Determinants of Norm Compliance: Moral Similarity and Group Identification

Appendix

Alexander Schneeberger Erin L. Krupka

July 19, 2026

Appendix A Deviations from the Pre-Analysis Plan

We preregistered our analysis at the American Economic Association's registry for randomized controlled trials (https://doi.org/g5qw, https://doi.org/mcw3, and https://doi.org/m675). We deviated from the Pre-Analysis Plans in the following ways:

1. We renamed the behavioral experiment the choice experiment.
2. We renamed the communal index the binding index. This change also applied to words related to the index (e.g., communal treatment was changed to binding treatment).
3. For reasons of clarity, we restated some of our hypotheses without changing their contents.[1]
4. To improve readability, we moved the analysis of the original Hypothesis 2 ("According to the measured group norms, the social appropriateness is linearly increasing in the number of balls placed in the norm compliant bucket") to the appendix (see Appendix B.2).
5. While examining the original Hypothesis 2, we used a Tobit regression model instead of the preregistered OLS regression model. We did this to take into account the censored nature of the social appropriateness ratings. However, we can report that the coefficient on the number of balls in the rule-compliant bucket was positive and significant in the OLS regression model ($b = 0.033$, $p < 0.001$).
6. While examining Hypothesis 2, we only reported the result of the preregistered one-sided t-tests. We did not report the results of the one-sided Wilcoxon-Mann-Whitney tests. This is because a one-sided Wilcoxon-Mann-Whitney test requires that the shapes of the distributions in the two samples are identical. This assumption might have been unrealistic in our case. However, the one-sided Wilcoxon-Mann-Whitney tests would have been significant (in all versions: $p < 0.001$).
7. We renamed Hypothesis 4a and 4b into Hypothesis 3 and 4.
8. While examining Hypothesis 3, we only reported the results of a preregistered comparison of the number of balls in the rule-compliant bucket between treatments. We did not report the results of a conditional (fixed-effect) logit model (McFadden, 1974). We deviated from the Pre-Analysis plan because the correlation between two explanatory variables made a reliable estimation of coefficients impossible. Since

[1] For reference, the preregistered hypothesis were: "**Hypothesis 1 (Group Norms):** According to individualizing subjects, the group norms present in the modified rule following task are identical in the individualizing and the communal treatment", "**Hypothesis 2 (Group Norms):** According to the measured group norms, the social appropriateness is linearly increasing in the number of balls placed in the norm compliant bucket", "**Hypothesis 3 (Group Identification):** Group identification is significantly higher for individualizing individuals that are assigned to the individualizing treatment compared to those that are assigned to the communal treatment." and "**Hypothesis 4 (Norm Compliance) / Hypothesis 4a:** Compliance with the group norms is higher in the individualizing treatment compared to the communal treatment. / **Hypothesis 4b:** This treatment effect is mediated by changes in group identification."

both group norms were identical, an action's monetary earnings and average social appropriateness were highly correlated.

9. While examining Hypothesis 3, we only reported the result of one-sided t-tests. For the same reason as described in the paragraph on Hypothesis 2, we did not report the results of one-sided Wilcoxon-Mann-Whitney tests. However, both tests provided the same results: A one-sided Wilcoxon-Mann-Whitney test would have been insignificant in the all control questions version ($p = 0.168$), significant in the limited control questions version ($p = 0.009$), and insignificant in the no control questions version ($p = 0.189$).

10. While examining Hypothesis 4, we only reported the results of a Tobit regression model. For the same reason as described in the paragraph on Hypothesis 3, we did not report the results of a conditional (fixed-effect) logit model.

11. While examining Hypothesis 4, we did not report the results of the preregistered regression models. Instead, we reported the results of models that did not include the binding treatment dummy. However, both model specifications provided the same results: The coefficient of group identification would have also been positive and significant in the preregistered regression model in the all control questions version ($b = 14.821$, $p = 0.001$), the limited control questions version ($b = 7.190$, $p = 0.024$), and the no control questions version ($b = 8.994$, $p = 0.003$).

12. We did not report the results of all preregistered exploratory analyses that examine how robust our results are to variations in the norm compliance measure (e.g., more than 0 balls in the rule-compliant bucket). The reason for this omission was that group norms were linearly increasing in the number of balls in the rule-compliant bucket and did not abruptly change at some cutoff point (see Appendix B.2).

13. To improve understanding, we moved the effect of distance progressivism on group identification, which was only preregistered as an exploratory analysis, to the main results section (see Section 6.2.1).

14. While examining the effect of distance progressivism on group identification, we did not report the results of the preregistered regression models. Instead, we reported the results of models that did not include the binding treatment dummy. However, both model specifications provided the same results: The coefficient of distance progressivism would have also been negative and significant in the preregistered regression model in the all control questions version ($b = -0.082$, $p < 0.001$), the limited control questions version ($b = -0.073$, $p < 0.001$), and the no control questions version ($b = -0.061$, $p < 0.001$).

15. We did not report the results of all preregistered exploratory analyses that examined the robustness of the relationship between moral similarity and group identification. We only reported those results of the Pre-Analysis Plan that did yield interesting results. Furthermore, to be consistent with the analyses in the main results section, we removed the binding treatment dummy from our regression models in Appendix B Additional Analyses.

Appendix B Additional Analyses

B.1 Comparison of Moral and Socio-Demographic Data between Treatments

Table 1: Individual and Aggregated Moral Foundations Scores by Experiment, Version, and Treatment

	Choice experiment (all CQs.)			Choice experiment (limited CQs.)		
	Individualizing treatment	Binding treatment	P-value[a]	Individualizing treatment	Binding treatment	P-value[a]
Progressivism	1.72	1.59	0.247	1.36	1.39	0.744
Individualizing index	3.89	3.83	0.596	3.85	3.83	0.770
Binding index	2.17	2.25	0.459	2.50	2.44	0.473
Harm	3.84	3.79	0.451	3.88	3.85	0.539
Fairness	3.94	3.88	0.409	3.83	3.82	0.794
Ingroup	2.14	2.10	0.886	2.22	2.19	0.826
Authority	2.38	2.52	0.292	2.79	2.73	0.397
Purity	1.98	2.13	0.315	2.48	2.40	0.492
	Choice experiment (no CQs.)			Norm elicitation experiment		
	Individualizing treatment	Binding treatment	P-value[a]	Individualizing treatment	Binding treatment	P-value[a]
Progressivism	1.51	1.47	0.817	1.49	1.59	0.600
Individualizing index	3.81	3.81	0.741	3.82	3.85	0.716
Binding index	2.30	2.33	0.824	2.33	2.26	0.468
Harm	3.84	3.80	0.460	3.77	3.81	0.529
Fairness	3.79	3.81	0.838	3.88	3.88	0.957
Ingroup	2.07	2.09	0.775	2.15	2.18	0.820
Authority	2.55	2.64	0.366	2.62	2.54	0.634
Purity	2.30	2.27	0.887	2.22	2.06	0.246

Notes: Progressivism could take on values from -5 to 5. All other indices could take on values from 0 to 5. In the table 'CQ' refers to 'control question'.

[a] The p-values are based on a two-sided Wilcoxon-Mann-Whitney test that compared the scores between treatments.

This subsection compares the moral foundations and social-demographic data of participants of all versions of the choice experiment and of the norm elicitation experiment between treatments. We found no differences between treatments. Consequently, we attributed any treatment effect to our treatment variation.

To examine this, we report in Table 1, separated by experiment, version, and treatment, participants' average scores in the five individual moral foundations (harm, fairness, ingroup, authority, purity) and the three aggregated indices that we used (individualizing index, binding index, progressivism). Furthermore, in Tables 2, 3, 4, and 5, we report, separated by experiment, version, and treatment, the average answer

Table 2: Answers to the Socio-Demographic Questions by Treatment, Version, and Experiment (Part 1)

	Choice experiment (all CQs.)			Choice experiment (limited CQs.)		
	Individualizing treatment	Binding treatment	P-value[a]	Individualizing treatment	Binding treatment	P-value[a]
Age						
Mean	35.26	35.07	.	37.30	37.98	.
Total	159	160	0.947	233	233	0.526
Gender						
Female	87	87	.	145	130	.
Male	69	71	.	84	95	.
Other	3	2	.	4	8	.
Total	159	160	0.943	233	233	0.260
Ethnicity						
White	123	118	.	131	135	.
Asian	12	18	.	28	31	.
Hispanic	6	8	.	13	14	.
Black	14	10	.	50	45	.
Other	5	6	.	11	8	.
Total	160	160	0.681	233	233	0.915
Degree						
None	1	0	.	1	2	.
High school diploma	68	50	.	72	69	.
Bachelor degree	65	70	.	0	0	.
Master's degree	20	32	.	0	0	.
Doctorate degree	6	8	.	12	8	.
Total	160	160	0.111	233	233	0.688
Employment						
Full time employed	62	82	.	109	117	.
Part time employed	17	24	.	34	31	.
Unemployed	20	9	.	37	27	.
Self-employed	22	23	.	21	24	.
Student	19	11	.	10	16	.
Housewife / husband	8	5	.	18	10	.
Retired	11	6	.	3	8	.
Total	159	160	0.053	232	233	0.237

Notes: Answering the Socio-Demographic Questions was voluntary. In the table 'CQ' refers to 'control question'.

[a] The p-values are based on two-sided Wilcoxon-Mann-Whitney (age) or Fisher's exact tests (all other variables) that compared the answers between treatments.

Table 3: Answers to the Socio-Demographic Questions by Treatment, Version, and Experiment (Part 2)

	Choice experiment (all CQs.)			Choice experiment (limited CQs.)		
	Individualizing treatment	Binding treatment	P-value[a]	Individualizing treatment	Binding treatment	P-value[a]
Income						
$0 - $10,000	40	25	.	21	33	.
$10,000 - $25,000	26	28	.	17	19	.
$25,000 - $50,000	39	44	.	56	47	.
$50,000 - $75,000	24	31	.	67	48	.
$75,000 - $100,000	11	14	.	33	35	.
$100,000 - $150,000	14	13	.	26	32	.
More than $150,000	3	3	.	11	18	.
Total	157	158	0.520	231	232	0.171
Community						
Urban	44	55	.	87	78	.
Suburban	86	85	.	126	127	.
Rural	29	20	.	19	28	.
Total	159	160	0.243	232	233	0.336
Religion						
Christian	49	55	.	109	100	.
Atheist	57	50	.	59	50	.
Jew	1	3	.	3	7	.
Buddhist	3	2	.	2	9	.
Muslim	2	0	.	4	7	.
Other	47	49	.	52	58	.
Total	159	159	0.619	233	232	0.124
Politics						
Very liberal	44	38	.	41	31	.
Liberal	53	60	.	90	93	.
Moderate	45	52	.	67	68	.
Conservative	14	8	.	30	32	.
Very conservative	4	2	.	4	9	.
Total	160	160	0.462	232	233	0.500

Notes: Answering the Socio-Demographic Questions was voluntary. In the table 'CQ' refers to 'control question'.

[a] The p-values are based on two-sided Fisher's exact tests that compared the answers between treatments.

Table 4: Answers to the Socio-Demographic Questions by Treatment, Version, and Experiment (Part 3)

	Choice experiment (no CQs.)			Norm elicitation experiment		
	Individualizing treatment	Binding treatment	P-value[a]	Individualizing treatment	Binding treatment	P-value[a]
Age						
Mean	41.31	40.90	.	36.04	35.35	.
Total	233	233	0.936	99	100	0.323
Gender						
Female	117	111	.	53	54	.
Male	109	115	.	45	46	.
Other	7	7	.	1	0	.
Total	233	233	0.868	99	100	1.000
Ethnicity						
White	166	153	.	79	69	.
Asian	19	26	.	6	12	.
Hispanic	19	21	.	5	9	.
Black	23	28	.	5	6	.
Other	6	5	.	3	4	.
Total	233	233	0.679	98	100	0.405
Degree						
None	2	4	.	1	0	.
High school diploma	80	91	.	34	45	.
Bachelor degree	110	100	.	45	43	.
Master's degree	32	29	.	11	10	.
Doctorate degree	9	9	.	8	1	.
Total	233	233	0.737	99	99	0.067
Employment						
Full time employed	109	104	.	46	46	.
Part time employed	36	29	.	14	11	.
Unemployed	32	26	.	10	8	.
Self-employed	19	36	.	12	15	.
Student	8	12	.	9	16	.
Housewife / husband	13	12	.	3	2	.
Retired	16	12	.	5	2	.
Total	233	231	0.226	99	100	0.643

Notes: Answering the Socio-Demographic Questions was voluntary. In the table, 'CQ' refers to 'control question'.

[a] The p-values are based on two-sided Wilcoxon-Mann-Whitney (age) or Fisher's exact tests (all other variables) that compared the answers between treatments.

Table 5: Answers to the Socio-Demographic Questions by Treatment, Version, and Experiment (Part 4)

	Choice experiment (no CQs.)			Norm elicitation experiment		
	Individualizing treatment	Binding treatment	P-value[a]	Individualizing treatment	Binding treatment	P-value[a]
Income						
$0 - $10,000	29	31	.	18	23	.
$10,000 - $25,000	31	35	.	15	22	.
$25,000 - $50,000	73	61	.	30	24	.
$50,000 - $75,000	39	41	.	20	18	.
$75,000 - $100,000	26	35	.	4	7	.
$100,000 - $150,000	20	19	.	5	4	.
More than $150,000	14	10	.	7	2	.
Total	232	232	0.753	99	100	0.392
Community						
Urban	69	70	.	29	33	.
Suburban	124	130	.	53	54	.
Rural	40	33	.	17	13	.
Total	233	233	0.688	99	100	0.700
Religion						
Christian	108	97	.	43	37	.
Atheist	57	65	.	31	28	.
Jew	3	4	.	3	2	.
Buddhist	6	4	.	2	2	.
Muslim	2	1	.	0	1	.
Other	55	62	.	20	29	.
Total	232	233	0.788	99	99	0.646
Politics						
Very liberal	39	43	.	19	18	.
Liberal	93	77	.	35	38	.
Moderate	68	76	.	29	33	.
Conservative	28	29	.	12	10	.
Very conservative	5	8	.	4	1	.
Total	233	233	0.585	99	100	0.708

Notes: Answering the Socio-Demographic Questions was voluntary. In the table 'CQ' refers to 'control question'.

[a] The p-values are based on Fisher's exact tests that compared the answers between treatments.

or the frequency distribution of answers to all socio-demographic questions (age, gender, ethnicity, degree, employment, income, community, religion, politics). We compared all moral and socio-demographic characteristics between the individualizing and binding treatment of all versions of the choice experiment and the norm elicitation experiment. In doing so, we used the appropriate test to either test for differences in mean (two-tailed Wilcoxon-Mann-Whitney test) or frequency distribution (Fisher's exact test). To account for multiple testing (68 comparisons in total / 17 per experiment), we applied the Bonferroni correction to our statistical significance threshold of $p = 0.05$. The new threshold became $p = 0.0007$. We did not detect any significant differences between treatments in any of the versions of the choice experiment or in the norm elicitation experiment (in all cases, $p > 0.053$).

B.2 Effect of the Number of Balls in the Blue Bucket on Social Appropriateness Ratings

Table 6: The Effect of Rule Compliance on Social Appropriateness

	Social appropriateness
Balls in the blue bucket	0.048^{***}
	(0.007)
Constant	-0.373^{***}
	(0.080)
Log-likelihood	-1,213
BIC	2,446
Number of cases	1,000

Notes: We used Tobit regressions with standard errors clustered at the level of the individual. The standard errors are displayed in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

As an intuition check, we also tested whether the social appropriateness of the actions in the group rule-following task were linearly increasing with the number of balls placed in the rule-compliant blue bucket. Placing less than 20 balls into the blue bucket constituted partial rule compliance and should have been perceived as proportionally less socially appropriate. We found evidence that the social appropriateness increased with the number of balls placed in the rule-compliant blue bucket.

To examine this issue, We estimated one Tobit regression model with standard errors clustered at the individual level.[2] In the model, we regressed the social appropriateness of an action in the group rule-following

[2]We chose a Tobit regression model to handle the censoring at -1 and 1. However, the coefficient of "Balls in the blue bucket"

task against the number of balls placed in the blue bucket. We report the results in Table 6. As can be seen, the coefficient of the number of balls in the blue bucket was positive and highly significant ($p < 0.001$).

B.3 Effect of Progressivism's Sub-components on Group Identification: Individualizing Index and Binding Index, and Individual Moral Foundations

Table 7: Decomposition of the Effect of Moral Similarity on Group Identification (Individualizing and Binding Indices)

	Identification (distance)		
	All control questions	Limited control questions	No control questions
Individualizing index	0.098^{***}	0.035	0.043
	(0.025)	(0.022)	(0.023)
Binding index	0.028	0.037^{**}	0.025
	(0.017)	(0.014)	(0.014)
Distance individualizing index	-0.185^{***}	-0.128^{***}	-0.090^{***}
	(0.030)	(0.026)	(0.026)
Distance binding index	-0.098^{***}	-0.081^{***}	-0.115^{***}
	(0.024)	(0.019)	(0.019)
Constant	0.486^{***}	0.718^{***}	0.678^{***}
	(0.101)	(0.084)	(0.086)
Log-likelihood	-33	-54	-48
BIC	100	146	132
Number of cases	320	466	466

Notes: We used Tobit regressions. The standard errors are displayed in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

The progressivism index is a composite index. It describes the degree to which individuals endorse individualizing foundations (harm and fairness) over binding foundations (ingroup, authority, and purity). We replicated our analysis from the main section with the *sub-components* contained in the progressivism index. We found that the distance in the individualizing and the binding index both influenced group identification. Furthermore, this effect was mainly driven by the distance in the fairness foundation and, to a lesser degree, by the distance in the ingroup and purity foundations.

We first examined how distance in the individualizing and the binding index influenced group identification. To do so, we estimated three Tobit regression models. One for each choice experiment version. In all models, we regressed the distance group identification measure of the "Inclusion of Ingroup in the Self Scale"

would have been significant and positive in an OLS regression model as well (see Appendix A).

Table 8: Decomposition of the Effect of Moral Similarity on Group Identification (Harm, Fairness, Ingroup, Authority, and Purity)

	Identification (distance)		
	All control questions	Limited control questions	No control questions
Harm	0.061^{*}	0.006	0.043
	(0.026)	(0.024)	(0.023)
Fairness	0.046	0.022	0.003
	(0.028)	(0.023)	(0.023)
Ingroup	0.003	0.049^{**}	0.006
	(0.023)	(0.017)	(0.019)
Authority	0.024	−0.012	0.012
	(0.025)	(0.020)	(0.021)
Purity	−0.011	−0.000	0.001
	(0.019)	(0.015)	(0.015)
Distance harm	−0.023	−0.034	−0.029
	(0.031)	(0.027)	(0.027)
Distance fairness	-0.165^{***}	-0.099^{***}	-0.081^{**}
	(0.033)	(0.028)	(0.028)
Distance ingroup	−0.028	-0.053^{**}	-0.046^{*}
	(0.022)	(0.017)	(0.019)
Distance authority	−0.034	0.004	−0.042
	(0.030)	(0.024)	(0.025)
Distance purity	−0.040	-0.042^{*}	-0.045^{*}
	(0.024)	(0.019)	(0.018)
Constant	0.496^{***}	0.799^{***}	0.728^{***}
	(0.107)	(0.087)	(0.089)
Log-likelihood	-32	-48	-43
BIC	133	170	160
Number of cases	320	466	466

Notes: We used Tobit regressions. The standard errors are displayed in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

against the subject's individualizing index, the subject's binding index, the absolute difference between the individual's individualizing index and the social group's implied individualizing index, and the absolute difference between the individual's binding index and the social group's implied binding index. We report the results in Table 7. As can be seen, the coefficients for the distance in the individualizing and the binding index were negative and significant in all models ($p < 0.001$). This result suggests that both the similarity in the individualizing and the binding moral values affected group identification.

To examine how the distance in the individual moral foundations influenced group identification, we estimated three additional Tobit regression models. Again, one for each choice experiment version. In all models, we regressed the distance group identification measure of the "Inclusion of Ingroup in the Self Scale" against the subject's scores in each of the five moral foundations (harm, fairness, authority, ingroup, and purity) and the absolute difference between the individual's score and the implied social group's score in each of the five moral foundations. We report the results in Table 8. As can be seen, distance fairness was negative and significant in all models ($p < 0.004$). Furthermore, distance ingroup ($p < 0.015$), and distance purity ($p < 0.023$) were significant and negative in two models each.[3]

B.4 Effect of Alternative Moral Similarity Indices on Group Identification: Euclidean Distance and Chebyshev Distance

This subsection examines the robustness of the effect of moral similarity on group identification by using different aggregation methods than our progressivism index. We found that the effect of moral similarity on group identification was robust to changes in the conceptualization of moral similarity.

In the study's main section, we conceptualized morality with the five foundations from moral foundations theory (harm, fairness, ingroup, authority, and purity) and used the progressivism index. That index aggregates all five foundations into one index by subtracting the individualizing foundations (harm and fairness) from the binding foundations (ingroup, authority, and purity). To create one index of (inverse) moral similarity, we calculated the absolute difference between the individual's progressivism and the social group's implied progressivism. However, we could have also calculated different moral similarity indices based on moral foundations theory. We could have calculated the Euclidean[4] and Chebyshev[5] distance between the

[3] Dehghani et al. (2016) used the moral foundations dictionary to measure the moral foundations of Twitter users as implied by their Tweets. They found that the distance in the purity, but not in any other moral foundation, predicted the distance between two people in the social network. Using lab studies, they showed that the distance in purity between two subjects predicted the preferred self-reported social and physical distance. Contrary to these results, we only observed a relatively weak effect of the distance in purity on group identification in two choice experiment version.

[4] Euclidean distance puts equal weight on the distance of all five moral foundations. Unlike progressivism, Euclidean distance does not impose a distinction between individualizing and binding moral values.

[5] Chebyshev distance only considers the moral foundation that differs the most between the individual and the social group.

Table 9: The Effect of Alternative Moral Similarity Metrics on Group Identification

	Identification (distance)		
	All control questions	Limited control questions	No control questions
Progressivism	0.006	-0.027^{*}	-0.010
	(0.016)	(0.013)	(0.012)
Euclidean distance	-0.091^{***}	-0.070^{***}	-0.086^{***}
	(0.017)	(0.013)	(0.013)
Constant	0.913^{***}	0.982^{***}	0.949^{***}
	(0.046)	(0.036)	(0.036)
Log-likelihood	-49	-61	-51
BIC	121	146	126
Number of cases	320	466	466
	Identification (distance)		
	All control questions	Limited control questions	No control questions
Progressivism	-0.004	-0.030^{*}	-0.015
	(0.016)	(0.013)	(0.014)
Chebyshev distance	-0.094^{***}	-0.092^{***}	-0.106^{***}
	(0.028)	(0.021)	(0.020)
Constant	0.874^{***}	0.974^{***}	0.930^{***}
	(0.051)	(0.038)	(0.040)
Log-likelihood	-56	-63	-56
BIC	134	150	137
Number of cases	320	466	466

Notes: We used Tobit regressions. The standard errors are displayed in parentheses.

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

individual's moral position and the social group's moral position.

To examine how using these distance measures would have influenced our results, we estimated six Tobit regression models with robust standard errors. One for each moral similarity index and for each choice experiment version. In all models, we regressed the distance group identification measure of the "Inclusion of Ingroup in the Self Scale" against the subject's progressivism and the respective moral similarity index (Euclidean or Chebyshev distance between the individual's moral position and the social group's moral position). We report the results in Table 9. As can be seen, both the Euclidean distance and the Chebyshev distance were negative and highly significant in all models ($p < 0.001$).

B.5 Effect of Treatment on Alternative Group Identification Measures: Overlap and Survey Question

This subsection examines the robustness of the effect of treatment on group identification. In Hypothesis 2, we conceptualized group identification with the distance measure of the "Inclusion of Ingroup in the Self Scale." Using two other group identification measures, we replicated the initial analysis. We did not find any noticeable difference due to changes in the conceptualization of group identification.

Instead of the distance measure of the "Inclusion of Ingroup in the Self Scale", we could have used its overlap measure. This measure measures group identification as the percentage of **overlap** between the "You" and the "Group A" circle. The overlap identification measure can take on values ranging from 0 (weak identification / the circles do not overlap) to 1 (strong identification / the "You" circle is contained in the "Group A" circle).[6] We could have also used the identification question from the **questionnaire** at the end of our study (see Appendix C.8). In it, we asked participants to rate how closely attached they felt to Group A. The normalized question identification measure could take on values ranging from 0 ("Not at all attached") to 1 ("very strongly attached").[7]

To test whether different measures of group identification are consequential to our results, we replicated our analysis from Hypothesis 2 with the overlap and question identification measures. In particular, we determined the average identification according to each measure for each treatment of each choice experiment version. Subsequently, we used a one-sided t-test to compare the respective identification measure between

---

Therefore, it only takes into account the most salient difference between the individual and the social group.

[6] The literature of identity fusion (Jiménez et al., 2016; Swann, Gómez, Seyle, Morales, & Huici, 2009) assumes that an overlap between the circles symbolizes that personal and social image are aligned about the group attitude. Since participants in our experiments were introduced to an unknown group, this appeared to be unlikely. However, this measure presumably captures an extreme form of group identification.

[7] The question was unincentivized and took place after the experiment. Consequently, while answering, participants might have desired to appear consistent with their behavior in the group rule-following task.

Table 10: Comparison of Alternative Group Identification Measures Between Individualizing and Binding Treatments in the Choice Experiment Versions

	Identification (overlap)		
Choice experiment version	Individualizing treatment	Binding treatment	P-value[a]
All control questions	0.28	0.07	0.001
Limited control questions	0.34	0.17	0.001
No control questions	0.30	0.12	0.001
	Identification (question)		
Choice experiment version	Individualizing treatment	Binding treatment	P-value[a]
All control questions	0.54	0.34	0.001
Limited control questions	0.54	0.39	0.001
No control questions	0.46	0.35	0.001

Notes: The tables display the average group identification according to the overlap measure from the "Inclusion of Ingroup in the Self Scale" and the identification question from the choice experiment's questionnaire.

[a] The p-values are based on the preregistered one-sided t-tests that compared answers between treatments.

treatments of a choice experiment version. We report the results in Table 10. As can be seen, the average overlap identification measure is significantly higher in the individualizing treatment (0.28 - 0.34) compared to the binding treatment (0.07 - 0.17) in all choice experiment versions ($p < 0.001$). Furthermore, the average question identification measure is significantly higher in the individualizing treatment (0.46 - 0.54) compared to the binding treatment (0.34 - 0.39) in all choice experiment versions ($p < 0.001$).

B.6 Effect of Alternative Group Identification Measures on Rule Compliance: Overlap and Survey Question

This subsection examines the robustness of the effect of group identification on rule compliance. In Hypothesis 4, we conceptualized group identification with the distance measure of the "Inclusion of Ingroup in the Self Scale." Using two other group identification measures, we replicated the initial analysis. The replications suggest that our results are largely unaffected by the conceptualization of group identification.

To examine how group identification's conceptualization influenced our results, we replicated our analysis from Hypothesis 4 with the **overlap** measure from the same scale or the identification question from the **questionnaire** at the end of our study (for more information, see Appendix B.5). We estimated six Tobit regression models. One for each identification measure and for each choice experiment version. In all models,

Table 11: The Effect of Alternative Group Identification Measures on Rule Compliance

	Balls in the blue bucket		
	All control questions	Limited control questions	No control questions
Identification (overlap)	6.428^{*}	6.147^{**}	3.527
	(2.954)	(1.945)	(2.061)
Constant	4.258^{***}	6.131^{***}	4.455^{***}
	(1.169)	(0.864)	(0.851)
Log-likelihood	-742	-1,156	-1,129
BIC	1,500	2,330	2,276
Number of cases	320	466	466
	Balls in the blue bucket		
	All control questions	Limited control questions	No control questions
Identification (question)	27.506^{***}	19.265^{***}	22.522^{***}
	(3.475)	(2.466)	(2.460)
Constant	-6.648^{***}	-1.271	-4.005^{**}
	(1.876)	(1.348)	(1.268)
Log-likelihood	-710	-1,129	-1,085
BIC	1,436	2,277	2,188
Number of cases	320	466	466

Notes: We used Tobit regressions. The standard errors are displayed in parentheses.

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

we regressed the number of balls in the rule-compliant bucket of the group rule-following task against the respective identification measure (overlap or question). We report the results in Table 11. As can be seen, the overlap identification measure was significant in the all control questions and limited control questions versions ($p < 0.030$), but not the no control questions version ($p = 0.088$). The question identification measure was significant in all choice experiment versions ($p < 0.001$). However, while almost all models have significant coefficients, the size of the estimated coefficients changes with the used identification measure. If group identification increased by 1, the number of balls in the rule-compliant bucket, ceteris paribus, increased between 3.53 and 6.43 with the overlap measure, between 8.69 and 14.13 with the distance measure (see Hypothesis 4), or between 19.26 and 27.51 with the question measure. The reasons for these findings might have been that the overlap measure only captured an extreme form of group identification and that the question measure also captured the desire to appear consistent with previous choices in the group rule-following task. Furthermore, the overlap identification measure might have been insignificant in the no control questions choice experiment version because the lack of control questions reduced the salience of the moral position of the group. However, in general, it seems that our findings were mostly robust to changes in the conceptualization of group identification.

B.7 Analysis of the Behavior of Participants with a Progressivism Score Above 1.17 in the Limited Control Questions Choice Experiment Version

In this subsection, we replicate the analysis with observations from the limited control questions version, which had a progressivism score above 1.17. We preregistered this analysis for the following reason: Participants that had a progressivism score between 0 and 1.17 were morally more similar to the binding rather than to the individualizing social group (implied progressivism score of the individualizing social group: 2.33 / implied progressivism score of the binding social group: 0). Consequently, those participants should have identified more strongly with the social group and placed more balls in the blue bucket of the rule-following task if assigned to the binding rather than the individualizing social group. Such a behavior would have counteracted the predicted treatment effect described in Hypothesis 2 and 3. The reason why we used such an experimental design to test our hypothesis was that we initially assumed that the progressivism score of the binding group in the rule elicitation survey would be much lower and, thus, that participants would always be morally closer to the individualizing rather than the binding social group. We only report the results in the appendix since restricting the sample reduces the number of observations from 466 to 246.

- **Hypothesis 1 (Comparison of the Group Norms By Treatment):** We did not collect additional sessions of the norm elicitation experiment.
- **Hypothesis 2 (Comparison of Self-Reported Group Identification By Treatment):** The average group identification was 0.86 and 0.66 in the individualizing and binding treatments, respectively. The 0.20 point difference was significant according to a one-sided t-test ($p < 0.001$).

- **Section 6.2.1 (The Effect of Moral Similarity on Group Identification):** We re-estimated the Tobit regression model. The coefficient of progressivism was insignificant ($b = 0.011$, $p = 0.690$), whereas the coefficient of distance progressivism was negative and significant ($b = -0.129$, $p < 0.001$).
- **Hypothesis 3 (Comparison of the Number of Balls in the Rule-Compliant Bucket):** The participants placed, on average, 9.38 and 7.71 balls in the rule-compliant bucket of the individualizing and the binding treatment, respectively. The 1.67 balls difference was insignificant according to a one-sided t-test ($p = 0.052$). We attribute the lack of significance to the reduced number of observations.[8]
- **Hypothesis 4 (Testing The Effect of Group Identification on Norm Sensitivity):** We re-estimated Tobit regression model. The coefficient of identification (distance) was positive and significant ($b = 9.855$, $p = 0.026$).
- **Appendix A 6. Hypothesis 2 (Comparison of Self-Reported Group Identification By Treatment) with Wilcoxon-Mann-Whitney tests** The difference in identification is also significant when using a one-sided Wilcoxon-Mann-Whitney test ($p < 0.001$).
- **Appendix A 9. Hypothesis 3 (Comparison of the Number of Balls in the Rule-Compliant Bucket) with Wilcoxon-Mann-Whitney tests** The difference in the number of balls is also insignificant when using a one-sided Wilcoxon-Mann-Whitney test ($p = 0.071$).
- **Appendix A 11. Hypothesis 4 (Testing the Effect of Group Identification on Norm Sensitivity) with the binding treatment dummy)** The coefficient of group identification is no longer significant when adding a binding treatment dummy ($b = 8.442$, $p = 0.080$).
- **Appendix A 14. Effect of Distance Progressivism on Group Identification with the Binding Treatment Dummy** The coefficient of distance progressivism is also significant when adding a binding treatment dummy ($b = -0.090$, $p = 0.013$).
- **Appendix B.3 Effect of Progressivism's Sub-components on Group Identification: Individualizing Index and Binding Index, and Individual Moral Foundations:** We re-estimated the Tobit regression model of both decompositions. In the first decomposition (into the individualizing index and binding index), the coefficients of the the following variables were insignificant: binding index ($b = 0.015$, $p = 0.614$). Whereas, the coefficients of the following variables were positive and significant: individualizing index ($b = 0.091$, $p = 0.027$). Furthermore, the coefficients of the following variables were negative and significant: distance individualizing index ($b = -0.183$, $p < 0.001$) and distance binding

[8] While excluding observations below the progressivism threshold had little effect on the effect size (all observations: 0.209 / above threshold: 0.208), it almost halved the number of usable observations per treatment (all observations: 233 / above threshold: 123). As a consequence, the power of the t-test decreased from 0.728 in the comparison with all observations to 0.493 in the comparison with only those observations above the progressivism threshold.

index ($b = -0.123$, $p < 0.001$). In the second decomposition (into harm, fairness, authority, ingroup, and purity), the coefficients of the following variables were insignificant: harm ($b = 0.029$, $p = 0.452$), fairness ($b = 0.039$, $p = 0.307$), ingroup ($b = 0.041$, $p = 0.109$), authority ($b = -0.012$, $p = 0.644$), purity ($b = -0.015$, $p = 0.519$), distance harm ($b = -0.015$, $p = 0.736$), distance authority ($b = -0.025$, $p = 0.440$), and distance purity ($b = -0.051$, $p = 0.071$). Whereas the following variables were negative and significant: Distance fairness ($b = -0.091$, $p < 0.001$), and distance ingroup ($b = -0.091$, $p < 0.001$).

- **Appendix B.4 Effect of Alternative Moral Similarity Indices on Group Identification: Euclidean Distance and Chebyshev Distance:** We re-estimated the Tobit regression model. In the first model, the coefficient of progressivism was insignificant ($b = 0.013$, $p = 0.643$), whereas the coefficient of Euclidean distance was negative and significant ($b = -0.144$, $p < 0.001$). In the second model, the coefficient of progressivism was insignificant ($b = -0.018$, $p = 0.513$), whereas the coefficient of Chebyshev distance was negative and significant ($b = -0.180$, $p < 0.001$).

- **Appendix B.5 Effect of Treatment on Alternative Group Identification Measures: Overlap and Survey Question:** The average overlap identification measure was 0.39 and 0.12 in the individualizing and binding treatments, respectively. The 0.27 point difference was significant according to a one-sided t-test ($p < 0.001$). The average question identification measure was 0.53 and 0.35 in the individualizing and binding treatments, respectively. The 0.19 point difference was significant according to a one-sided t-test ($p < 0.001$).

- **Appendix B.6 Effect of Alternative Group Identification Measures on Rule Compliance: Overlap and Survey Question:** We re-estimated the Tobit regression models. In the first model, the coefficient of identification (overlap) was positive and significant ($b = 8.457$, $p = 0.006$). In the second model, the coefficient of identification (question) was positive and significant ($b = 24.111$, $p < 0.001$).

Appendix C Experimental Materials

In the following subsections, we report the instructions of our online experiments. With this, it should be noted that: 1.) Everything below a screen tag (e.g., "**Screen 1:**") was displayed on the same screen. 2.) Everything below a horizontal separation line was displayed in one text box. If the separation line includes centered text, the text was used as a text box title. 3.) Input items are always described by stating the used questions ("Q") and all possible answers among which the participants could choose ("A"). 4.) In some cases we include comments (e.g., "[*Example comment*]") that were not included in the online experiments. 5.) We do not include small details (e.g., buttons and their labels).

C.1 Moral Foundations Questionnaire

Screen 1:

Information:

In Part 1, we want you to answer a questionnaire. Please read the questions carefully and answer truthfully.

When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking? Please rate each statement using this scale:

[5] = Extremely relevant (This is one of the most important factors when I judge right and wrong)

[4] = Very relevant

[3] = Somewhat relevant

[2] = Slightly relevant

[1] = Not very relevant

[0] = Not at all relevant (This consideration has nothing to do with my judgments of right and wrong)

Q1: "Whether or not someone suffered emotionally"

Q2: "Whether or not some people were treated differently than others"

Q3: "Whether or not someone's action showed love for his or her country"

Q4: "Whether or not someone showed a lack of respect for authority "

Q5: "Whether or not someone violated standards of purity and decency"

Q6: "Whether or not someone was good at math"

Q7: "Whether or not someone cared for someone weak or vulnerable"

Q8: "Whether or not someone acted unfairly"

Q9: "Whether or not someone did something to betray his or her group"

Q10: "Whether or not someone conformed to the traditions of society"

Q11: "Whether or not someone did something disgusting"

Q12: "Whether or not someone was cruel"

Q13: "Whether or not someone was denied his or her rights"

Q14: "Whether or not someone showed a lack of loyalty"

Q15: "Whether or not an action caused chaos or disorder"

Q16: "Whether or not someone acted in a way that God would approve of"

A: "Extremely relevant", "Very relevant", "Somewhat relevant", "Slightly relevant", "Not very relevant" or "Not at all relevant"

Screen 2:

Information:

In Part 1, we want you to answer a questionnaire. Please read the questions carefully and answer truthfully.

Please read the following sentences and indicate your agreement or disagreement:

Q17: "Compassion for those who are suffering is the most crucial virtue"

Q18: "When the government makes laws, the number one principle should be ensuring that everyone is treated fairly"

Q19: "I am proud of my country's history"

Q20: "Respect for authority is something all children need to learn"

Q21: "People should not do things that are disgusting, even if no one is harmed"

Q22: "It is better to do good than to do bad"

Q23: "One of the worst things a person could do is hurt a defenseless animal"

Q24: "Justice is the most important requirement for a society"

Q25: "People should be loyal to their family members, even when they have done something wrong"

Q26: "Men and women each have different roles to play in society"

Q27: "I would call some acts wrong on the grounds that they are unnatural"

Q28: "It can never be right to kill a human being"

Q29: "I think it's morally wrong that rich children inherit a lot of money while poor children inherit nothing"

Q30: "It is more important to be a team player than to express oneself"

Q31: "If I were a soldier and disagreed with my commanding officer's orders, I would obey anyway because that is my duty"

Q32: "Chastity is an important and valuable virtue"

A: "Strongly agree", "Moderately agree", "Slightly agree", "Slightly disagree", "Moderately disagree" or "Strongly disagree"

[*We adopted the moral foundations questionnaire from Graham et al. (2011). The questionnaire consisted of two parts: The moral relevance section and the moral judgment section. We converted the categorical ratings of both sections into integers from* 0 *("Not at all relevant" / "Strongly disagree") to* 5 *("Extremely relevant" / "Strongly agree"). The questions were either included to measure the harm foundation (Q1, Q7, Q12, Q17, Q23, Q28), the fairness foundation (Q2, Q8, Q13, Q18, Q24, Q29), the ingroup foundation (Q3, Q9, Q14, Q19, Q25, Q30), the authority foundation (Q4, Q10, Q15, Q20, Q26, Q31), the purity foundation (Q5, Q11, Q16, Q21, Q27, Q32) or as attention checks (Q6, Q22). The foundations were calculated by taking the mean of all contained items. In addition, we calculated the individualizing index by averaging the score of the harm and fairness foundations and the binding index by averaging the scores of the ingroup, authority and purity foundations. Lastly, we calculated the progressivism index by subtracting the individualizing index from the binding index (Clark et al., 2017). This index could take on values ranging from* -5 *(not at all individualizing & extremely binding) to* $+5$ *(extremely individualizing & not at all binding).*]

C.2 Rule Elicitation Survey

Screen 1:

Information:

In Part 1, we want you to read about a situation that other people will be facing in the future. While you yourself will **not** be in this situation, we want you to read about it, and then we will ask you a question. At the following stage, you will be able to reopen the text describing the situation.

Please take a moment to read about this situation:

In today's study, you will decide how to allocate 20 balls between two buckets. Your task is to put each of the balls, one-by-one, into one of the two buckets: the blue bucket or the yellow bucket. The balls will appear in the center of your screen, and you can allocate each ball by clicking and dragging it to the bucket of your choice. For each ball you put in the blue bucket, you will receive £0.05, and for each ball you put in the yellow bucket, you will receive £0.10.

The rule is to put the balls in the blue bucket

Your payment will be based on your decisions: it is the sum of payments from the blue and yellow buckets.

Screen 2:

Information:

In Part 1, we want you to answer a question. Please read the question carefully and answer truthfully.

Q: **"If you had to describe the rule to other study participants, how would you describe it?"**

A: "The rule is to put the balls in the blue bucket" or "The rule is to put the balls in the yellow bucket"

C.3 Introduction to Group A

Screen 1:

Information:

In Part 1, we want to introduce you to another group of participants. During the rest of this study, we will call this group **"Group A"**. Some time ago, Group A took part in another study. Among other things, this study included a questionnaire. Below you can see the answers of Group A to a small selection of five questions. In each case, the answer is the rating that is closest to the average rating that we received from all members of Group A. We want you to carefully read and memorize the answers provided by members of Group A. [The following three sentences are only included in the all and limited control questions version.] On the next page, we will show you one question, and we will ask you to select the same response as members of Group A. If you choose the correct answer, you will receive a bonus of £0.50! We will tell you whether you were right at the end of the study.

When deciding whether something is right or wrong, to what extent are the following considerations relevant to the thinking of Group A? They rated each statement with the following scale:

[5] = Extremely relevant (This is one of the most important factors when I judge right and wrong)

[4] = Very relevant

[3] = Somewhat relevant

[2] = Slightly relevant

[1] = Not very relevant

[0] = Not at all relevant (This consideration has nothing to do with my judgments of right and wrong)

---

1. Whether or not someone cared for someone weak or vulnerable: [*In the individualizing treatment:* Very Relevant / *in the binding treatment:* Somewhat Relevant]
2. Whether or not some people were treated differently from others: [*In the individualizing treatment:* Very Relevant / *in the binding treatment:* Somewhat Relevant]
3. Whether or not someone's action showed love for his or her country: [*In the individualizing treatment:* Not very relevant / *in the binding treatment:* Somewhat Relevant]
4. Whether or not someone showed a lack of respect for authority: [*In the individualizing treatment:* Slightly relevant / *in the binding treatment:* Somewhat Relevant]
5. Whether or not someone violated standards of purity and decency: [*In the individualizing treatment:* Slightly relevant / *in the binding treatment:* Somewhat Relevant]

Screen 2:

[This screen is only included in the all and limited control questions version.]

Information:

In the box below, you can see one of the five questions presented on the last page. Please, select the rating that is closest to the average rating that we received from all members of Group A. If you choose the correct answer, you will receive a bonus of £0.50! We will tell you whether you were right at the end of the study.

When deciding whether something is right or wrong, to what extent is the following consideration relevant to the thinking of Group A? Please rate the statement in the same way as members of Group A did by using this scale:

[5] = Extremely relevant (This is one of the most important factors when they judge right and wrong)

[4] = Very relevant

[3] = Somewhat relevant

[2] = Slightly relevant

[1] = Not very relevant

[0] = Not at all relevant (This consideration has nothing to do with their judgments of right and wrong)

Q: "Whether or not someone cared for someone weak or vulnerable:", "Whether or not some people were treated differently from others:", "Whether or not someone's action showed love for his or her country:", [The following question is only included in the all control questions version.] "Whether or not someone

showed a lack of respect for authority:" or "Whether or not someone violated standards of purity and decency:"

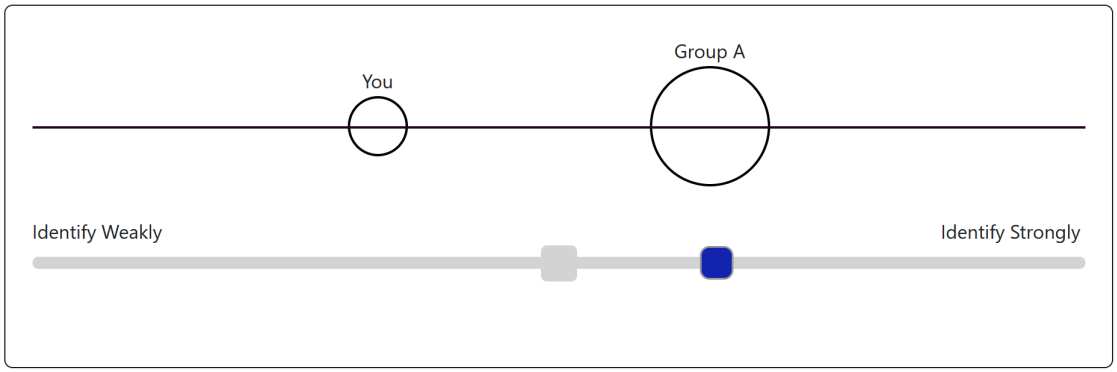
A: "Extremely relevant", "Very relevant", "Somewhat relevant", "Slightly relevant", "Not very relevant" or "Not at all relevant."

C.4 Group Identification Task

Screen 1:

Information:

In Part 2, we want to know your identification with Group A. On the scale below you can see two circles: One representing you and one representing Group A. You can change the distance between the two circles by moving the blue slider on the scale. Moving the slider to the left moves the circles further apart and signals that you identify very little with Group A. Moving the slider to the right moves the circles closer to each other and signals that you identify very much with Group A. Please move the slider to a position in which the distance between the two circles most accurately describes your identification with Group A. Please zoom out if the scale is only displayed partially.



C.5 Rule Following Task

Screen 1:

Information:

In Part 3, you are going to play a game. You will receive the instructions for this game on the next page. Some parts of these instructions have been select by members of Group A. In addition to answering the questions that were presented to you in Part 1, the members of Group A also were asked to select a description for the game you are about to play. The description that was chosen by the majority was included in the instruction

you are about to receive.

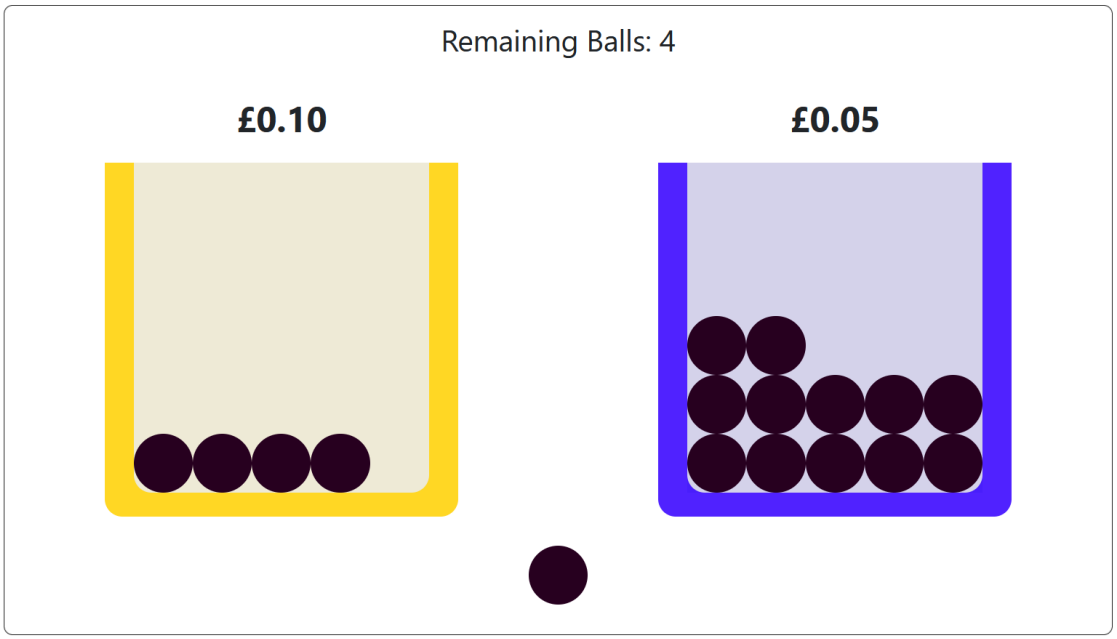
Screen 2:

Information:

In Part 3 of this study, you will decide how to allocate 20 balls between two buckets. Your task is to put each of the balls, one-by-one, into one of the two buckets: the blue bucket or the yellow bucket. The balls will appear in the center of your screen, and you can allocate each ball by clicking and dragging it to the bucket of your choice. For each ball you put in the blue bucket, you will receive £0.05, and for each ball you put in the yellow bucket, you will receive £0.10.

According to the members of Group A, the rule is to put the balls into the blue bucket

Your payment from Part 3 will be based on your decisions: it is the sum of payments from the blue and yellow buckets. After you have completed the task, a "Continue" button will be displayed at the bottom of the page.



C.6 Norm Elicitation

Screen 1:

Information:

In Part 2, we ask you to evaluate certain behaviors that may have occurred in another study. Up until this point, the participants of the other study had a very similar experience to you: Before participating

in the study, they answered the same questionnaire that you answered (more on this later), and they were introduced to the same Group A that you were introduced to in Part 1. We want you to read the instructions for one of their subsequent tasks. While you yourself will not play this game, you should read the instructions carefully. After you have read those instructions, you will receive further instructions that will detail your own task.

You will be able to reopen these instructions when you make your final decisions!

Please take a moment to read about the task that you have to evaluate:

Page 1:

In this part, you are going to play a game. You will receive the instructions for this game on the next page. Some parts of these instructions have been select by members of Group A. In addition to answering the questions that were presented to you in Part 1, the members of Group A also were asked to select a description for the game you are about to play. The description that was chosen by the majority was included in the instruction you are about to receive.

Page 2:

In today's study, you will decide how to allocate 20 balls between two buckets. Your task is to put each of the balls, one-by-one, into one of the two buckets: the blue bucket or the yellow bucket. The balls will appear in the center of your screen, and you can allocate each ball by clicking and dragging it to the bucket of your choice. For each ball you put in the blue bucket, you will receive £0.05, and for each ball you put in the yellow bucket, you will receive £0.10.

According to the members of Group A, the rule is to put the balls into the blue bucket

Your payment will be based on your decisions: it is the sum of payments from the blue and yellow buckets.

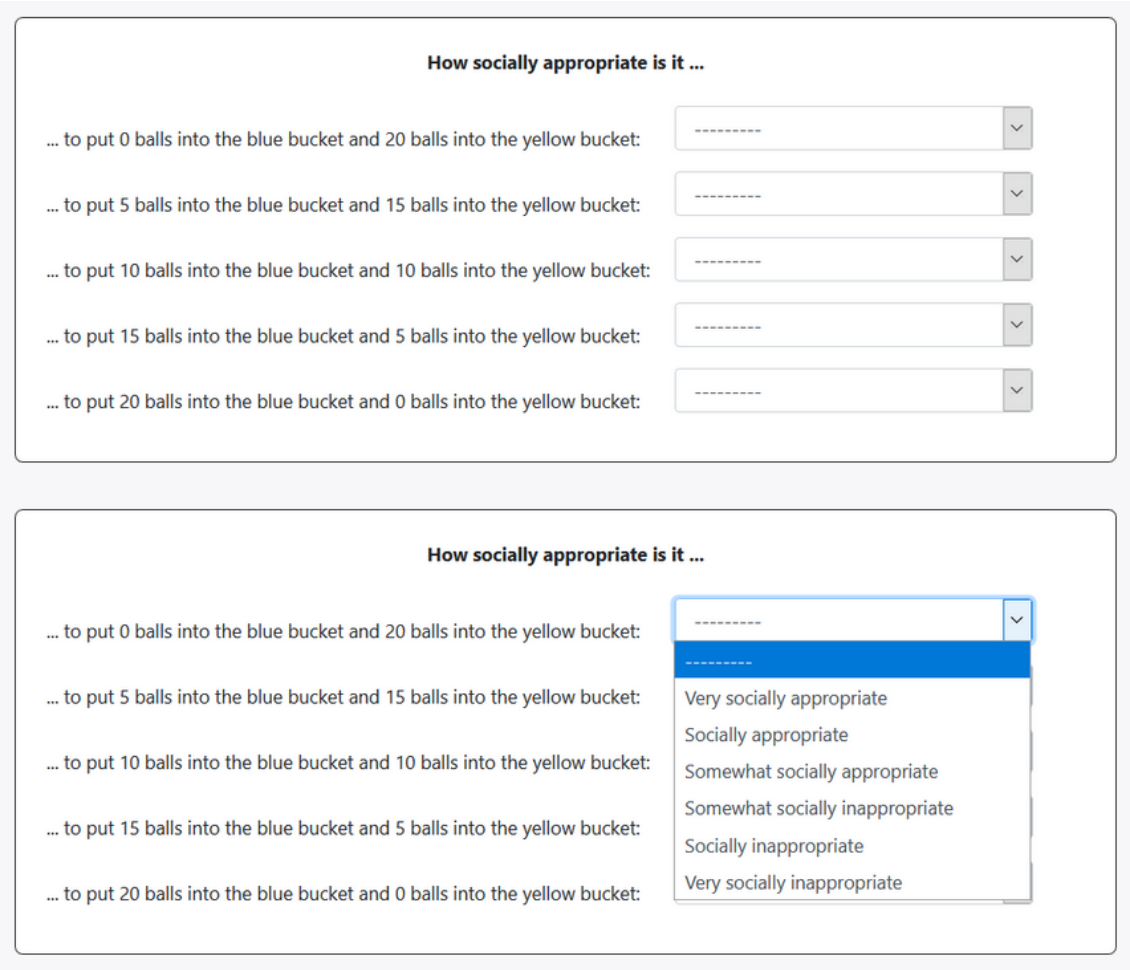
Screen 2:

Information:

Your task in Part 2 is to evaluate how socially appropriate it is to put 0, 5, 10, 15, or 20 balls into the blue bucket while placing the remaining balls into the yellow bucket. In your judgment, you can rate any behavior on a scale that ranges from "Very socially inappropriate" to "Very socially appropriate". Hereby, "Very socially inappropriate" indicates that the behavior is inconsistent with what most people expect an Individual ought to do, whereas "Very socially appropriate" indicates that the behavior was consistent with what most people expect an Individual ought to do. To further support you in your decision: A behavior might be considered as socially appropriate if most people agree that it is the "proper" thing to do, whereas a behavior might be considered as socially inappropriate if people might get angry if they observe a person doing it. You can find screenshots of your task and of your possible answers in the box below.

You will be able to reopen these instructions when you make your final decisions!

Screenshots of your upcoming task:



Screen 3:

Information:

But wait, there's more! We don't want you to tell us whether you personally believe that the behavior is socially appropriate or inappropriate. Instead, we want you to guess how most other's from a specific group of people would rate the behavior. In the following, we will call this specific group of people, the **"Matching Group"**. You should note that the Matching Group and Group A consist of different people. We will tell you more about the Matching Group on the next page.

Learning about your Matching Group is important because your earnings will depend on it. At the end of the study, we will randomly select one of the five situations (put 0, 5, 10, 15, or 20 balls into the blue bucket) for all participants. Then we will compare your evaluation in this situation with the most frequent evaluation that we received from all members of your Matching Group (this includes you). If you selected the same appropriateness rating as was most frequent in your Matching Group, you will receive a bonus of £2.00. If

you selected a different rating, you will earn no bonus.

You will be able to reopen these instructions when you make your final decisions!

Screen 4:

Information:

Now we will explain to you who the members of your Matching Group are. Before participating in this study, each participant answered the same questionnaire. This questionnaire contained two types of questions (**category 1** and **category 2**). The questions included in the two categories differ in that they are concerned with different aspects of morality. For some people, the considerations expressed in category 1 are more important than those expressed in category 2, whereas for other people, the opposite is true. Lastly, some people might value both categories equally. Your answers indicate that:

For you, the considerations of category 1 are more important to those of category 2!

We call the group of participants, for whom the statements from category 1 were more relevant and agreeable than those of category 2, your **Matching Group**. Please read the questions contained in each category carefully and try to understand what both categories express. It might help you to imagine what kind of person would prefer either category. It is crucial that you understand who is in your Matching Group since it will help you to give us the same response as to what most people from your Matching Group believe constitutes socially appropriate or socially inappropriate behavior. Consequently, it will make it more likely for you to receive the additional £2.00.

You will be able to reopen these instructions when you make your final decisions!

Category 1:

To what extent are the following considerations relevant to your thinking about right and wrong:

- Whether or not someone suffered emotionally
- Whether or not some people were treated differently than others
- Whether or not someone cared for someone weak or vulnerable
- Whether or not someone acted unfairly
- Whether or not someone was cruel
- Whether or not someone was denied his or her rights

Indicate your agreement or disagreement:

- Compassion for those who are suffering is the most crucial virtue.

- When the government makes laws, the number one principle should be ensuring that everyone is treated fairly.
- One of the worst things a person could do is hurt a defenseless animal.
- Justice is the most important requirement for a society.
- It can never be right to kill a human being.
- I think it's morally wrong that rich children inherit a lot of money while poor children inherit nothing.

Category 2:

To what extent are the following considerations relevant to your thinking about right and wrong:

- Whether or not someone's action showed love for his or her country
- Whether or not someone showed a lack of respect for authority
- Whether or not someone violated standards of purity and decency
- Whether or not someone did something to betray his or her group
- Whether or not someone conformed to the traditions of society
- Whether or not someone did something disgusting
- Whether or not someone showed a lack of loyalty
- Whether or not an action caused chaos or disorder
- Whether or not someone acted in a way that God would approve of

Indicate your agreement or disagreement:

- I am proud of my country's history.
- Respect for authority is something all children need to learn.
- People should not do things that are disgusting, even if no one is harmed.
- People should be loyal to their family members, even when they have done something wrong.
- Men and women each have different roles to play in society.
- I would call some acts wrong on the grounds that they are unnatural.
- It is more important to be a team player than to express oneself.
- If I were a soldier and disagreed with my commanding officer's orders, I would obey anyway because that is my duty.

- Chastity is an important and valuable virtue.

Screen 5:

Information:

Please, rate now how socially appropriate it is to put 0, 5, 10, 15, or 20 balls into the blue bucket while placing the remaining balls into the yellow bucket. Should you have any questions remaining, you can reopen any of the previous instructions by clicking on the buttons below.

How socially appropriate is it ...

Q1: "... to put 0 balls into the blue bucket and 20 balls into the yellow bucket:"

Q2: "... to put 5 balls into the blue bucket and 15 balls into the yellow bucket:"

Q3: "... to put 10 balls into the blue bucket and 10 balls into the yellow bucket:"

Q4: "... to put 15 balls into the blue bucket and 5 balls into the yellow bucket:"

Q5: "... to put 20 balls into the blue bucket and 0 balls into the yellow bucket:"

A: "Very socially appropriate", "Socially appropriate", "Somewhat socially appropriate", "Somewhat socially inappropriate", "Socially inappropriate" or "Very socially inappropriate"

C.7 Socio-Demographic Questionnaire

Screen 1:

Information:

Before we inform you about your earnings and transfer you back to Prolific, we would like to ask you some optional questions. In doing so, please consider that all your data is collected anonymously and will not affect your payment!

---

1. [*Age*]
   Q: "What is your age?'
   A: Any integer number
2. [*Gender*]
   Q: "What is your gender?"
   A: "Male", "Female" or "Other"
3. [*Ethnicity*]
   Q: "Which of the following best describes your ethnicity?"

A: “White”, “Black”, “Hispanic”, “Asian” or “Other”

4. [*Degree*]

   Q: “What is your highest obtained degree?”

   A: “None”, “High school diploma”, “Bachelor’s degree”, “Master’s degree” or “Doctorate degree”

5. [*Employment Status*]

   Q: “What is your current employment status?”

   A: “Full time employed”, “Part time employed”, “Unemployed”, “Self-employed”, “Student”, “Housewife / husband” or “Retired”

6. [*Income*]

   Q: “Approximately, how high is your yearly income in US Dollar (after taxes)?”

   A: “$0 - $10,000”, “$10,000 - $25,000”, “$25,000 - $50,000”, “$50,000 - $75,000”, “$75,000 - $100,000”, “$100,000 - $150,000” or “More than $150,000”

7. [*Community*]

   Q: “Which type of community do you live in?”

   A: “Urban”, “Suburban” or “Rural”

8. [*Religion*]

   Q: “How would you describe your religious views?”

   A: “Christian”, “Jew”, “Buddhist”, “Muslim”, “Hindu”, “Atheist” or “Other”

9. [*Politics*]

   Q: “How would you describe your political affiliation?”

   A: “Very liberal”, “Liberal”, “Moderate”, “Conservative” or “Very conservative”

C.8 Behavioral Questionnaire

Screen 1:

Information:

Before we inform you about your earnings and transfer you back to Prolific, we would like to ask you some additional questions. In doing so, please consider that all your data is collected anonymously and will not affect your payment!

1. [*Alternative Group Identification Measure*]

   Q: “On a scale from 1 to 10, please rate how closely attached you felt to Group A throughout the experiment (1: Not at all / 10: Very strongly)”

   A: Any integer number between 1 and 10 (including 1 and 10)

2. [*Reason for Rule Following*]

   Q: "In Part 3, why did or didn't you follow the rule to put all the balls into the blue bucket?"

   A: Open form text input

C.9 User Experience Questionnaire

Screen 1:

Information:

Before we inform you about your earnings and transfer you back to Prolific, we would like to ask you some additional questions. In doing so, please consider that all your data is collected anonymously and will not affect your payment!

1. [*Device*]

   Q: "What device did you use during this study?"

   A: "Desktop PC", "Notebook", "Tablet", "Cellphone" or "Other"

2. [*Browser*]

   Q: "What browser did you use during this study?"

   A: "Chrome", "Safari", "Firefox", "Other" or "Unknown"

The following questions are optional and can be left blank:

3. [*Understanding Problems*]

   Q: "If you had any problems of understanding during the experiment, please let us know in this field:"

   A: Open form text input

4. [*Technical Problems*]

   Q: "If you had any technical problems during the experiment, please let us know in this field:"

   A: Open form text input

5. [*Comments*]

   Q: "Do you have any further comments?"

   A: Open form text input

References

Clark, C. B., Swails, J. A., Pontinen, H. M., Bowerman, S. E., Kriz, K. A., & Hendricks, P. S. (2017). A behavioral economic assessment of individualizing versus binding moral foundations. *Personality and Individual Differences*, *112*, 49–54. doi:10/gfxk2c

Dehghani, M., Johnson, K., Hoover, J., Sagi, E., Garten, J., Parmar, N. J., … Graham, J. (2016). Purity homophily in social networks. *Journal of Experimental Psychology: General*, *145*(3), 366–375. doi:10/f8bqq8

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. doi:10/cq64hc

Jiménez, J., Gómez, Á., Buhrmester, M. D., Vázquez, A., Whitehouse, H., & Swann, W. B. (2016). The dynamic identity fusion index: A new continuous measure of identity fusion for web-based questionnaires. *Social Science Computer Review*, *34*(2), 215–228. doi:10/f8cwpx

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

Swann, W. B., Gómez, Á., Seyle, D. C., Morales, J. F., & Huici, C. (2009). Identity fusion: The interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology*, *96*(5), 995–1011. doi:10/dx5mf2