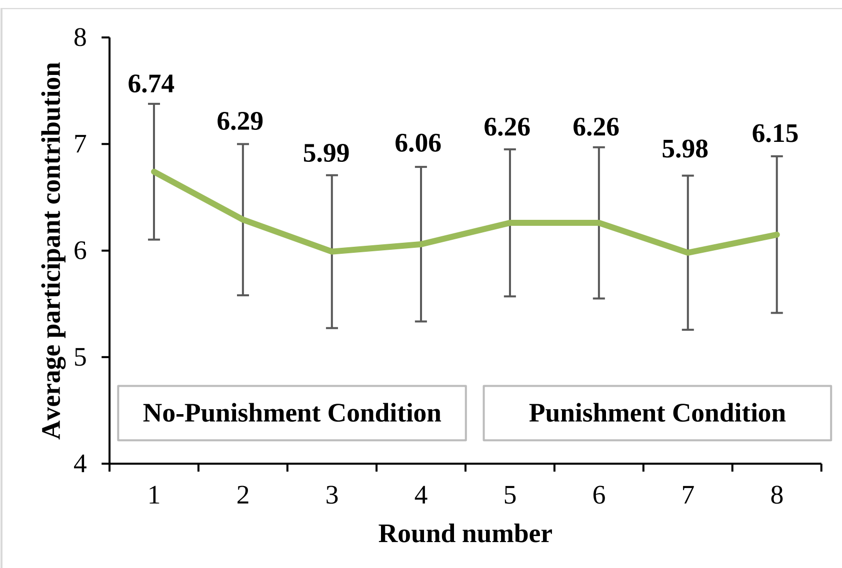

Figure 4. Average contributions (0-10 MUs) across PGG rounds among participants who first did *not* have a punishment option and then did (No-Punishment/ Punishment Treatment).

Cooperation and Punishment in the Public Goods Game across Power Conditions

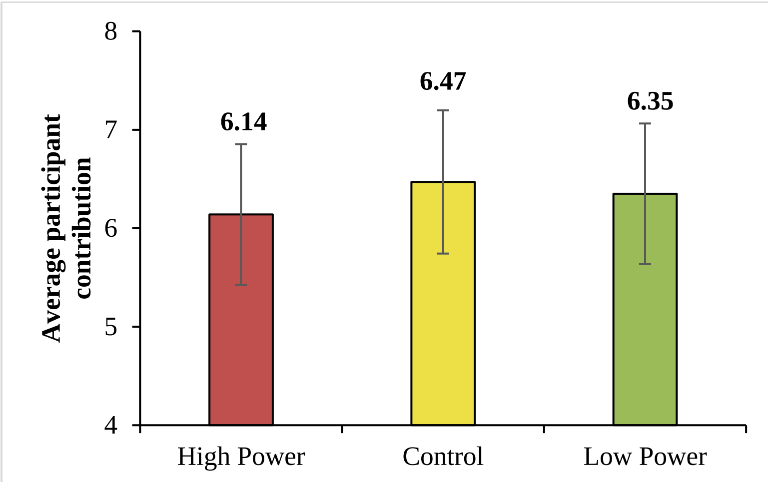

Figure 5. Average contributions (0-10 MUs) across power conditions in the PGG.

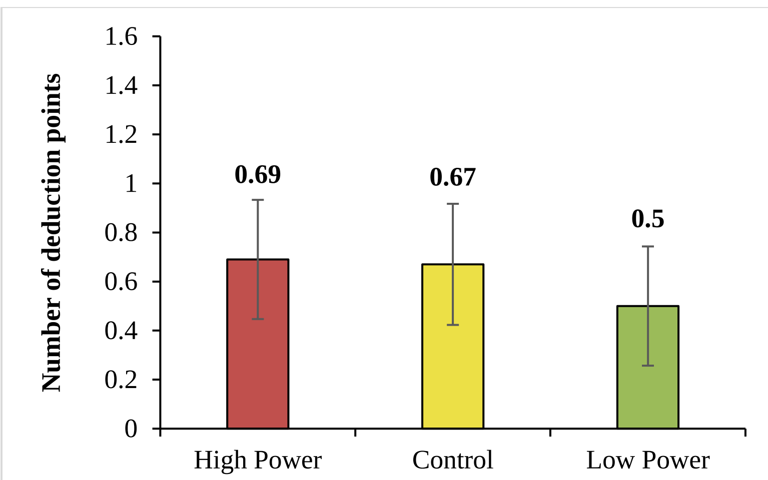

Figure 6. Average punishment (0-5 deduction points assigned) across power conditions in the PGG.

Models Predicting Cooperation in the Public Goods Game

Main Analyses

We coded cooperation as 0 (0-4 MUs), 1 (5 MUs), and 2 (6-10 MUs), and ran an ordinal logistic regression testing the effects of time, order of punishment conditions (0 = NP/P; 1 = P/NP), power, punishment option, and the Power × Punishment option interaction on cooperation (N = 246; k = 1,949). The significance tests and parameter estimates are presented in Tables S1a and S1b.

Table S1a. Significance tests from the ordinal logistic regression on cooperation

Variable	Wald χ^2	df	*p*
Time	5.11	1	.024
Order	0.02	1	.874
Power	0.60	2	.741
Punishment option	0.48	1	.487
Power × Punishment option	3.29	2	.193

Note. Cooperation was coded as 0 (0-4 MUs), 1 (5 MUs), and 2 (6-10 MUs).

Table S1b. Parameter estimates from the ordinal logistic regression on cooperation

Variable	*b*	95% CI
Time	-0.04	[-0.08, -0.01]
Order	0.03	[-0.36, 0.42]
Low power	0.39	[-0.17, 0.96]
Control	0.31	[-0.23, 0.85]
Punishment option	0.18	[-0.13, 0.49]
Low Power × Punishment option	-0.44	[-0.95, 0.08]
Control × Punishment option	-0.31	[-0.74, 0.12]

Notes. Cooperation was coded as 0 (0-4 MUs), 1 (5 MUs), and 2 (6-10 MUs). High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables. Punishment option (0 = no punishment, 1 = punishment) was dummy coded.

Analyses with Continuous Measure of Cooperation

We ran a linear regression on the continuous measure of cooperation (i.e., 0 to 10 MUs) testing the effects of time, order of punishment conditions (0 = NP/P; 1 = P/NP), power, punishment option, and the Power × Punishment option interaction (*N* = 246; *k* = 1,949 observations). The significance tests and parameter estimates are presented in Tables S2a and S2b.

Table S2a. Significance tests from the linear regression on cooperation

Variable	Wald χ^2	*df*	*p*
Time	8.96	1	.003
Order	0.03	1	.852
Power	0.32	2	.853
Punishment option	0.01	1	.903
Power × Punishment option	3.16	2	.206

Note. Cooperation was used in its original form: 0-10 MUs contributed to collective account.

Table S2b. Parameter estimates from the linear regression on cooperation

Variable	*b*	95% CI
Time	-0.11	[-0.18, -0.04]
Order	-0.08	[-0.88, 0.73]
Low power	0.43	[-0.67, 1.54]
Control	0.55	[-0.50, 1.60]
Punishment option	0.41	[-0.04, 0.86]
Low Power × Punishment option	-0.61	[-1.46, 0.23]
Control × Punishment option	-0.55	[-1.28, 0.18]

Notes. Cooperation was used in its original form: 0-10 MUs contributed to collective account. High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables. Punishment option (0 = no punishment, 1 = punishment) was dummy coded.

Analyses with Binary Measure of Cooperation

We coded cooperation as 0 (0 MUs) and 1 (1-10 MUs), and ran a binary logistic regression testing the effects of time, order of punishment conditions (0 = NP/P; 1 = P/NP), power, punishment option, and the Power × Punishment option interaction on cooperation (*N* = 246; *k* = 1,949 observations). The significance tests and parameter estimates are presented in Tables S3a and S3b.

Table S3a. Significance tests from the binary logistic regression on cooperation

Variable	Wald χ^2	*df*	*p*
Time	27.02	1	< .001
Order	0.46	1	.498
Power	0.81	2	.666
Punishment option	1.38	1	.240
Power × Punishment option	1.22	2	.542

Note. Cooperation was coded as 0 (0 MUs) and 1 (1-10 MUs).

Table S3b. Parameter estimates from the binary logistic regression on cooperation

Variable	*b*	95% CI
Time	-0.11	[-0.16, -0.07]
Order	-0.18	[-0.71, 0.34]
Low power	0.01	[-0.67, 0.68]
Control	0.21	[-0.49, 0.92]
Punishment option	0.34	[-0.07, 0.75]
Low Power × Punishment option	-0.34	[-0.95, 0.27]
Control × Punishment option	-0.19	[-0.83, 0.45]

Notes. Cooperation was coded as 0 (0 MUs) and 1 (1-10 MUs). High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables. Punishment option (0 = no punishment, 1 = punishment) was dummy coded.

Models Predicting Punishment Behavior in the Public Goods Game

Main Analyses

We coded punishment behavior as 0 (0 deduction MUs) and 1 (1-5 deduction MUs), and conducted a binary logistic regression testing the effects of time, order of punishment conditions (0 = NP/P; 1 = P/NP), and power on punishment behavior (N = 246; k = 980 observations). The model also controlled for the other group member's cooperation level, as well as the positive and negative deviations of the other's contribution from participant's contribution. The significance tests and parameter estimates are presented in Tables S4a and S4b.

Table S4a. Significance tests from the binary logistic regression on punishment behavior

Variable	Wald χ^2	*df*	*p*
Time	0.29	1	.588
Order	0.64	1	.423
Power	2.98	2	.225
Group member's contribution	7.89	1	.005
Positive deviation	18.38	1	< .001
Negative deviation	6.15	1	.013

Note. Punishment behavior was coded as 0 (0 deduction MUs) or 1 (1-5 deduction MUs).

Table S4b. Parameter estimates from the binary logistic regression on punishment behavior

Variable	*b*	95% CI
Time	-0.03	[-0.14, 0.08]
Order	0.28	[-0.40, 0.95]
Low power	-0.56	[-1.22, 0.10]
Control	-0.11	[-0.72, 0.51]
Group member's contribution	-0.10	[-0.17, -0.03]
Positive deviation	0.16	[0.08, 0.23]
Negative deviation	0.09	[0.02, 0.17]

Note. Punishment behavior was coded as 0 (0 deduction MUs) and 1 (1-5 deduction MUs). High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables.

Analyses with Continuous Measure of Punishment Behavior

We ran a linear regression on the continuous measure of punishment behavior (i.e., 0 to 5 deduction MUs) testing the effects of time, order of punishment conditions (0 = NP/P; 1 = P/NP), and power (N = 246; k = 980 observations). The model also controlled for the other group member's cooperation level, as well as the positive and negative deviations of the other's contribution from participant's contribution. The significance tests and parameter estimates are presented in Tables S5a and S5b.

Table S5a. Significance tests from the linear regression on punishment behavior

Variable	Wald χ^2	*df*	*p*
Time	0.54	1	.462
Order	0.11	1	.738
Power	1.99	2	.369
Group member's contribution	6.70	1	.008
Positive deviation	13.98	1	< .001
Negative deviation	9.19	1	.002

Note. Punishment behavior was used in its original form, ranging from 0-5 deduction MUs.

Table S5b. Parameter estimates from the linear regression on punishment behavior

Variable	*b*	95% CI
Time	0.02	[-0.04, 0.09]
Order	0.06	[-0.30, 0.43]
Low power	-0.16	[-0.47, 0.13]
Control	0.04	[-0.28, 0.36]
Group member's contribution	-0.05	[-0.09, -0.01]
Positive deviation	0.09	[0.04, 0.13]
Negative deviation	0.10	[0.03, 0.16]

Note. Punishment behavior was used in its original form, ranging from 0 to 5 deduction MUs. High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables.

Behavior and Earnings across Experimental Games

To better understand how high-power participants' allocation decisions in the DG were related to their own and their group member's behavior in the PGG, we ran some additional analyses focusing on high-power participants in our sample ($N = 83$). We first looked at the bivariate correlations of high-power participants' and their low-power counterparts' average contributions to the collective account in the PGG with high-power participants' allocated MUs to the other in the DG. We observed a positive correlation between high-power individuals' PGG contributions and their DG allocations ($r = .45$, $p < .001$). We also observed a positive correlation between the low-power group member's PGG contributions and the high-power individuals' DG allocations ($r = .38$, $p = .001$). However, when regressing the DG allocations on high-power individuals' and their low-power partner's contributions, we only found a statistically significant effect of the high-power individuals' average contribution, $b = 3.06$, $t(77) = 2.38$, $p = .020$, whereas low-power group member's average contribution was not a significant predictor, $b = 0.67$, $t(77) = 0.51$, $p = .615$.

Finally, we analyzed differences in participants' total earnings depending on their power condition using an ANOVA. We observed a significant positive effect of power on participants' total earnings, $F(1, 239) = 28.13$, $p < .001$, $\eta_p^2 = .19$. More specifically, participants in the high-power condition earned more MUs compared to those in the control condition ($b = -24.04$, $p < .001$), and those in the low-power condition ($b = -44.81$, $p < .001$).

Study 2

Participant Recruitment

In Study 2, all participants completed a pre-study survey where they signed up for specific timeslots. In total, 920 individuals completed the sign-up. Besides scheduling purposes, the sign-up served to exclude those MTurk workers who already participated in Study 1.

Cooperation, Punishment, and Gossip in the Public Goods Game across Time

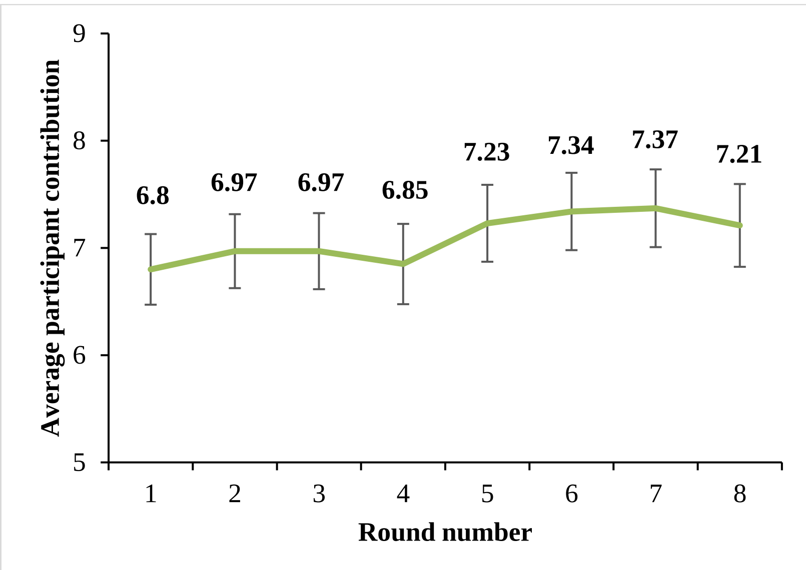

Figure 7. Average contributions (0-10 MUs) across PGG rounds.

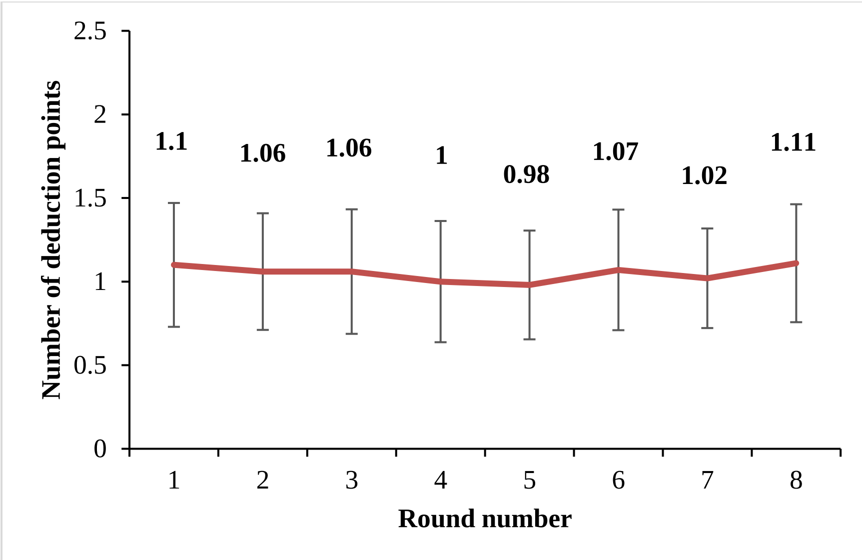

Figure 8. Average punishment (0-5 deduction points assigned) across PGG rounds.

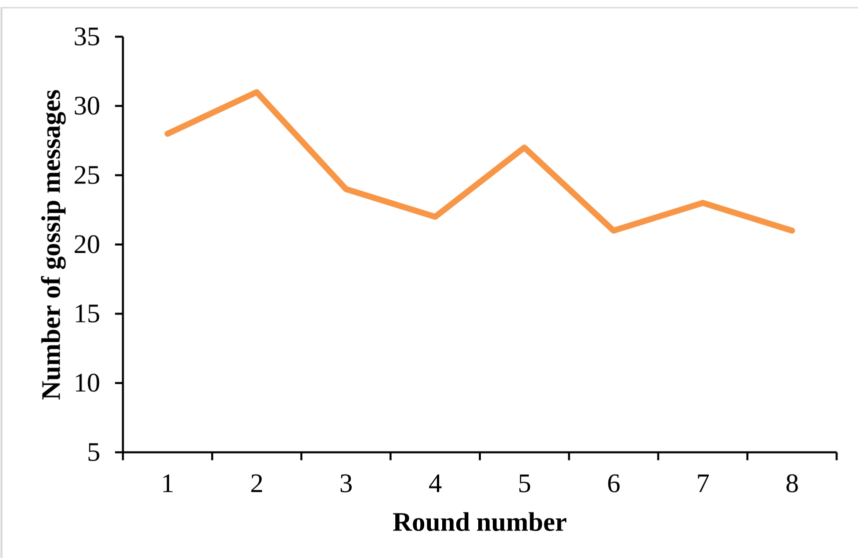

Figure 9. Total number of gossip messages sent across PGG rounds.

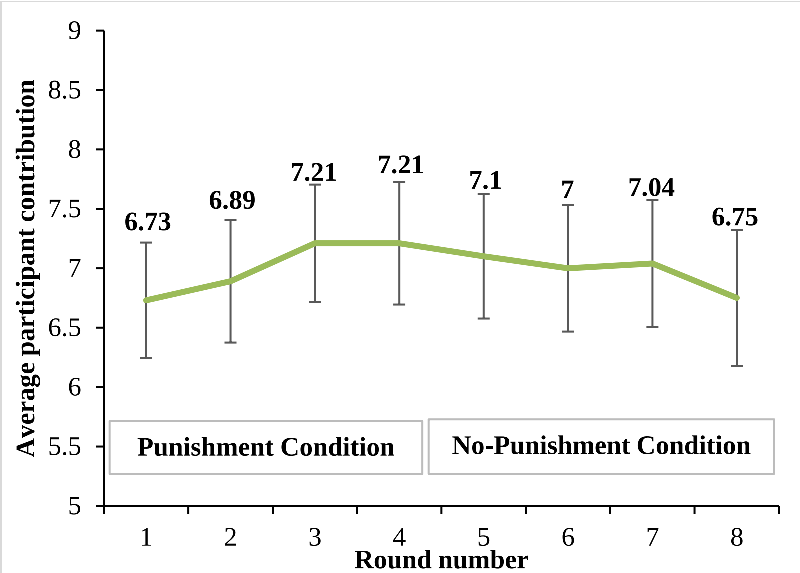

Figure 10. Average contributions (0-10 MUs) across PGG rounds among participants who first had a punishment option and then did not (Punishment/No-Punishment Treatment).

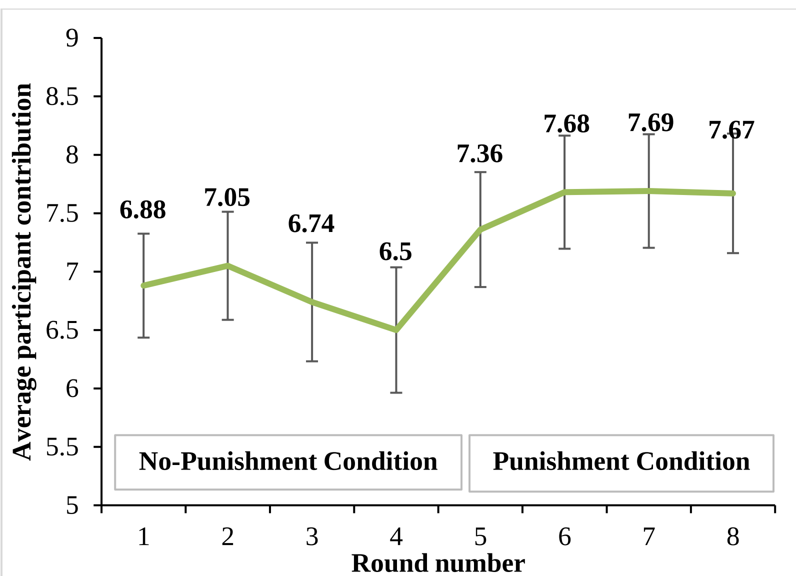

Figure 11. Average contributions (0-10 MUs) across PGG rounds among participants who first did *not* have a punishment option and then did (No-Punishment/Punishment Treatment).

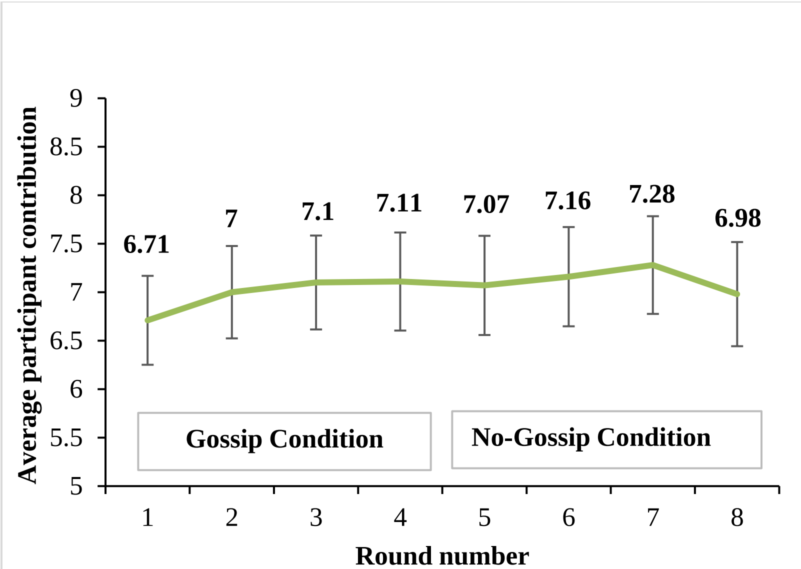

Figure 12. Average contributions (0-10 MUs) across PGG rounds among participants who first had a gossip option and then did not (Gossip/No-Gossip Treatment).

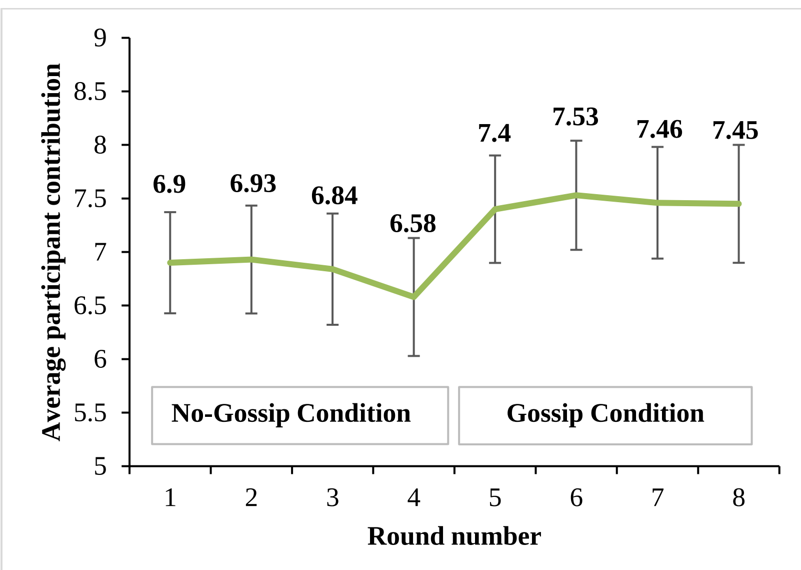

Figure 13. Average contributions (0-10 MUs) across PGG rounds among participants who first did *not* have a gossip option and then did (No-Gossip/Gossip Treatment).

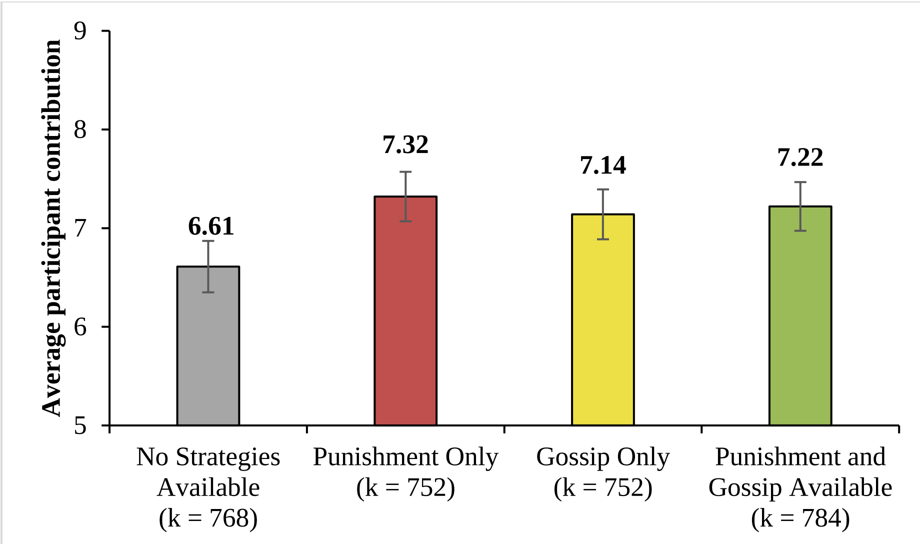

Figure 14. Contributions (0-10 MUs) in the PGG depending on the availability of punishment and gossip options. **Note:** means are calculated based on all repeated observations of cooperation.

Cooperation, Punishment, and Gossip in the Public Goods Game across Power Conditions

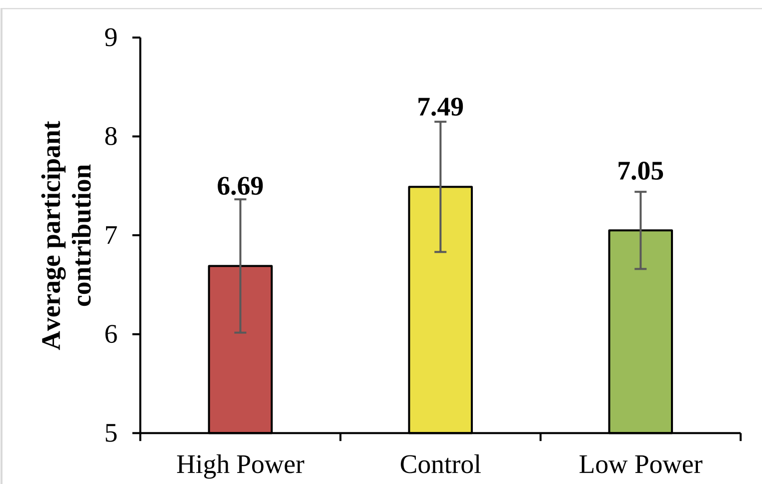

Figure 15. Average contributions (0-10 MUs) across power conditions in the PGG.

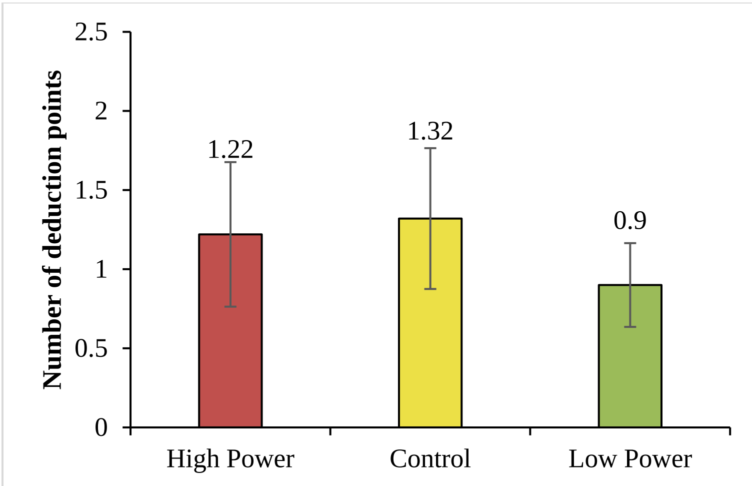

Figure 16. Average punishment (0-5 deduction points assigned) across power conditions in the PGG.

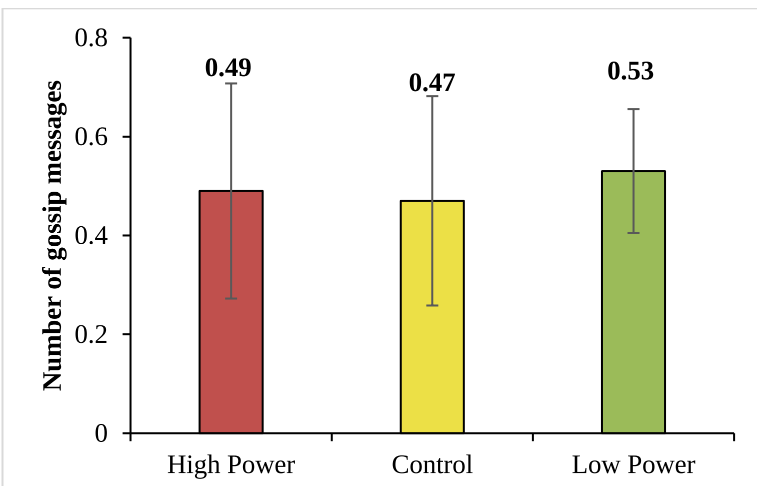

Figure 17. Average gossip messages (0-4) sent across power conditions in the PGG.

Models Predicting Cooperation in the Public Goods Game

Main Analyses

We coded cooperation as 0 (0-4 MUs), 1 (5 MUs), and 2 (6-10 MUs), and ran an ordinal logistic regression testing the effects of time, order of punishment conditions (0 = NP/P; 1 = P/NP), order of gossip conditions (0 = NG/G; 1 = G/NG), power, punishment option, and the Power × Punishment option and Power × Gossip option interactions on cooperation (*N* = 371; *k* = 2,952 observations). The significance tests and parameter estimates are presented in Tables S7a and S7b.

Table S6a. Significance tests from the ordinal logistic regression on cooperation

Variable	Wald χ^2	*df*	*p*
Time	7.17	1	.007
Punishment order	2.76	1	.097
Gossip order	3.65	1	.056
Power	5.36	2	.069
Punishment option	7.47	1	.006
Gossip option	3.84	1	.050
Power × Punishment option	1.44	2	.485
Power × Gossip option	0.57	2	.751

Note. Cooperation was coded as 0 (0-4 MUs), 1 (5 MUs), and 2 (6-10 MUs).

Table S6b. Parameter estimates from the ordinal logistic regression on cooperation

Variable	*b*	95% CI
Time	0.05	[0.01, 0.08]
Punishment order	0.27	[-0.05, 0.60]
Gossip order	0.32	[-0.01, 0.64]
Low power	0.36	[-0.17, 0.88]
Control	0.35	[-0.26, 0.97]
Punishment option	0.12	[-0.27, 0.51]
Gossip option	0.11	[-0.28, 0.50]
Low Power × Punishment option	0.12	[-0.33, 0.57]
Control × Punishment option	0.31	[-0.21, 0.83]
Low Power × Gossip option	0.16	[-0.28, 0.61]
Control × Gossip option	0.07	[-0.45, 0.59]

Note. Cooperation was coded as 0 (0-4 MUs), 1 (5 MUs), and 2 (6-10 MUs). High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables. Punishment option (0 = no punishment, 1 = punishment) and gossip option (0 = no gossip, 1 = gossip) were dummy coded.

Analyses with Continuous Measure of Cooperation

We ran a linear regression on the continuous measure of cooperation (i.e., 0 to 10 MUs) testing the effects of time, order of punishment conditions (0 = NP/P; 1 = P/NP), order of gossip conditions (0 = NG/G; 1 = G/NG), power, punishment option, and the Power × Punishment option and Power × Gossip option interactions (N = 371; k = 2,952 observations). The significance tests and parameter estimates are presented in Tables S8a and S8b.

Table S7a. Significance tests from the linear regression on cooperation

Variable	Wald χ^2	*df*	*p*
Time	6.40	1	.011
Punishment order	1.34	1	.247
Gossip order	0.56	1	.456
Power	2.53	2	.282
Punishment option	19.72	1	< .001
Gossip option	8.42	1	.004
Power × Punishment option	5.43	2	.066
Power × Gossip option	0.68	2	.710

Note. Cooperation was used in its original form: 0-10 MUs contributed to collective account.

Table S7b. Parameter estimates from the linear regression on cooperation

Variable	*b*	95% CI
Time	0.06	[0.01, 0.11]
Punishment order	0.35	[-0.24, 0.95]
Gossip order	0.23	[-0.37, 0.82]
Low power	0.62	[-0.32, 1.56]
Control	0.53	[-0.55, 1.62]
Punishment option	0.53	[0.01, 1.05]
Gossip option	0.38	[-0.14, 0.91]
Low Power × Punishment option	-0.25	[-0.83, 0.33]
Control × Punishment option	0.37	[-0.33, 1.07]
Low Power × Gossip option	-0.13	[-0.71, 0.45]
Control × Gossip option	0.08	[-0.61, 0.78]

Note. Cooperation was used in its original form: 0-10 MUs contributed to collective account. High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables. Punishment option (0 = no punishment, 1 = punishment) and gossip option (0 = no gossip, 1 = gossip) were dummy coded.

Analyses with Binary Measure of Cooperation

We coded cooperation as 0 (0 MUs) and 1 (1-10 MUs), and ran a binary logistic regression testing the effects of from time, order of punishment conditions (0 = NP/P; 1 = P/NP), order of gossip conditions (0 = NG/G; 1 = G/NG), power, punishment option, and the Power × Punishment option and Power × Gossip option interactions on cooperation (N = 371; k = 2,952 observations). The significance tests and parameter estimates are presented in Tables S9a and S9b.

Table S8a. Significance tests from the binary logistic regression on cooperation

Variable	Wald χ^2	*df*	*p*
Time	12.23	1	< .001
Punishment order	0.82	1	.366
Gossip order	1.09	1	.295
Power	4.07	2	.131
Punishment option	6.80	1	.009
Gossip option	3.58	1	.058
Power × Punishment option	3.45	2	.178
Power × Gossip option	1.09	2	.579

Note. Cooperation was coded as 0 (0 MUs) and 1 (1-10 MUs).

Table S8b. Parameter estimates from the binary logistic regression on cooperation

Variable	*b*	95% CI
Time	-0.09	[-0.14, -0.04]
Punishment order	0.23	[-0.27, 0.73]
Gossip order	-0.27	[-0.78, 0.24]
Low power	0.42	[-0.26, 1.10]
Control	0.47	[-0.35, 1.28]
Punishment option	0.54	[0.001, 1.08]
Gossip option	0.18	[-0.35, 0.70]
Low Power × Punishment option	-0.44	[-1.06, 0.18]
Control × Punishment option	0.24	[-0.77, 1.24]
Low Power × Gossip option	0.02	[-0.59, 0.63]
Control × Gossip option	0.48	[-0.50, 1.47]

Note. Cooperation was coded as 0 (0 MUs) and 1 (1-10 MUs). High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables. Punishment option (0 = no punishment, 1 = punishment) and gossip option (0 = no gossip, 1 = gossip) were dummy coded.

Models Predicting Punishment Behavior in the Public Goods Game

Main Analyses

We coded punishment behavior as 0 (0 deduction MUs) and 1 (1-15 deduction MUs), and conducted a binary logistic regression testing the effects of time, order of punishment conditions (0 = NP/P; 1 = P/NP), and power on punishment behavior (N = 371; k = 1,484 observations). The model also controlled for the average cooperation level of group members, as well as the positive and negative deviations of group members' contributions from participant's contribution. The significance tests and parameter estimates are presented in Tables S10a and S10b.

Table S9a. Significance tests from the binary logistic regression on punishment behavior

Variable	Wald χ^2	df	p
Time	0.01	1	.945
Punishment order	0.04	1	.844
Power	2.66	2	.264
Group members' contributions	21.04	1	< .001
Positive deviation	5.94	1	.015
Negative deviation	29.81	1	< .001

Note. Punishment behavior was coded as 0 (0 deduction MUs) and 1 (1-15 deduction MUs).

Table S9b. Parameter estimates from the binary logistic regression on punishment behavior

Variable	*b*	95% CI
Time	0.003	[-0.08, 0.09]
Punishment order	0.05	[-0.46, 0.57]
Low power	-0.06	[-0.56, 0.43]
Control	0.32	[-0.25, 0.88]
Group members' contributions	-0.15	[-0.22, -0.09]
Positive deviation	0.08	[0.01, 0.14]
Negative deviation	0.25	[0.16, 0.34]

Note. Punishment behavior was coded as 0 (0 deduction MUs) and 1 (1-15 deduction MUs). High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables.

Analyses with Continuous Measure of Punishment Behavior

We ran a linear regression on the continuous measure of punishment behavior (i.e., 0 to 15 deduction MUs) testing the effects of time, order of punishment conditions (0 = NP/P; 1 = P/NP), and power (N = 371; k = 1,484 observations). The model also controlled for the average cooperation level of group members, as well as the positive and negative deviations of group members' contributions from participant's contribution. The significance tests and parameter estimates are presented in Tables S11a and S11b.

Table S10a. Significance tests from the linear regression on punishment behavior

Variable	Wald χ^2	df	p
Time	2.24	1	.135
Punishment order	0.46	1	.498
Power	1.81	2	.405
Group members' contributions	7.39	1	.007
Positive deviation	4.34	1	.037
Negative deviation	34.68	1	< .001

Note. Punishment behavior was used in its original form: 0-15 deduction MUs assigned to others.

Table S10b. Parameter estimates from the linear regression on punishment behavior

Variable	*b*	95% CI
Time	0.07	[-0.02, 0.16]
Punishment order	-0.18	[-0.70, 0.34]
Low power	-0.21	[-0.77, 0.34]
Control	0.11	[-0.57, 0.79]
Group members' contributions	-0.07	[-0.12, -0.02]
Positive deviation	0.10	[0.01, 0.19]
Negative deviation	0.35	[0.23, 0.46]

Note. Punishment behavior was used in its original form: 0-15 deduction MUs assigned to others. High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables.

Model Predicting Gossip Behavior in the Public Goods Game

We coded gossip behavior as 0 (no gossip sent) and 1 (gossip sent), and ran a binary logistic regression on gossip behavior testing the effects of time, order of gossip conditions (0 = NG/G; 1 = G/NG), and power (N = 371; k = 1,484 observations). The significance tests and parameter estimates are presented in Tables S12a and S12b.

Table S11a. Significance tests from the binary logistic regression on gossip behavior

Variable	Wald χ^2	*df*	*p*
Time	2.44	1	.119
Gossip order	0.75	1	.388
Power	0.22	2	.897

Table S11b. Parameter estimates from the binary logistic regression on gossip behavior

Variable	*b*	95% CI
Time	-0.09	[-0.21, 0.02]
Gossip order	0.28	[-0.36, 0.93]
Low power	0.07	[-0.51, 0.65]
Control	-0.05	[-0.74, 0.63]

Note. High power was the reference group when coding low power (1 = low power, 0 = high power) and control (1 = control, 0 = high power) as two dummy variables.

Behavior and Earnings Across Experimental Games

To better understand how high-power participants' allocation decisions in the DG were related to their own and their group members' behavior in the PGG, we ran additional analyses focusing on high-power participants in our sample (*N* = 75). We first looked at the correlations of high-power participants' and their low-power counterparts' average contributions to the collective account in the PGG with the MUs allocated to group members in the DG. As in Study 1, we observed a positive correlation between high-power individuals' PGG contributions and their DG allocations to group members ($r = .63$, $p < .001$). We also observed a positive correlation between the low-power members' PGG contributions and the high-power individuals' DG allocations ($r = .30$, $p = .008$). However, when regressing the DG allocations on high-power individuals and their low-power partners' PGG contributions, we only found a significant effect of high-power individuals' average contributions, $b = 11.26$, $t(71) = 5.99$, $p < .001$, whereas their low-power individuals' average contribution was not a significant predictor, $\beta = 0.96$, $t(71) = 0.39$, $p = .694$.

Finally, we analyzed differences in participants' total earnings depending on their power condition using an ANOVA. As in Study 1, we observed a significant positive effect of power on participants' total earnings, $F(1, 331) = 46.46$, $p < .001$, $\eta_p^2 = .22$. More specifically, participants in the high-power condition earned more MUs compared to those in the control condition ($b = -63.85$, $p < .001$), and those in the low-power condition ($b = -83.26$, $p < .001$).