NOT FOR PUBLICATION

Supplementary Online Appendices

Women, Class, and Cooperation in Groups: Evidence from a Public Goods Experiment in Lebanon

Appendices

A Experimental Design **1**
- A.1 Main Study . . . 1
- A.2 Randomization . . . 1
- A.3 Recruitment . . . 2
- A.4 Possible concerns about selection into participation . . . 4
- A.5 Treatment Assignment Probabilities . . . 5

B Balance Checks **8**

C Comparison of Discussion Sample to Lebanese Population **10**

D Public Goods Game Implementation **13**
- D.1 Introducing the public goods game . . . 13
- D.2 Payoff structure of the game . . . 13
- D.3 Mitigating demand effects . . . 14

E Data and Estimation **16**
- E.1 Relation to the pre-analysis plan . . . 16

F Control Variables **17**

G Robustness Checks **19**

H Mechanism Analysis **20**
- H.1 Uncertainty or class bias . . . 20
- H.2 Lack of a common gender identity . . . 22

I Class versus Sectarian Differences **25**

A Experimental Design

A.1 Main Study

The public goods experiment featured in this paper was played as part of the baseline data collection for a separate study on the effects of political discussion in homogeneous versus heterogeneous class and sectarian groups on support for ethnic versus programmatic politics. This main experiment was pre-registered with the Evidence in Governance and Politics (EGAP) network.[1] The strategy for recruitment and randomization described below was driven by the needs for this main study (e.g. the fact that we have more men than women in the study, despite our focus on women's cooperation). The analysis in this paper focuses on the *class* composition treatments although we do some additional analysis with the cross-cutting *sectarian* composition treatments and therefore describe the complete experimental design below.

A.2 Randomization

We organized 120 discussion groups in the Beirut and Mount Lebanon areas in the spring of 2016. Individuals with different sectarian (Christian, Sunni, and Shia) and economic (lower and upper income) profiles were randomly assigned to participate in discussions that varied in their class and sectarian compositions. Assignment to the two treatments was orthogonal following a 2x2 factorial design with 30 groups in each cell. Specifically, participants were randomly assigned to one of four discussion group types: (1) homogeneous sect and class, (2) mixed sect, homogeneous class, (3) homogeneous sect, mixed class, and (4) mixed sect and class.

In homogeneous sectarian groups all six participants were either Christian, Sunni, or Shia. In mixed sectarian groups, two participants were Christian, two were Sunni, and two were Shia. In homogeneous class groups, all six participants were either lower or upper income. In mixed class groups, three participants were lower income and three were upper income (to see how we determined participant class, see Appendix A.3). Table A.1 provides a summary of the randomization while Table A.2 shows how class and sect combine for each of the 24 discussions in a set. The 120 discussion groups were organized in five sets of 24 discussion sessions (6 sessions x 4 group types).[2]

To obtain the correct group compositions, we in fact recruited and randomly assigned 1200 individuals of which 720 would ultimately participate and 480 would be 'back-ups'. Upon arrival at their scheduled discussion session, participants were checked in by staff and informed consent was administered. Participants were not designated as 'main' or 'backup' in advance and if extra participants arrived, those that were asked to stay were randomly selected. This was essential to ensure that those who participated in each discussion were a random sample of those who were assigned to that treatment condition. There were some issues in how the scheduling was implemented that could raise concerns about non-comparability of the treatment and control groups. We describe the issue below in Appendix A.4.

[1] While our analysis employs the same estimation strategy as pre-registered (see Appendix E), we did not pre-register that we were going to analyze the results of the public goods experiment by gender.

[2] A set of discussions was completed every 2-3 weeks between February and April 2016.

		Sectarian discussion	
		Homog.	Hetero.
Class discussion	Homog.	**Group 1** groups = 30 n = 180 Sect comp: 6 Sunni or 6 Christian or 6 Shia Class comp: All poor or all rich	**Group 2** groups = 30 n = 180 Sect comp: 2 Sunni, 2 Christian, and 2 Shia Class comp: All poor or all rich
	Hetero.	**Group 3** groups = 30 n = 180 Sect comp: 6 Sunni or 6 Christian or 6 Shia Class comp: 3 poor and 3 rich	**Group 4** groups = 30 n = 180 Sect comp: 2 Sunni, 2 Christian, and 2 Shia Class comp: 1 poor and 1 rich of each sect

Table A.1: Summary of randomization

Group type 1: Same sect, same class						**Group type 2: Mixed sect, same class**					
1	2	3	4	5	6	7	8	9	10	11	12
P. Sun.	P. Shi.	P. Chr.	R. Sun.	R. Shi.	R. Chr.	P. Sun.	P. Sun.	P. Sun.	R. Sun.	R. Sun.	R. Sun.
P. Sun.	P. Shi.	P. Chr.	R. Sun.	R. Shi.	R. Chr.	P. Sun.	P. Sun.	P. Sun.	R. Sun.	R. Sun.	R. Sun.
P. Sun.	P. Shi.	P. Chr.	R. Sun.	R. Shi.	R. Chr.	P. Shi.	P. Shi.	P. Shi.	R. Shi.	R. Shi.	R. Shi.
P. Sun.	P. Shi.	P. Chr.	R. Sun.	R. Shi.	R. Chr.	P. Shi.	P. Shi.	P. Shi.	R. Shi.	R. Shi.	R. Shi.
P. Sun.	P. Shi.	P. Chr.	R. Sun.	R. Shi.	R. Chr.	P. Chr.	P. Chr.	P. Chr.	R. Chr.	R. Chr.	R. Chr.
P. Sun.	P. Shi.	P. Chr.	R. Sun.	R. Shi.	R. Chr.	P. Chr.	P. Chr.	P. Chr.	R. Chr.	R. Chr.	R. Chr.
Group type 3: Same sect, mixed class						**Group type 4: Mixed sect, mixed class**					
13	14	15	16	17	18	19	20	21	22	23	24
P. Sun.	P. Sun.	P. Shi.	P. Shi.	P. Chr.	P. Chr.	P. Sun.	P. Sun.	P. Sun.	P. Sun.	P. Sun.	P. Sun.
P. Sun.	P. Sun.	P. Shi.	P. Shi.	P. Chr.	P. Chr.	R. Sun.	R. Sun.	R. Sun.	R. Sun.	R. Sun.	R. Sun.
P. Sun.	P. Sun.	P. Shi.	P. Shi.	P. Chr.	P. Chr.	P. Shi.	P. Shi.	P. Shi.	P. Shi.	P. Shi.	P. Shi.
R. Sun.	R. Sun.	R. Shi.	R. Shi.	R. Chr.	R. Chr.	R. Shi.	R. Shi.	R. Shi.	R. Shi.	R. Shi.	R. Shi.
R. Sun.	R. Sun.	R. Shi.	R. Shi.	R. Chr.	R. Chr.	P. Chr.	P. Chr.	P. Chr.	P. Chr.	P. Chr.	P. Chr.
R. Sun.	R. Sun.	R. Shi.	R. Shi.	R. Chr.	R. Chr.	R. Chr.	R. Chr.	R. Chr.	R. Chr.	R. Chr.	R. Chr.

Table A.2: Individual profiles by group type

A.3 Recruitment

Identifying lower and upper income participants. To determine whether potential participants were rich or poor for the class randomization, following extensive piloting, eight questions about economic status were asked on the screening survey and these were used to create an index (see Table A.3). The screening survey recorded answers about income, assets, leisure travel and dining, and electricity usage. Responses for each question were re-coded into three categories where one equaled poor, two equaled middle class, and three equaled rich. These scores were summed across the eight questions such that individuals with scores of 8-13 were considered lower income, individuals with scores of 19-24 were considered upper income and individuals with scores of 14-18 were middle income and were excluded from eligibility.

Obtaining target numbers of participants. There were only a few instances in which sessions proceeded with fewer than six individuals or with individuals with different demographic profiles than anticipated. This includes seven instances in which groups proceeded with five rather than six individuals, either because an insufficient number showed up or because a participant left before the session was concluded. This affected three same/same groups, 2 mixed sect/same class groups, 1

Screening Survey Questions Included in Economic Status Index			
Question No.	**Question Text**	**Answer Options**	**Scoring for Index (1-3)**
1	When you think of the total number of your household acquirings (houses, lands, cars, mobile phones, computers and laptops, household appliances, valuable furniture/decoration items, jewelry, etc.) what is, roughly, their estimated total value?	0 - 250,000 USD	1
		250,001 500,000 USD	2
		500,001+ USD	3
		Don't know/Refuse/NA	0
2	What is the estimated area of your main place of residency?	Less then 150m2	1
		150 to 250m2	2
		More then 250m2	3
		Don't know/Refuse/NA	0
3	Do you own a summer house? (Including chalets in seaside resorts)	No	1
		It happens that we rent a place for summer but not consistently	2
		Yes	3
		Don't know/Refuse/NA	0
4	When faced with power shortage, what alternatives do you resort to?	Nothing, we don't have money to buy power	1
		We buy power from a private generator 5 A	1
		We buy power from a private generator 10 A	2
		We buy power from a private generator 15 A+	3
		We own a private generator	3
		Don't know/Refuse/NA	0
5	In general, can you afford to travel on a leisure trip with your family at least once a year?	No we can't afford it	1
		Yes, but only to cheaper destinations, or on tour offers	2
		Yes, we can go wherever we want	3
		Don't know/Refuse/NA	0
6	In a typical month, how often can you afford to go with your family for lunch or dinner to restaurants (for bills totaling at least 100 USD)?	0	1
		1 to 2	2
		3+	3
		Don't know/Refuse/NA	0
7	What is your family's net monthly income? (Shown here in US Dollars but both options were provided in the original questionnaire.)	0	1
		1 - 120	1
		121 - 333	1
		334 - 667	1
		668 - 1,333	1
		1,334 - 2,667	1
		2,668 - 4,000	2
		4,001 - 5,333	2
		5,334 - 7,333	3
		7,334 - 9,333	3
		9,334 - 12,667	3
		12,668 - 16,667	3
		16,668 - 26,667	3
		26,668 - 53,333	3
		53,334 or more	3
		Don't know/Refuse	0
8	Which of the following is the best description of your family's monthly income?	The family income does not cover our needs and we face major problems making ends meet	1
		The family income barely covers our needs and we sometimes face problems making ends meet	1
		The family income covers our needs but we cannot afford luxury items or any extra leisure activities	2
		The family income covers our needs without us facing any major difficulties	3
		The family income very well covers our needs and we can also save some of it.	3
		Don't know/Refuse/NA	0
Creating the Index:			
The minimum score is 8 (1 point on each question above)		Score between 8 and 13 = lower income individuals	
The maximum score is 24 (3 points on each question above)		Score between 14 and 18 = middle class individuals disregarded	
		Score between 19 and 24 = upper middle class individuals	
Scores below 8 means that at least one question was not responded to.			
If more than two questions are not responded to, the screener is disregarded.			
If 1 or 2 questions are not responded to, the following scoring applies:			
		Score between 6 and 9 = lower income individuals	
		Score between 10 and 13 = middle class individuals disregarded	
		Score between 14 and 18 = upper middle class individuals	

Table A.3: Screening Survey Questions and Index Creation.

same sect/mixed class group, and 1 mixed/mixed group. The effects of the imbalance are plausibly the greatest for the groups that are not homogeneous. To address concerns, we control for the number of session participants in each group, described in Appendix F. We also checked to make sure that we did not accidentally have individuals who knew each other in the same session. While 41 individuals in 26 sessions reported that they knew at least one person in their session group

prior to the session, only 15 of those 41 were women participants. Upon further investigation with the session organizers, we learned that these were mostly cases in which individuals had been transported together or met casually just prior to the session. We nonetheless control for the total number of people in the session that each participant reported knowing beforehand (see Appendix F).

A.4 Possible concerns about selection into participation

The way in which individuals were scheduled to participate in the sessions gives some cause for concern that there was selection into participation in a way that could have introduced imbalances in pre-treatment characteristics for individuals in different arms. Recall that we recruited a total of 1200 individuals in order to obtain 720 participants and 480 back-ups. To obtain the 1200, we recruited 40 individuals of each of the six profile types (e.g. poor Sunni, rich Sunni, poor Shia, rich Shia, poor Christian, rich Christian) for each of the five sets. For each set we block randomized individuals by profile type with the the goal of obtaining 24 participants and 16 extras for each session. Panel A of Table A.4 shows how the 40 individuals of each profile type were assigned and Panel B of Table A.4 shows the target number of participants per treatment. The targets were set this way because we anticipated needing a different number of backups for each experimental condition.[3]

Table A.4: Illustration of potential selection into participation

		Mixed sectarian	
		N	*Y*
Mixed class	*N*	9	9
	Y	10	12

Panel A: Treatment assignment ($n = 40$)

		Mixed sectarian	
		N	*Y*
Mixed class	N	6	6
	Y	6	6

Panel B: Target participated ($n = 24$)

		Mixed sectarian	
		N	*Y*
Mixed class	*N*	4.5/9 = .50	4.5/9 = .50
	Y	5/10 = .50	6/12 = .50

Panel C: Proportion of those assigned who are 'very enthusiastic'

		Mixed sectarian	
		N	*Y*
Mixed class	*N*	4.5/6 = .75	4.5/6 = .75
	Y	5/6 = .83	6/6 = 1

Panel D: Proportion of those who participated who are 'very enthusiastic'

Implementing randomization in this way would still yield unbiased estimates of treatment effects as long as those who actually participated in the session were a random sample of the pool that was assigned. We worked with the implementing partner to design a procedure to try to ensure that this would be the case. First, the partner pre-screened all eligible participants for willingness to participate in a discussion on political and economic issues (without providing any information on the differing sectarian compositions of the groups). This resulted in a pool of potentia partici-

[3]Specifically, we planned to over-recruit by 50 percent. For example, for poor Sunnis in homogeneous groups there was one session and we needed six participants and 3 backups ($6\times1 + 3\times1 = 9$). The mixed sect/same class treatment required two poor Sunnis for three sessions and one backup for each session ($2\times3 + 1\times3 = 9$). The same sect/mixed class treatment required three poor Sunnis for two sessions and two backups for each session ($3\times2 + 2\times2 = 10$). And the fully mixed treatment required one poor Sunni for six sessions plus one backup for each session ($1\times6 + 1\times6 = 12$).

pants who were all willing to join in the activity. We asked our implementing partner to schedule the sessions such that every person in the pool would show up at one in accordance with their treatment assignment, ensuring that we always had more individuals than necessary of each profile type at each session. The implementing partner was then supposed to randomly select (for each profile type) who would actually stay to participate and who would be asked to go home (after receiving compensation) or invited to a different session. In actuality, however, the partner typically ended up getting only the target number of participants to show up for each session, which introduces the possibility that there was some differential selection into who ended up participating.

To see why this is an issue, assume that there is some (unobserved) variable like *enthusiasm* that affects willingness to participate. Assume also that treatment assignment achieved balance in this variable across the four experimental conditions. For illustrative purposes, we assume that 50 percent of all assigned individuals are *very* enthusiastic and the rest were only moderately enthusiastic. Panel C of Table A.4 shows the proportion of individuals assigned who were very enthusiastic and we can see that this is balanced across the four experimental conditions. Assume then that all very enthusiastic individuals were the easiest to schedule and were therefore more likely to participate (regardless of their treatment assignment, which they did not know before arrival). Panel D of Table A.4 shows how, if this were the case, the enthusiasm proportion would now be imbalanced across the treatment conditions among those who actually participated. We emphasize that this issue is not related to the treatment assignment itself but rather to the fact that we assigned a varying number of individuals in each experimental condition in order to reach our target of six participants of each profile.

One way to avoid this problem would have been to over-recruit even more individuals, for instance if we had a pool of 48 of each profile type rather than 40 (meaning that we would have had 12 people assigned to each experimental condition rather than the configuration shown in Panel A). This would have required the partner to over-recruit an additional 8 individuals from 6 profiles for each of 5 sets for a total of recruiting an additional 240 people. At the time of design our implementing partner strongly preferred the plan described above because they felt it would be more manageable and cost-effective that over-recruiting even more as they were already at the maximum of what they felt they could do.

So, how concerned should we be? After we discovered this, we discussed extensively with our partner and it seems that in most cases attendance was driven by idiosyncratic scheduling factors rather than systematic differences. Moreover, for this to be a problem, there would have to be not only non-trivial differential participation but also that this disparity would have to have non-trivial impacts on cooperation. While we think this unlikely, some might find this only somewhat reassuring. We are further reassured by the fact that the checks in Appendix B suggest balance on a large number of pre-treatment covariates between treatment and control. We include covariates in all analysis to address concerns.

A.5 Treatment Assignment Probabilities

Our main analysis employs inverse probability weights to correct for unequal treatment assignment probabilities. We use two different weights. As described above, we block randomly assigned participants based on profile and set using the same probabilities in each block ($\frac{9}{40}$, $\frac{9}{40}$, $\frac{10}{40}$, and $\frac{12}{40}$). In practice we stratified treatment assignment not only on set and profile type but, where possible, we created even smaller strata using additional information on recruiter and participant

neighborhood and randomly assigned individuals using proportional probability assignment within these small strata. We used these small strata to minimize the chances that discussion participants would know each other, which was more likely if they came from the same neighborhood and/or same recruiter network. In going from our pool of 40 of each type to our 24 participants, we lose observations in small strata cells, resulting in a large number of empty cells. Panel A of A.5 provides an illustration of this, showing the number of participants as assigned in small strata (left) and the number of participants that actually took part in the discussions (right).

We address this issue through post-stratification where we collapse the strata until we have no empty cells and then create new weights so that those who participated are weighted up to reflect the pool of potential participants originally assigned. We create two versions of weights based on two ways of collapsing the strata. First we created new 'smaller' strata where we collapsed cells such that we had no empty cells but where we retained information on recruiter or neighborhood were possible. Panel B in Table A.5 provides an example of how this was done. We then construct probability weights to weight individuals who participated up to reflect the 'population' as assigned. Second, we create 'bigger' strata where we collapse such that strata are formed by profile and set only, as in Panel C. We again create weights to weight those who participated up to the population of those assigned.

Our main analysis uses weighted least squared regression employing the weights created for the smaller strata. In Appendix G we check the robustness of results to several additional specifications, including estimates of treatment effects on the sample, estimates using the weights for bigger strata, and estimates with block fixed effects using smaller and bigger strata.

Table A.5: Example of post-stratification

Panel A: Example of treatment assignment and participation in small strata

	Assigned (n=40)		Participated (n=24)	
Small strata 1	1	1	1	0
	1	1	0	1
Small strata 2	1	1	0	0
	1	2	0	1
Small strata 3	1	1	1	1
	2	2	1	1
Small strata 4	2	2	2	2
	2	2	2	1
Small strata 5	2	2	0	2
	2	2	1	2
Small strata 6	2	2	2	1
	2	3	2	0

Panel B: Example of treatment assignment and participation in 'smaller' strata after collapsing strata

	Assigned (n=40)		Participated (n=24)	
New small strata	6	6	3	3
(collapsed 1, 2, 5, 6)	6	8	3	4
Small strata 3	1	1	1	1
	2	2	1	1
Small strata 4	2	2	2	2
	2	2	2	1

Panel C: Example of treatment assignment and participation in 'bigger' strata

	Assigned (n=40)		Participated (n=24)	
'Big' strata	9	9	6	6
	10	12	6	6

B Balance Checks

We use data from the screening survey and self-administered baseline survey (which participants completed after giving consent but before the session began) to check balance. The screening and pre-treatment surveys contain 50 variables that can be used to check balance. While we check balance using the individual covariates, we also use inverse covariance weighting to create pre-specified indices for measures that capture a common underlying concept (we use the indices as controls rather than the individual variables in the main analysis). Table B.1 presents results for the balance tests for both women and men for the individual covariates as well as for the indices.[4] We test for balance with a weighted least squares regression of the variable on a binary indicator for treatment status (mixed-class = 1) with robust standard errors.

The balance tests in Table B.1 show that five out of 50 covariates are significant at the 95 percent confidence level for women. While this is slightly higher than what we would expect by chance, these imbalances could be due to the relatively small sample. To address these imbalances we include all variables as controls in our main regressions. We also note that our main results for women are significant at the 99 percent confidence level and we have very few chance imbalances at that level of significance. With respect to men, only two of the 50 covariates are significant at the 95 percent confidence level, which is what we would expect by chance. Taken together, these balance tests suggest that the randomization procedure was largely effective in ensuring that characteristics are likely to be evenly distributed across our treatment (mixed-class) and control (same-class) groups.

[4] We note that if there is an imbalance in an index component there is likely to be an imbalance in the index itself.

	Mixed class (women)		Mixed class (men)	
	b	**p**	**b**	**p**
Panel A: Individual Variables				
Demographics				
Age	0.66	0.625	0.32	0.696
Marital status	0.01	0.842	0.05	0.296
Post-secondary education	0.01	0.857	-0.03	0.568
Christian	0.00	1.000	0.00	0.923
Sunni	0.00	0.957	0.00	1.000
Shia	0.00	0.956	0.00	0.922
Economic wealth index	-0.05	0.663	0.05	0.661
Assets (screening)	0.01	0.931	-0.04	0.577
HH area (screening)	-0.06	0.472	-0.02	0.737
Summer house (screening)	-0.02	0.859	0.05	0.602
Electricity (screening)	0.05	0.655	0.10	0.315
Vacation (screening)	-0.01	0.893	-0.05	0.586
Dineout (screening)	-0.04	0.687	0.03	0.652
Household income (screening)	-0.05	0.801	0.04	0.756
Income subjectie (screening)	-0.02	0.913	0.04	0.760
Household income (pre-treatment)	-0.26	0.241	0.10	0.611
Self-identified class (pre-treatment)	-0.02	0.827	0.07	0.357
Students (%)	0.01	0.839	-0.01	0.734
Homemaker (%)	0.06	0.313	0.01	0.180
Prejudice index	0.14	0.262	0.04	0.662
Marrying someone from a diff confession	-0.06	0.686	0.01	0.923
Diff confession as physician	0.10	0.235	0.09	0.160
Dif confession as neighbor	0.04	0.636	0.02	0.830
Discussion politics with diff confession	0.15	0.272	0.02	0.864
Discussing social or econ issues with diff confession	0.05	0.630	0.04	0.602
Supervised by diff confession	0.22	0.025	-0.02	0.811
Friends with diff confession	0.06	0.432	-0.01	0.836
Political action index	-0.15	0.184	0.14	0.196
Discuss issues	-0.13	0.032	0.04	0.369
Talked to party members	-0.07	0.083	0.02	0.627
Signed a petition	0.03	0.177	0.02	0.497
Attended protest	-0.01	0.862	0.07	0.208
Social homgeneity index	0.08	0.547	-0.10	0.273
Friends from same class	-0.08	0.488	-0.09	0.351
Friends from same sect	0.23	0.114	-0.07	0.456
How often do you discuss when disagree	-0.07	0.463	-0.06	0.495
Sectarian identity index	-0.29	0.016	0.01	0.920
Willing to change sect	-0.21	0.030	-0.11	0.213
Support sectarian political party	-0.17	0.007	0.01	0.778
Strong sectarian identity	0.10	0.688	0.27	0.208
Well connected to sectarian elite index	-0.12	0.306	0.06	0.552
Help from zaim	-0.15	0.161	0.15	0.163
Help from religious leader	-0.05	0.666	-0.04	0.740
Strength of age group identity	0.36	0.142	0.12	0.555
Strength of gender identity	0.05	0.854	0.19	0.357
Strength of class identity	0.11	0.648	0.01	0.976
Strength of occupational identity	0.21	0.401	0.11	0.545
Strength of Lebanese identity	0.02	0.947	-0.38	0.054
Panel B: Implementtion variables				
Moderator 1 (of 2)	0.04	0.456	0.06	0.219
Groups with six participants	0.04	0.122	0.06	0.018
Knew people in group	0.08	0.015	-0.03	0.618
Days until the municipal election	-4.59	0.108	-6.39	0.049
Mixed sect treatment assignment	0.00	1.000	0.00	1.000
Answered all practice problems correctly (%)	0.03	0.454	-0.01	0.872
Correctly answered amount earned from group pot	-0.01	0.519	-0.03	0.126
Corrently answered group pot share	0.01	0.851	0.01	0.621
Correctly answered total earned	0.02	0.598	0.01	0.647

Notes: *P-values are from a two-tailed test.*

Table B.1: Balance checks

C Comparison of Discussion Sample to Lebanese Population

To assess the comparability of our participants to the population of the Beirut/Mount Lebanon area and the entire country, we take advantage of the fact that we also conducted a nationally representative survey of Lebanese citizens in early 2016, just prior to implementing the public goods experiments. We can benchmark the characteristics of our sample against what we know about the population from this survey.

We first briefly summarize the survey methodology here. More detailed information is available from the authors upon request. The survey was conducted with 2,496 adult Lebanese citizens (18-65 years of age). Respondents were selected through multi-stage cluster sampling. Primary sampling units (PSUs) were villages in rural areas and cities or neighborhoods in urban areas. PSUs were randomly sampled—within strata defined by district, population size, and predominant sect—using simple random sampling. Households (and individuals within households) were randomly sampled within PSUs, with one respondent per household. To achieve a similar number of men and women in the sample, a target sex was set for each household. To draw population level inferences we employ a number of different design and post-survey weighting strategies. The analysis presented here uses entropy balancing as a re-weighting method as in Hainmueller (2012), although we get similar results if we use design weights, raked weights, or entropy balancing weights.

Appendix Tables C.1 and C.2 uses 25 comparable questions on both the nationally representative survey and the self-administered survey completed just before the public goods game was played to compare our women and men participants to the population in the Beirut/Mt Lebanon areas as well as nationwide. It shows, for instance, that women in our sample were more likely to be married than the average woman in the Beirut and Mount Lebanon area, less likely to have post-secondary education, and perhaps more likely to have socially homogeneous social networks. We also emphasize that one of the differences between the sample and the population is that the sample intentionally excludes individuals who are middle-class, which could explain differences on a number of other measures correlated with class.

While it is difficult to speculate on the extent to which the results for our sample would generalize to the population, we hope that this comparison provides a basis for readers who might have a particular interest in a particular characteristic. For instance, insofar as women in our sample have more homogeneous class networks than women in the Beirut/Mt Lebanon population (mean of 2.90 versus 2.48)—and the negative effects of mixed class discussion hold for women on average but are stronger for those with homogeneous networks (see Appendix H—this suggests that the average effects in the population might still be negative but smaller in magnitude than those found in the sample.

	Variable Range		Discussion sample			Beirut/Mt. Lebanon population			All Lebanon population		
	Min	Max	Mean	SD	N	Mean	SD	N	Mean	SD	N
Demographics											
Age	19	60	36	11	285	41	16	373	39	14	1230
Married	0	1	0.73	0.45	285	0.51	0.50	373	0.52	0.50	1230
At least secondary education	0	1	0.61	0.49	285	0.71	0.45	373	0.60	0.49	1230
Christian	0	1	0.33	0.47	285	0.57	0.49	373	0.44	0.50	1230
Sunni	0	1	0.33	0.47	285	0.09	0.29	373	0.24	0.43	1230
Shia	0	1	0.34	0.47	285	0.21	0.41	373	0.25	0.43	1230
Economic welfare											
Income (scale 1-15)	1	11	6.06	1.76	285	5.24	0.76	373	4.91	1.01	1230
Employed (at least part-time)	0	1	0.30	0.46	284	0.46	0.50	373	0.42	0.49	1230
Perceived economic class	1	5	3.06	0.73	278	2.47	0.65	373	2.44	0.62	1230
Subjective income	1	5	3.27	1.28	285	2.32	0.67	373	2.29	0.68	1230
Unemployment is top three concern	0	1	0.29	0.46	285	0.25	0.43	373	0.27	0.44	1230
Rising prices are a top three concern	0	1	0.25	0.44	285	0.47	0.50	373	0.50	0.50	1230
Clientelist connections											
Connected to Zaim	1	4	1.74	0.91	285	1.53	0.77	373	1.73	0.90	1230
Connected to religious leader	1	4	2.02	0.99	285	1.77	0.94	373	1.89	1.00	1230
Political action											
Talk to party members/MPs/Zaim	0	1	0.10	0.30	285	0.06	0.24	373	0.04	0.20	1230
Signed a petition	0	1	0.04	0.18	285	0.13	0.34	373	0.08	0.27	1230
Attended protest	0	1	0.27	0.44	285	0.27	0.44	373	0.15	0.36	1230
Comfortable [] a non co-sectarian											
Marrying	1	4	2.53	1.09	285	2.60	1.15	373	2.87	1.14	1230
Being neighbors with	1	4	1.52	0.72	285	1.47	0.75	373	1.64	0.85	1230
Being supervised by	1	4	1.65	0.78	285	1.41	0.72	373	1.68	0.91	1230
Being friends with	1	4	1.35	0.64	285	1.30	0.62	373	1.61	0.88	1230
Network Homogeneity											
Proportion of friends from a different sect	1	5	2.79	1.17	285	2.31	0.80	373	2.89	1.10	1230
Proportion of friends from a different class	1	5	2.90	0.98	285	2.48	0.70	373	2.93	1.02	1230
Discuss with those with whom you disagree	1	4	2.36	0.82	285	2.53	0.76	373	2.84	0.84	1230

Table C.1: Comparison of discussion participants to Lebanese population (women)

	Variable Range		Discussion sample			Beirut/Mt. Lebanon population			All Lebanon population		
	Min	Max	Mean	SD	N	Mean	SD	N	Mean	SD	N
Demographics											
Age	18	65	29	8	428	43	14	394	40	14	1265
Married	0	1	0.36	0.48	428	0.61	0.49	394	0.64	0.48	1265
At least secondary education	0	1	0.71	0.45	428	0.54	0.50	394	0.47	0.50	1265
Christian	0	1	0.33	0.47	428	0.42	0.49	394	0.34	0.47	1265
Sunni	0	1	0.34	0.47	428	0.20	0.40	394	0.32	0.47	1265
Shia	0	1	0.33	0.47	428	0.24	0.43	394	0.25	0.43	1265
Economic welfare											1265
Income (scale 1-15)	1	12	6.36	1.87	428	5.22	0.73	394	4.95	0.88	1265
Employed (at least part-time)	0	1	0.81	0.39	428	0.88	0.33	394	0.88	0.32	1265
Perceived economic class	1	5	3.04	0.78	412	2.46	0.59	394	2.44	0.58	1265
Subjective income	1	5	3.31	1.34	428	2.33	0.62	394	2.29	0.67	1265
Unemployment is top three concern	0	1	0.29	0.45	428	0.36	0.48	394	0.34	0.47	1265
Rising prices are a top three concern	0	1	0.25	0.44	428	0.44	0.50	394	0.45	0.50	1265
Clientelist connections											1265
Connected to Zaim	1	4	2.02	1.04	428	1.55	0.83	394	1.72	0.94	1265
Connected to religious leader	1	4	2.19	1.00	428	1.62	0.82	394	1.85	0.99	1265
Political action											1265
Talk to party members/MPs/Zaim	0	1	0.13	0.34	428	0.19	0.40	394	0.14	0.34	1265
Signed a petition	0	1	0.08	0.28	428	0.10	0.31	394	0.09	0.28	1265
Attended protest	0	1	0.41	0.49	428	0.37	0.48	394	0.23	0.42	1265
Comfortable [] a non co-sectarian											1265
Marrying	1	4	2.08	1.00	428	2.35	1.18	394	2.64	1.21	1265
Being neighbors with	1	4	1.52	0.74	428	1.27	0.54	394	1.51	0.76	1265
Being supervised by	1	4	1.68	0.87	428	1.27	0.53	394	1.60	0.85	1265
Being friends with	1	4	1.34	0.64	428	1.15	0.39	394	1.49	0.80	1265
Network Homogeneity											1265
Proportion of friends from a different sect	1	5	2.58	1.02	428	2.05	0.70	394	2.70	1.12	1265
Proportion of friends from a different class	1	5	2.81	0.97	428	2.19	0.71	394	2.70	1.03	1265
Discuss with those with whom you disagree	1	4	2.31	0.79	428	1.94	0.71	394	2.48	0.95	1265

Table C.2: Comparison of discussion participants to Lebanese population (men)

D Public Goods Game Implementation

This appendix details the implementation of the public goods game. A detailed field manual with all instructions for the moderator teams and with the forms used to ensure comprehension and record contributions is available from the authors upon request.

D.1 Introducing the public goods game

Upon arrival at the study site, participants were asked to provide informed consent and fill out the self-administered pre-survey questionnaire. After filling out the survey, participants were invited to sit together at a table where everyone could see one another, as well as the trained session moderator. To ensure that participants were aware of their group composition before playing the public goods game, the moderator provided this information during her introductory remarks using the following script, which was primarily used to introduce the discussion activity that would follow the baseline data collection, of which the public goods experiment was a part:

> *We are meeting today to discuss the recent developments in the country, mainly the protests that recently began in Lebanon. Many persons consider that these protests may present an important moment to reflect about the future of this country regardless of their outcome.*
>
> *We have invited you here today to engage in a discussion with members from [SAME/ DIFFERENT] sectarian groups and [SAME/DIFFERENT] economic classes so that you can share with each other your thoughts and feelings about your economic and political hopes and concerns. Some of what we discuss today could be sensitive and at times people might disagree—that is ok. We just ask that you engage with one another with honesty and respect so that we can all learn more about how people who we do not know personally are thinking and feeling on the issues that we all face.*

Participants were then asked to introduce themselves and offer basic personal information (e.g. on their jobs or neighborhoods) that would confirm their profiles to all other members of the group. After the session was introduced, a *different* member of the moderation team administered the public goods game. This was done to mitigate any potential social desirability bias that might arise from trying to 'impress' the moderator before engaging in the discussion.

D.2 Payoff structure of the game

Participants played with 10,000 Lebanese pounds (LBP) that they earned for completing a pre-survey upon arrival at the site. Participants were allowed to contribute any amount in 1,000 LBP increments to the group pot. To indicate their choice, participants circled a contribution amount on a slip of paper,[5] inserted the paper into an envelope labeled with their participant identification number, and then passed the envelope back to the assistant moderator. Payoffs were determined as follows: the total amount contributed to the group pot was multiplied by 1.5 and divided evenly among all six participants, regardless of whether they contributed or not. Thus, the payoff function for each subject i was:

[5] All participants had the option to contribute from 0 to 10,000 LBP in increments of 1,000.

$$\pi_i = 10,000 - c_i + 0.25 \cdot \sum_{j=1}^{6} c_j \tag{1}$$

where c_i is the contribution to the public good (group pot) of subject i, in any group whose 6 members are indexed by j. The marginal per capita return (MPCR) from the public good was 0.25 (1.5 times total contributions divided by 6). Participants were not informed of the final results of the game until just prior to exiting the facility, after completing a post-survey questionnaire and signing receipts for payment.

A total of 713 subjects participated in the experiment.[6] The average amount earned in the public goods game was $7.85 USD.[7] The maximum amount earned in the public goods game was $14.00 USD while the minimum amount earned in the public goods game was $2.50 USD.[8] For women specifically, the average total payoff from the public goods game was $7.78 USD, with a minimum earned of $2.67 USD and a maximum of $12.17 USD.[9]

Prior to playing the game, the public goods game moderator completed exercises with each participant to ensure their comprehension of the payoff structure. For an example of the practice problems used, see Figure D.1. The main variable equals 1 if the participants correctly answered all three questions on the example exercise worksheet provided prior to playing Round 1 of the game. As can be seen in Table D.1, comprehension scores for both women and men were high—88 percent of women answered all three questions correctly as did 82 percent of men.

	Women (n=285)				**Men (n=428)**			
	Min	*Max*	*Mean*	*SD*	*Min*	*Max*	*Mean*	*SD*
Comprehension variables								
Answered all practice problems correctly (%)	0	1	0.88	0.33	0	1	0.82	0.39
Correctly answered amount earned from group pot	0	1	0.98	0.14	0	1	0.97	0.18
Corrently answered group pot share	0	1	0.93	0.25	0	1	0.90	0.30
Correctly answered total earned	0	1	0.94	0.25	0	1	0.91	0.29

Table D.1: Summary Statistics for Control Variables

D.3 Mitigating demand effects

We took steps to ensure that the study results are not an artifact of social desirability bias, experimenter effects, or moderator effects. All 120 sessions were introduced by one of two moderators. Moreover, as mentioned above, the public goods experiment itself was administered by an assistant moderator not involved in the main discussion activity that followed the baseline data collection

[6] The average group size of six participants in our study is in line with standard public goods game designs where groups typically include 3-6 participants (Kurzban, Burton-Chellew and West, 2015, 585).

[7] For reference, the hourly minimum wage in Lebanon is about $3.78 USD.

[8] In Lebanese currency: the average amount earned in the public goods game was 11,769 LBP, with a minimum of 3,750 LBP and a maximum of 21,000 LBP earned for the full sample of participants. The amount earned in the game was combined with a $20 USD show-up fee for participation in all activities involved in the experiment, including the discussion portion not analyzed here, to yield each individual's total compensation for participation in the approximately 90-minute study.

[9] For men, the average total payoff from the game was about $7.89 USD, with a minimum of $2.50 USD and a maximum of $14.00 USD.

Examples Worksheet
Group ID:
Participant ID:

Example 1:

Contributors	To Group Pot	Keep privately	Earned from Group Pot	Total Earned
Participant 1	8,000	2,000	4,000	6,000
Participant 2	0	10,000	4,000	14,000
Participant 3	0	10,000	4,000	14,000
Participant 4	2,000	8,000	4,000	12,000
Participant 5	2,000	8,000	4,000	12,000
Participant 6	4,000	6,000	4,000	10,000

Total in group pot	16,000
Multiplied by 1.5	24,000
Each person's share from Group Pot	24,000/6= 4,000

Example 2:

Contributors	To Group Pot	Keep privately	Earned from Group Pot	Total Earned
Participant 1	1,000	9,000	12,000	21,000
Participant 2	9,000	1,000	12,000	13,000
Participant 3	9,000	1,000	12,000	13,000
Participant 4	9,000	1,000	12,000	13,000
Participant 5	10,000	0	12,000	12,000
Participant 6	10,000	0	12,000	12,000

Total in group pot	48,000
Multiplied by 1.5	72,000
Each person's share from Group Pot	72,000/6= 12,000

Practice Problem

Contributors	To Group Pot	Keep privately	Earned from Group Pot	Total Earned
Participant 1	6,000	4,000		
Participant 2	5,000	5,000		
Participant 3	5,000	5,000		
Participant 4	7,000	3,000		
Participant 5	7,000	3,000		
Participant 6	10,000	0		

Total in group pot	40,000
Multiplied by 1.5	60,000
Each person's share from Group Pot?	60,000/6 =

Figure D.1: **Comprehension exercises**

to mitigate the possibility that participants behaved in a way to ensure that participants did not behave in a way to try to curry favor with the main moderator in the subsequent discussion. To control for any differences in abilities among moderators to introduce the sessions, we take advantage of the fact that both moderators led both same- and mixed-class groups to run regressions with moderator fixed effects. We also guarded against moderator fatigue by making sure that treatment and control groups were organized in a roughly alternating fashion. Additionally, to mitigate social desirability bias, all participants also recorded their contributions in private. Finally, we note that neither the moderators nor the subjects knew the hypotheses of the study in advance—indeed these results are based on exploratory analysis for which we had no *a priori* expectations.

E Data and Estimation

Our main analysis involves a weighted least squares repression of the following form:

$$Y_{ij} = \alpha + \beta T + X'_{ij}\gamma + \epsilon_i \tag{2}$$

where Y_i is the contribution to the group pot of individual i; T is a treatment indicator for whether a participant was in a mixed-class (versus same-class) group; X_{ij} is a vector of individual and group j controls (see Appendix F); and ϵ_i is the individual-level random error term. We note that we estimate results separately for men and women as well as by economic class.

E.1 Relation to the pre-analysis plan

As noted in Appendix A.1, the public goods experiment was part of the baseline for a separate study on the effects of discussion in homogeneous and heterogeneous class and sectarian groups on political preferences. While the specific tests performed in this paper were not pre-registered, we did follow the pre-analysis plan registered with the Evidence in Governance and Politics (EGAP) network with respect to data preparation and estimation. Specifically, we pre-specified that we would perform 10 rounds of missing data imputation using multivariate imputation via chained equations. We also pre-specified that we would use inverse covariance weighting to create indices using variables that measure the same construct, which we use to create control indices (see Appendix F), and that we would estimate results with and without controls. As can be seen in the main text and Appendix G, results are not sensitive to the inclusion or exclusion of control variables in the estimation or to the specific weights we use in the analysis.

F Control Variables

We estimate all results in the main text and in Appendix G with and without control variables. The control variables come from both the screening survey and a self-administered survey that was completed before the start of the public goods experiment. We include these measures to improve precision, check for chance imbalances (see Appendix B), and account for the fact that neither gender nor class are randomly assigned and as such there could be a number of potentially confounding factors.

Table F.1 presents summary statistics for all control indices and their components. Control indices were pre-registered. Specifically, we include two sets of controls. First, we include *individual* controls for factors like demographics, political engagement, sectarian prejudice, and network homogeneity (Panel A). We note that some of these pre-treatment covariates have been shown to affect cooperation in public goods games specifically, including (1) level of education (Candelo, Croson and Li, 2017), (2) university student status (Gächter, Herrmann and Thöni, 2004), (3) age (Martinsson, Villegas-Palacio and Wollbrant, 2015), (4) marital status (Tognetti et al., 2016), and (6) domestic laborer status (Carpenter, Daniere and Takahashi, 2004). Including these and other covariates collected through our pre-treatment survey instruments allows us to have greater confidence that the effects are due to heterogeneous group treatment exposure and not other non-random variation that exists between men and women in our sample.

Second, we include *implementation* controls that account for things related to the specific session or group (Panel B). These include moderator fixed effects (to control for differences in ability); whether the session had six participants; whether any participants new each other in advance; the number of days to the upcoming municipal elections (to control for timing effects); and whether the group was also randomly assigned to be a homogeneous or heterogeneous sectarian group (see Appendix A).

	Women (n=285)				Men (n=428)			
	Min	*Max*	*Mean*	*SD*	*Min*	*Max*	*Mean*	*SD*
Panel A: Individual Variables								
Demographics								
Age	19	60	36.07	11.17	18	65	28.56	8.36
Marital status	0	1	0.72	0.45	0	1	0.36	0.48
Post-secondary education	0	1	0.61	0.49	0	1	0.71	0.45
Christian	0	1	0.33	0.47	0	1	0.34	0.47
Sunni	0	1	0.33	0.47	0	1	0.33	0.47
Shia	0	1	0.33	0.47	0	1	0.33	0.47
Economic wealth index	-2.04	2.27	-0.09	0.90	-2.22	2.30	0.05	1.05
Assets (screening)	1	3	1.61	0.71	1	3	1.65	0.79
HH area (screening)	1	3	1.72	0.68	1	3	1.71	0.75
Summer house (screening)	1	3	2.03	0.98	1	3	2.04	0.97
Electricity (screening)	1	5	2.49	0.88	1	5	2.60	0.99
Vacation (screening)	1	3	1.83	0.84	1	3	1.91	0.85
Dineout (screening)	1	3	2.06	0.72	1	3	2.14	0.74
Household income (screening)	1	10	6.63	1.59	1	10	6.80	1.44
Income subjectie (screening)	1	5	3.27	1.28	1	5	3.31	1.34
Household income (pre-treatment)	1	11	6.07	1.77	1	12	6.37	1.85
Self-identified class (pre-treatment)	0	2	1.04	0.65	0	2	1.02	0.73
Students (%)	0	1	0.11	0.31	0	1	0.15	0.36
Homemaker (%)	0	1	0.57	0.50	0	1	0.00	0.06
Prejudice index	-1.26	4.16	0.15	0.98	-1.26	4.16	-0.11	1.01
Marrying someone from a diff confession	1	4	2.52	1.09	1	4	2.08	0.98
Diff confession as physician	1	4	1.36	0.66	1	4	1.32	0.60
Dif confession as neighbor	1	4	1.52	0.72	1	4	1.52	0.76
Discussion politics with diff confession	1	4	2.18	1.11	1	4	1.93	1.03
Discussing social or econ issues with diff confession	1	4	1.70	0.86	1	4	1.57	0.80
Supervised by diff confession	1	4	1.64	0.78	1	4	1.68	0.87
Friends with diff confession	1	4	1.35	0.65	1	4	1.35	0.64
Political action index	-1.20	3.25	-0.17	0.94	-1.20	3.25	0.13	1.03
Discuss issues	0	1	0.66	0.47	0	1	0.74	0.44
Talked to party members	0	1	0.10	0.30	0	1	0.13	0.34
Signed a petition	0	1	0.04	0.19	0	1	0.08	0.27
Attended protest	0	1	0.27	0.44	0	1	0.43	0.50
Social homgeneity index	-2.18	2.78	0.13	1.06	-2.18	2.78	-0.09	0.95
Friends from same class	1	5	2.92	0.98	1	5	2.81	0.97
Friends from same sect	1	5	2.81	1.18	1	5	2.56	1.01
How often do you discuss when disagree	1	4	2.34	0.82	1	4	2.30	0.80
Sectarian identity index	-2.74	1.54	0.09	0.97	-2.74	1.54	-0.07	1.02
Willing to change sect	1	4	3.52	0.78	1	4	3.34	0.86
Support sectarian political party	0	1	0.48	0.50	0	1	0.47	0.50
Strong sectarian identity	1	7	4.41	2.01	1	7	4.22	2.13
Well connected to sectarian elite index	-1.15	2.25	-0.19	0.92	-1.15	2.25	0.11	1.03
Help from zaim	1	4	1.69	0.87	1	4	2.00	1.03
Help from religious leader	1	4	1.98	0.97	1	4	2.20	1.01
Strength of different identities								
Age group	1	7	4.49	1.94	1	7	4.57	1.85
Gender	1	7	5.38	1.87	1	7	4.88	1.83
Class	1	7	3.78	1.84	1	7	3.93	1.83
Occupation	1	7	4.26	1.91	1	7	4.88	1.77
Lebanese	1	7	5.48	1.88	1	7	5.68	1.78
Panel B: Implementtion variables								
Moderator 1 (of 2)	0	1	0.40	0.49	0	1	0.40	0.49
Groups with six participants	0	1	0.95	0.23	0	1	0.95	0.22
Knew people in group	0	3	0.06	0.29	0	5	0.11	0.52
Days until the municipal election	33	94	61.07	23.57	17	109	61.15	32.21
Mixed sect treatment assignment	0	1	0.50	0.50	0	1	0.50	0.50

Table F.1: Summary Statistics for Control Variables

G Robustness Checks

This appendix presents robustness checks for the main results. Columns three and four in Table G correspond to the results presented in the main table. The other columns show robustness of those results to the inclusion/exclusion of control variables, to the use of alternative weights to account for unequal treatment assignment probabilities (corresponding to weights used for the 'smaller' and 'bigger' strata described in Appendix A), and to the use of block fixed effects instead of IPW for the 'smaller' strata.[10] Table G shows that all main results are robust across specifications.

	Effect of mixed (versus same) class group							
	Sample		Smaller Strata				Bigger Strata	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Results for Women								
All women	-1330	-1332	-1374	-1383	-1285	-1548	-1310	-1327
	(362)	(376)	(367)	(390)	(357)	(441)	(363)	(375)
	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
N	285	285	285	285	285	285	285	285
Panel B: Results for Men								
All men	738	767	912	946	743	774	802	835
	(327)	(337)	(339)	(350)	(327)	(407)	(329)	(339)
	0.025	0.023	0.007	0.007	0.024	0.058	0.015	0.014
N	428	428	428	428	428	428	428	428
Controls	No	Yes	No	Yes	No	Yes	No	Yes
IPW	No	No	Yes	Yes	No	No	Yes	Yes
Fixed Effects	No	No	No	No	Yes	Yes	No	No

Tables present coefficients, standard errors, and p-values from two-tailed tests for separate regressions for men and women.

Table G.1: Robustness Checks

[10] For bigger strata probabilities or assignment are equal across all blocks.

H Mechanism Analysis

This appendix provides the supporting analysis for the discussion of possible explanations for the main findings. Specifically, we use additional data from the self-administered survey implemented before the public goods game experiment to assess three potential explanations for why women give less (and men give more) in mixed-class groups. Specifically, we consider whether women in mixed-class settings were more uncertain about how other women would behave in such settings; were more prone to in-class bias related to status-seeking; or were more distrustful due to competing conceptualizations of gender identity.

H.1 Uncertainty or class bias

As described in the main text, to evaluate the role of both uncertainty and class-bias, we use a question from a self-administered survey that captures the extent to which participants' real-world social networks are heterogeneous or homogeneous in their economic class. A homogeneous network could proxy for uncertainty *or* in-class bias insofar as those with less diverse networks might have fewer opportunities for cross-class interactions or be more biased against individuals from other classes (and thus select into more homogeneous networks). The survey question asked "Of your circle of family, friends, and acquaintances, how many of them belong to a different economic class from your own?" and we recode responses such that most/almost all=0, some=1, and a few/almost none=2.

Table H.1 presents results for women and men as well as disaggregated by class. We note the following patterns:

- Women on average, regardless of their social networks, are less likely to cooperate in mixed versus same-class groups.
- The negative effect on cooperation among women is biggest for those with homogeneous class networks, consistent with both the uncertainty and class-bias explanations.
- Men on average, regardless of their social networks, are more likely to cooperative in mixed versus same-class groups.
- Mixed-class groups have a bigger positive effect for poor men with homogeneous social networks, consistent with more of a status-seeking story.
- Mixed-class groups have a bigger positive effect for rich men with heterogeneous social networks, consistent with rich men being less uncertain about contributions in mixed settings or being more likely to seek status through 'competitive altruism' (as discussed in the main text).

Table H.1: Heterogeneous Effects of Mixed-Class Treatment by Type of Social Network

	All		Poor Only		Rich Only	
	Same class *mean*	Mixed class *b/se/p*	Same class *mean*	Mixed class *b/se/p*	Same class *mean*	Mixed class *b/se/p*
Panel A: Results for Women						
Almost all/most network from other class	4277	-1077	3966	-436	4662	-1826
		(687)		(916)		(1047)
		0.118		0.635		0.084
Some network from other class	3503	-781	3317	-378	3648	-1139
		(540)		(693)		(802)
		0.149		0.586		0.158
A few/almost none in network from other class	5014	-2716	4340	-1545	5758	-3937
		(742)		(1063)		(1021)
		0.000		0.148		0.000
Panel B: Results for Men						
Almost all/most friends from other class	3341	733	3645	166	3190	1095
		(594)		(1006)		(758)
		0.218		0.869		0.150
Some network from other class	3103	886	2635	686	3686	950
		(520)		(696)		(761)
		0.089		0.325		0.213
A few/almost none in network from other class	2766	1208	2590	1844	3091	80
		(710)		(906)		(1136)
		0.090		0.043		0.944

Robust standard errors in parentheses. P-values are from two-sided tests. All models incorporate weights that correct for unequal treatment assignment probabilities across strata and the full set of control variables.

H.2 Lack of a common gender identity

To investigate whether different views of gender identity might be driving mistrust among women in mixed-class groups, we use survey measures that capture the strength (salience) of gender identity as well as support for civil marriage. To examine strength of gender identity, we use a question from the self-administered survey that asked respondents to rank from a list the identities that they feel the strongest to weakest attachment to (with options for write-ins). Table H.2 shows the mean number of participants who view gender as their most important identity (column 1) or one of their top three identities (column 2). The results show that women generally feel more strongly about their gender identity than men and that poor and rich women tend to feel about equally strongly about their gender identity. This last observation is confirmed in the first two columns of Table H.3.

Table H.2: Summary Statistics on Strength of Gender Identity

	Gender is most important ID	Gender is top three ID
	mean/(s.e.)	*mean/(s.e.)*
Panel A: Women		
All	0.42	0.71
	(0.03)	(0.03)
Poor	0.43	0.76
	(0.04)	(0.04)
Rich	0.40	0.66
	(0.04)	(0.04)
Panel B: Men		
All	0.29	0.61
	(0.02)	(0.03)
Poor	0.29	0.58
	(0.03)	(0.04)
Rich	0.29	0.65
	(0.03)	(0.04)

Estimates incorporate weights that correct for unequal treatment assignment probabilities across strata.

The third and fourth columns in Table H.3 examine the association between class and support for civil marriage for both women and men. Civil marriage is widely viewed as a policy that would advance women's rights and protections under the law and as such has been supported by the women's movement in Lebanon (Zuhur, 2002). Under current law, women's access to economic, political, and social equality under the law varies depending on which of the various religious sects they belong to, since this determines which of the many religious courts they must enter for disputes related to marriage or family law. There is far more variation in the regulations across religious courts for women compared to men, which makes them doubly-disadvantaged by the current arrangement.

Our measure for support for civil marriage comes from the post-treatment survey used to collect outcome data for the main study described in Appendix A.1. As such, it reflects the effects of being assigned to a same- or mixed-class and same- or mixed-sect discussion (which we control for in our analysis). However, this is the only data from this study available to us that allows us to examine

whether support for civil marriage varies by class, consistent with the notion that there could be class differences among women in gender identity, what it means to be a woman, and what the role of women in society should be. The regression results presented in Table H.3 show, however, that there is no association between wealth and support for civil marriage for women. (Interestingly, upper class men are, however, significantly more likely to support civil marriage than lower class men). All in all, while we cannot rule out the possibility that class differences in other aspects of gender identity undermined women's cooperation, our analysis does not provide direct support for this mechanism.

Table H.3: Association between Class and Support for Civil Marriage by Gender

	Strength of Gender ID		Support for Civil Marriage	
	Women	Men	Women	Men
	b/se	*b/se*	*b/se*	*b/se*
Rich	0.10	0.03	0.00	0.15**
	(0.09)	(0.06)	(0.09)	(0.07)
Age	0.00	0.00	-0.01**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Married	-0.10	-0.07	-0.08	-0.01
	(0.10)	(0.05)	(0.09)	(0.06)
Post-secondary education	-0.12	-0.14*	-0.03	-0.01
	(0.10)	(0.07)	(0.10)	(0.07)
Shia	-0.15*	0.08	-0.24***	-0.18***
	(0.08)	(0.06)	(0.07)	(0.06)
Sunni	0.02	0.08	-0.34***	-0.23***
	(0.08)	(0.05)	(0.07)	(0.06)
Student	0.15	0.05	-0.26**	0.05
	(0.14)	(0.07)	(0.12)	(0.07)
Homemaker	0.03	0.07	-0.04	0.32*
	(0.08)	(0.12)	(0.08)	(0.18)
Sectarian prejudice index	-0.03	0.03	-0.14***	-0.03
	(0.04)	(0.03)	(0.03)	(0.03)
Political activity index	-0.06	0.00	0.03	0.03
	(0.04)	(0.03)	(0.03)	(0.02)
Homogeneity of social networks index	0.02	0.00	0.03	-0.04
	(0.03)	(0.02)	(0.03)	(0.03)
How often disagree	-0.02	0.02	0.05	0.00
	(0.04)	(0.03)	(0.04)	(0.03)
Sectarian ID index	-0.01	0.03	-0.08**	-0.09***
	(0.04)	(0.02)	(0.03)	(0.03)
Connectedness to sectarian leaders	0.03	0.03	-0.04	-0.03
	(0.04)	(0.03)	(0.04)	(0.03)
Strength of age ID	0.07***	0.08***	-0.01	0.00
	(0.02)	(0.01)	(0.02)	(0.01)
Strength of gender ID	–	–	0.03*	0.00
	–	–	(0.02)	(0.01)
Strength of econ ID	-0.01	0.01	0.01	-0.01
	(0.02)	(0.01)	(0.02)	(0.01)
Strength of Lebanese ID	-0.01	0.01	-0.03**	-0.01
	(0.02)	(0.01)	(0.02)	(0.01)
Strength of occupational ID	-0.02	0.02	0.00	0.03**
	(0.02)	(0.01)	(0.02)	(0.01)
Mixed-class treatment	0.01	0.08	-0.01	-0.04
	(0.06)	(0.05)	(0.06)	(0.05)
Mixed-sect treatment	-0.08	0.04	0.16***	0.01
	(0.07)	(0.04)	(0.06)	(0.05)
Constant	0.40	-0.45**	0.86***	0.37*
	(0.27)	(0.20)	(0.23)	(0.20)
N	285	428	285	428

Robust standard errors in parentheses. P-values are from two-sided tests. All models incorporate weights that correct for unequal treatment assignment probabilities across strata and the full set of control variables.

I Class versus Sectarian Differences

In this appendix we consider why class differences undermine cooperation more than sectarian differences for women. It is widely believed that ethnic or sectarian differences weaken cooperation in ethnically divided societies (Habyarimana et al., 2009), although evidence is mixed (Berge et al., 2016; Greig and Bohnet, 2009) and few studies examine gender differences in sectarian cooperation. One possible explanation for our results is that sectarian differences are less relevant for women than for men because of the sectarian and gendered way in which competition over resources takes place in Lebanon. Research suggests that, in societies where resources are distributed along ethnic lines through clientelistic networks, women are more likely than men to be denied direct access to these benefits (Benstead, 2016; Beall, 2005; Wantchekon, 2003). If sectarian identity is the foundation of competition over resources for men, and has little impact on access to goods and services for women, then sectarian differences might undermine cooperation among men more than among women. Moreover, such inter-group competition over resources could promote stronger norms of intra-group cooperation among men (Raihani and Bshary, 2015; Van Vugt, Cremer and Janssen, 2007).

To investigate this, we look at the impact of sectarian differences on cooperation among men, first overall and then conditional on the sectarian composition of the group. Table I.1 presents results for men corresponding to those for women presented in Table 2 in the main text. The negative coefficient suggests that men cooperate less in mixed-sect than in same-sect groups, although the difference is not statistically significant at conventional levels.

Table I.1: Contributions in Mixed-Sectarian Groups (Men only)

	Same sect	**Mixed sect**		
	mean	*b/(se)/p*		*N*
		Model 1	Model 2	
All men	3762	-398	-455	428
		(343)	(344)	
		0.247	0.187	

Robust standard errors in parentheses. P-values are from two-sided tests. All models incorporate weights that correct for unequal treatment assignment probabilities across strata. Model 1 has no controls; Model 2 includes all controls.

We dig deeper by taking advantage of the study's 2x2 factorial design in which we experimentally varied both the class and sectarian compositions of the groups (see Appendix A). This enables us to shed more light on how the sectarian composition of the group conditions cross-class cooperation. Table I.2 presents mean contribution levels separately for women and men in each of the four experimental arms: (1) same-sect/same-class, (2) mixed-sect/same-class, (3) same-sect/mixed-class, (4) and mixed-sect/mixed-class. We also estimate conditional average treatment effects.[11]

[11] We estimate effects by running a regression of the form: $Y_{ij} = \alpha + \beta_1 MS_i + \beta_2 MC_i + \beta_3 MS * MC_i + \epsilon_{ij}$ where Y_{ij} is the contribution by individual i in group j. MS is an indicator for whether a participant was assigned to a mixed-sect group and MC for whether they were assigned to a mixed-class group. Thus, β_1 captures the effect of being in a mixed-sect/same-class group (relative to a same-sect/same-class group); β_2 is the effect of being in a mixed-class/same-sect group; and β_3 is the interaction. We complete Table I.2 by calculating the remaining marginal effects. Standard errors are not clustered because treatment assignment was at the individual level (Abadie et al., 2017).

Table I.2: Contributions by Class and Sectarian Composition

Panel A: Women					Panel B: Men				
		Sect Composition					Sect Composition		
		Same	Mixed	**Diff**			Same	Mixed	**Diff**
Class	Same	3941	4364	422	Class	Same	2941	3273	332
	Mixed	3000	2556	-444		Mixed	4582	3455	-1127**
	Diff	-941*	-1807***	-866		**Diff**	1641***	183	-1459**

Notes: *Table shows mean contributions in each of the four experimental arms as well as tests of the differences between arms. Regressions to test differences incorporate weights that correct for unequal treatment assignment probabilities across strata and do not include controls.* $*p < .10, **p < .05, ***p < .01$.

The results for men in Panel B yield two notable findings. First, sectarian differences do not undermine cooperation *even among men*. Contribution levels in same-sect/same-class groups are similar to those in both mixed-sect/same-class and mixed-sect/mixed-class groups. Second, and even more striking, men cooperate significantly more in same-sect/mixed-class groups. Specifically, when men play with cosectarians, they contribute 1641 more in mixed-class groups than in same-class groups; there is no equivalent effect when men play with non-cosectarians. These results provide clear evidence that class differences strengthen cooperation among men but only among cosectarians. They also reveal that the negative coefficients Appendix Table I.1 arise not because sectarian differences undermine cooperation among men but rather because cross-class cooperation among cosectarians strengthens cooperation.

The results in Panel A show a noticeable difference for women. Class differences undermine cooperation for women regardless of the sectarian composition of the groups. Being in a mixed-class group reduces cooperation by 941 LBP among cosectarian women and by 1807 LBP among non-cosectarian women. While the negative (albeit statistically insignificant) coefficient on the interaction implies that class differences might reduce cooperation among non-cosectarians more than cosectarians, the results clearly show that class differences harm collective action capacity among women in general. Additional analysis in Table I.3 indicates that these results are driven almost entirely by the behavior of rich women, perhaps driven by the kinds of status-seeking behavior and class-bias described in the main text.[12]

Overall, these results support two main takeaways. First, sectarianism in Lebanon does not serve to undermine cooperation across sectarian lines either for men or for women. While a surprising finding, there are several possible explanations. It could be that ordinary Lebanese are simply less divided along sectarian lines than many believe. A similar lack of coethnic bias has been found in public goods games played in Kenya, another context where ethnic divisions are thought to be highly salient Berge et al. (2016). It is also possible that sectarian differences only undermine cooperation among certain subgroups in the population—for instance those with stronger identity attachments or access to coethnic elites (Marshall, 2019)—or only when ethnic identity is primed by political elites, for instance during periods of electoral competition (Posner, 2017, 2004).

[12]Note that the negative effect of mixed-class among poor women in heterogeneous sectarian groups is largely a result of their high levels of cooperation in same-sect, mixed-class groups.

Table I.3: Contributions by Class and Sectarian Composition

Panel A: Poor Women					**Panel B: Rich Women**				
		Sect Composition					Sect Composition		
		Same	Mixed	***Diff***			Same	Mixed	***Diff***
	Same	3184	4476	1292*		Same	4699	4252	-447
Class	Mixed	3325	2898	-427	Class	Mixed	2676	2204	-472
	Diff	141	-1577**	-1719*		***Diff***	-2023**	-2049***	-25

Panel C: Poor Men					**Panel D: Rich Men**				
		Sect Composition					Sect Composition		
		Same	Mixed	***Diff***			Same	Mixed	***Diff***
	Same	2358	3353	996		Same	3525	3193	-332
Class	Mixed	4249	3266	-983	Class	Mixed	4916	3645	-1271*
	Diff	1892***	-87	-1979**		***Diff***	1391**	453	-938

Notes: *Table shows mean contributions in each of the four experimental arms as well as tests of the differences between arms. Regressions to test differences incorporate weights that correct for unequal treatment assignment probabilities across strata and do not include controls.* $*p < .10, **p < .05, ***p < .01$.

The second takeaway is that sectarian differences in Lebanon could generate very different pressures for cross-class cooperation for men and for women. The results for men suggest that the chief impact of sectarian differences in Lebanon is not more out-group antagonism but rather more in-group cooperation among men who differ on other social dimensions. Critically, because women are generally excluded from competition over resources, they do not experience similar pressures for in-group solidarity. In this way, sectarian differences might strengthen cross-class ties among cosectarian men and fail to do the same for women. This indicates an important possible relationship between sectarian and class differences in their effect on cooperation that merits further investigation in future research.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" Working paper no. 17-030.
URL: *https://arxiv.org/pdf/1710.02926.pdf*

Beall, Jo. 2005. "Decentralizing Government and Decentering Gender: Lessons from Local Government Reform in South Africa." *Politics & Society* 33(2):253–276.
URL: *http://pas.sagepub.com/cgi/doi/10.1177/0032329205275194*

Benstead, Lindsay J. 2016. "Why Quotas Are Needed to Improve Women's Access to Services in Clientelistic Regimes." *Governance* 29(2):185–205.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Simon Galle, Edward Miguel, Daniel Posner, Bertil Tungodden and Kelly Zhang. 2016. "How Strong are Ethnic Preferences?".

Candelo, Natalia, Rachel T.A. Croson and Sherry Xin Li. 2017. "Identity and Social Exclusion: An Experiment with Hispanic Immigrants in the U.S." *Experimental Economics* 20:460–480.

Carpenter, Jeffrey P., Amrita G. Daniere and Lois M. Takahashi. 2004. "Social Capital and Trust in South-east Asian Cities." *Urban Studies* 41(4):853–874.

Gächter, Simon, Benedikt Herrmann and Christian Thöni. 2004. "Trust, Voluntary Cooperation, and Socio-Economic background: Survey and Experimental Evidence." *Journal of Economic Behavior and Organization* 55:505–531.

Greig, Fiona and Iris Bohnet. 2009. "Exploring Gendered Behavior in the Field with Experiments: Why Public Goods Are Provided by Women in a Nairobi Slum." *Journal of Economic Behavior and Organization* 70:1–9.

Habyarimana, James, Macartan Humphreys, Daniel Posner and Jeremy Weinstein. 2009. *Coethnicity: Diversity and the Dilemmas of Collective Action.* Russell Sage Foundation.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20(1):25–46.

Kurzban, Robert, Maxwell N. Burton-Chellew and Stuart A. West. 2015. "The Evolution of Altruism in Humans." *Annual Review of Psychology* 66(1):575–599.
URL: *http://www.annualreviews.org/doi/10.1146/annurev-psych-010814-015355*

Marshall, Leslie. 2019. Three Essays on the Possibilities for Realizing Women's 'Collective Interests' PhD thesis University of Pittsburgh.

Martinsson, Peter, Clara Villegas-Palacio and Conny Wollbrant. 2015. "Cooperation and Social Classes: Evidence from Colombia." *Social Choice and Welfare* 45(829-848).

Posner, Daniel. 2004. "The Political Salience of Cultural Difference: Why Chewas and Tumbukas are Allies in Zambia and Adversaries in Malawi." *American Political Science Review* 98(4):529–545.

Posner, Daniel N. 2017. "When and Why Do Some Social Cleavages Become Politically Salient Rather Than Others?" *Ethnic and Racial Studies* 40(12):2001–2019.

Raihani, Nichola J. and Redouan Bshary. 2015. "Why Humans Might Help Strangers." *Frontiers in Behavioral Neuroscience* 9(39):1–11.

Tognetti, Arnaud, Dimitri Dubois, Charlotte Faurie and Marc Willinger. 2016. "Men Increase Contributions to a Public Good When Under Sexual Competition." *Scientific Reports* 6(29819):1–10.

Van Vugt, Mark, David De Cremer and Dirk P. Janssen. 2007. "Gender Differences in Cooperation and Competition: The Male-Warrior Hypothesis." *Psychological Science* 18(1):19–23.

Wantchekon, Leonard. 2003. "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin." *World Politics* 55(03):399–422.

Zuhur, Sherifa. 2002. "Empowering Women or Dislodging Sectarianism? Civil Marriage in Lebanon." *Yale Journal of Law and Feminism* 14(1).