UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2017/163

Why do women co-operate more in women's groups?

James D. Fearon[1] and Macartan Humphreys[2]

September 2017

United Nations University World Institute for Development Economics Research
wider.unu.edu

Abstract: We examine a public goods game in 83 communities in northern Liberia. Women contributed substantially more to a small-scale development project when playing with other women than in mixed-gender groups, where they contributed at about the same levels as men. We try to explain this composition effect using a structural model, survey responses, and a second manipulation. Results suggest women in the all-women condition put more weight on co-operation regardless of value of public good, fear of discovery, or desire to match others' behaviour. Game players may have stronger motivation to signal public-spiritedness when primed to consider themselves representatives of the women of the community.

Keywords: public goods, gender, Bayesian estimation
JEL classification: H41, J16, D7, C11, C93

Acknowledgements: This paper uses data from an experiment implemented jointly with Jeremy Weinstein, to whom we are deeply indebted. We thank also Nicholai Lidow and Gwendolyn Taylor for leading the game and follow-up survey data collection teams in the field, and Andrea Abel, Jessica Gottlieb, and Amanda Robinson for their fieldwork on the public goods games. We thank the National Ex-Combatant Peace-Building Initiative for its research support in Liberia, in particular K. Johnson Borh and Morlee Zawoo, and Brian Coyne. AUSAID, through a grant to the Center for Global Development, provided the resources to implement the behavioural measures described in this paper; the Center for Democracy, Development, and the Rule of Law provided funding for fieldwork expenses, and the International Growth Center provided support for final data compilation and analysis.

[1] Stanford University, Stanford, CA, USA, jfearon@stanford.edu; [2] Columbia University, NY, USA, and WZB Berlin Social Science Center, Berlin, Germany, corresponding author: mh2245@columbia.edu.

This study has been prepared within the UNU-WIDER project on 'Gender and development'.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-389-9 https://doi.org/10.35188/UNU-WIDER/2017/389-9

Typescript prepared by Luke Finley.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Denmark, Finland, Sweden, and the United Kingdom.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

A substantial amount of development programming assumes that women have preferences or aptitudes that are more conducive to economic development. And indeed development funding is often channelled through women's groups rather than through all-male or mixed-gender traditional and modern authorities. For example, conditional cash transfer programmes commonly deliver funding to female household heads, and many microcredit schemes focus on women's savings groups. Programmes and reforms to increase women's empowerment in low-income countries are advocated on the grounds that, in addition to simple fairness, women in political office will spend more on public goods or on more-needed public goods, and empowered women in families will lead to more investment in agricultural productivity and children's health and education. Duflo (2012) reviews a growing literature assessing these and related hypotheses. Many studies find evidence that directing resources or political power towards women and women's groups increases children's wellbeing and public investments in clean water and (perhaps) other public goods.[1]

In this regard the views of the development community (Gates, 2014; World Bank, 2012) often reflect those of the rural poor in many low-income countries. For example, it is often held that extra cash is more likely to be spent on alcohol and grilled meat by men, versus household needs or productive investment by women. Women may also be seen as more community-minded on average—indeed our own survey data, described below, reflects such views.

What is not clear from these accounts, however, is whether these beliefs reflect views about the attitudes and behaviour of women versus men, or rather some features of women's groups in particular. Also not clear is the reason for any such differences in behaviour across genders or gender groups.

We examine these questions by analysing play in a public goods game in northern Liberia in which the gender composition of groups making collective decisions over public goods was randomly assigned.[2] We find that women did contribute substantially more than men, *though only when they knew that they were playing with other women*. In public goods problems involving equal numbers of men and women, men and women contributed similar amounts, and markedly less on average than in the groups with only women players.

This main finding partially supports the arguments of development practitioners who seek to engage communities through women-only groups. It does not provide clear evidence in favour of the assumption that women are per se more community-minded when asked to make decisions between private and social goods (though of course this could be the case for particular public goods in particular settings).

---

[1] Duflo cautions, however, that for economic development, women's empowerment 'is not the magic bullet it is sometimes made out to be' (2012: 76), stressing that women face so many constraints that any multiplier effects of alleviating any one may be limited or absent, and that women's empowerment can yield improvement in some dimensions at the expense of others.

[2] The data (Fearon et al., 2014) can be accessed at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/28006.

It is one thing to find that in a particular context all-female groups generate more collective action, and another to explain why. In the second part of the paper we use Bayesian methods, data from surveys of game players, and knowledge of a multiplier on contributions that was randomly varied across players, to estimate a simple structural model of individual decisions about how much to contribute. The goal is to gain insight into how the motivations of the women acting in the all-women groups differed from those of men and women in the mixed groups.

The model and data allow us to estimate the weight that people in our sample put on four different underlying motivations, or preferences, that could factor into different contribution decisions:

1. the value the individual places on the expected use of the total amount of money raised;
2. the individual's fear of discovery of their own contribution and thus possible sanctioning;
3. the individual's concern with matching what others do; and
4. the intrinsic value (or cost) the individual has for contributing, including both opportunity costs for the money donated and any positive value for signalling 'community spirit' by contributing.

The results suggest that women in the all-women groups had stronger intrinsic motivation to contribute than both women and men in the mixed groups. That is, they would have contributed a substantial share of their endowment irrespective of beliefs about what it would be spent on, fear of discovery, or desire to match others' contribution levels.

These patterns are consistent with field observations suggesting that women in the all-women communities may have contributed more in order to signal that the women of their community have powerful community spirit. Thus a sense of group solidarity and implicit competition—a social-identity effect—may have favoured collective action. If this is correct, it would suggest a mechanism that might favour channelling development aid through women-only groups in some contexts, but not necessarily others.

In Section 2 we describe the context of our study and provide a description of the experiment. In Section 3 we provide the basic results, showing differences in behaviour between gender groups and the effects of group composition. Section 4 compares our results with the most closely related results in the literature, found in Greig and Bohnet (2009). Section 5 explores mechanisms more thoroughly, using a simple structural model to assess the role of intrinsic incentives and three types of instrumental incentives, relating to conformity concerns and concerns over sanctioning. While all three of these appear to play a role, the differences in intrinsic incentives between conditions appear most salient. In the concluding Section 6 we speculate, based on the results presented and additional field observations, about possible reasons for the main finding. We suggest that the evidence for strong intrinsic motivation to contribute is consistent with the hypothesis that many game players thought they were signalling community quality to potential donors. Women in the all-women groups may have been more strongly motivated in this way due to a social-identity effect—a sense that they were playing for, or representing, 'team women' of the village.

2 A public goods game in post-conflict Liberia

We examine differences in the effects of the gender composition of groups by analysing play in a public goods game implemented in 2008 following an international development intervention in Liberia.

As described in Fearon et al. (2015), we used a public goods game to assess the effects of a community-driven reconstruction (CDR) programme implemented between 2006 and 2008 in 83 communities in two districts in northern Liberia, Voinjama and Zorzor. Funded by the UK's Department for International Development and implemented by the International Rescue Committee, the programme sought to foster reconciliation and improve local-level governance in the wake of a long civil war that ended in 2003, and that was particularly intense in these districts. Secondarily, the implementers and funders hoped that the funds provided would have positive economic effects. The primary goal of the intervention, however, was institution building to improve the collective action capacities of post-conflict communities.

Based on the popular model of 'community-driven development' (Mansuri and Rao 2012), CDR works by giving communities power in selecting how the offered aid will be used, and also in governing the implementation and management of projects that result. The 'catch' is that communities are required to construct and use donor-specified institutions for choosing and managing the projects. In particular, under aid agency guidance the communities elect community development committees (CDCs) in open elections that (at least in this case) exclude the main traditional leaders (chiefs). The CDCs then deliberate and hold town meetings to decide on how the development funds provided will be spent, subject to some parameters given by the aid agency.[3]

The public goods games, implemented in 2008 after the CDR interventions were (almost entirely) completed, were our primary measurement strategy for estimating whether the CDR programme had a causal impact on treated communities' ability to generate collective action after a devastating civil war. The games presented treated and control communities with a small-scale challenge of organizing to raise funds that we would match at an average rate of 250 per cent, for a project chosen entirely by the community in whatever manner they decided. Described to the communities as a 'small-scale development project' in which they could obtain matching funds depending on how much they raised from contributions made during the game, the measurement strategy was intended to approximate the kind of real-world collective-action problem communities had faced in the CDR programme (which also required matching contributions), or indeed any community collective problem.

The basic protocol was as follows: we held a community meeting in each of the 83 villages, announcing that a public goods game would be played the following week to determine how much financing would be provided to a community for use in any way the community desired. Attendees of the meeting were told that 24 households would be randomly selected and then a randomly chosen adult from each of those households would receive a sum of money (300 Liberian dollars, LD, or about US$5). The game player could then decide, in private, how much of this to keep and how much to contribute towards a public good. It was explained that after the game had been played, we would hold another public meeting to open the private contributions and then add our 'matching' contributions on top, according to a multiplier known in advance to each game player and identified on the envelope. The total amount raised would then be given to three 'community

[3] In our case, projects had to be for community-wide rather than private or narrowly targeted benefit, and purchase of capital equipment for income-generating projects (such as a rice mill) was not allowed. The projects chosen tended to involve construction of community facilities, such as community meeting houses and guest houses (approximately 35 per cent), latrines (30 per cent), and hand-dug wells (15 per cent). A few projects (less than 5 per cent) focused on school or health clinic construction; almost none were in agriculture, skills training and small business development, and other income-generating activities. The median value of total grants was about US$13,000 for a community, the specific amount depending on the community's size and their proposals.

representatives' who could be selected by the village in any manner they wished in the week between the first community meeting and game day.

Notice that communities knew that in the week between the initial meeting and game day, they (or their leaders, or whoever) had time to decide how the funding would be used and could engage in mobilization activities around participation in the game. They did not know which individuals would be picked to play and so could not lobby or pressure specific people.[4]

When we implemented the public goods game we included two variations that would allow us to assess differences in play between men and women, as well as the effects of group composition. The first variation, similar to that employed by Greig and Bohnet (2009), was that in half the communities we sampled 24 women to play the game and in half we sampled 12 men and 12 women. As a result we have data on the individual contributions of 504 men and 504 women in 42 communities who played knowing that other game players were 11 of the same gender and 12 of the other gender, and of 971 women in 41 communities who played knowing that all 23 other game players were women.

The second variation is in the multiplier applied to contributions. Within each village/gender block half the players had their contributions increased by a factor of two and the other half had theirs increased by a factor of five. Below, we use this variation to seek to assess the role of other-regarding preferences in contribution decisions.

In addition to the games data we also gathered data from a survey implemented immediately after a player made their private contribution choice. From this survey we constructed a set of measures used in the analysis that follows.

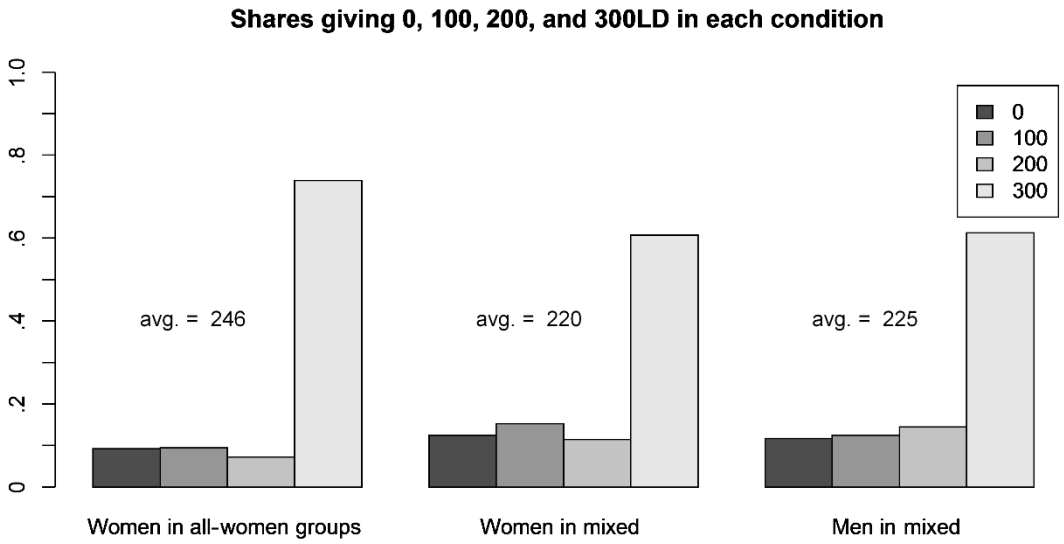
Our focus in this paper is on the main effect of the gender composition treatment, the difference between contribution levels in the 42 'mixed' communities where 12 men and 12 women played the game, and the 41 'all-women' communities where 24 women were selected to play. This gender composition treatment was assigned by us at random, orthogonally to the CDR treatment assignments. We should note that due to low power, we were not able to have a treatment arm of villages in which only *men* were selected to play the game, which limits our ability to draw some inferences, as discussed below.

3 Basic results

The raw game data, broken down by gender and treatment condition, is shown in Figure 1. We see that women in the mixed condition give about the same as men—indeed slightly less. But they give considerably more in the women-only condition. This latter effect can be interpreted causally as the effect of gender composition on women's contributions.

[4] For greater detail on the development intervention and the public goods game, see Fearon et al. (2015).

Figure 1: Contributions by treatment/gender group



Source: Authors' own construction.

Table 1 shows these results with estimation of treatment effects implemented at the village level, taking account of randomization blocks, including status with respect to the CDR intervention (see Fearon et al. 2015 for details on the randomization), together with Neyman standard errors and *p* values generated via randomization inference for the full sample of women (Column 1) as well as for the urban and non-urban sub-populations. We see a large and strongly statistically significant treatment effect: 30 LD is approximately 70 per cent of one standard deviation of the variation in average individual contributions in the mixed communities, and it is almost twice the size of the average effect of getting the 400 per cent interest rate versus the 100 per cent rate.

Table 1: Effects of composition

	All	In quarters	Outside quarters
Mixed villages	218.55	182.15	240.86
Homogeneous villages	248.62	232.91	252.61
Difference (average treatment effect)	30.06	50.75	11.75
N	82	28	54
p (ri)	0.00	0.00	0.15
s.e. (Neyman)	9.43	13.59	11.73

Source: Authors' own construction.

In the second and third columns of Table 1 we report evidence of an interesting and puzzling heterogeneity for which we have no confident explanation. Of the 83 communities studied, 28 (12 assigned to all-women and 16 to mixed) were 'quarters' of a larger town, and thus slightly more urbanized than the remaining 55 villages, which were rural. Average contributions in the public goods game were much lower in the quarters, reflecting lower levels of organization and mobilization capacity (Fearon et al., 2015). But we see that the gender composition effect is much larger in the quarters: mixed communities in quarters generated very low contributions, whereas the quarters where only women played did much better. In line with our argument and

interpretation of the evidence on the CDR treatment effect in Fearon et al. (2015), this difference could result from women's traditional organizations functioning better for mobilization in the quarters than the official chief and sub-chief system, which seemed less well established than in the rural communities. But this is speculation.

Our goal in what follows will be to make sense of the basic pattern of substantially greater collective action produced in the communities where women knew that only other women were making contribution decisions, versus communities where both men and women played the game.

4 Comparison with Greig and Bohnet (2009)

Greig and Bohnet (2009) implemented a study closely related to ours which allows for the possibility of assessing the consistency of findings in two locations on the continent.

Though our game was similar in many ways to the game studied in Kenya by Greig and Bohnet (2009), a number of important differences are worth highlighting. As summarized in Table 2, in the Kenya study all participants came from a single community whereas in our study communities were the unit of randomization. In Kenya, participants played in groups of four in a type of lab established at the community centre; we worked with much larger groups of 24 subjects that were randomly sampled from villages, with these subjects playing in their own homes. The endowments in the Kenya game were smaller—approx. US$0.64 rather than approx. US$5; the multiplier was 2 whereas ours was either 2 or 5.[5] Because of the different sizes of the groups, however, the private returns to the public investment were much smaller in our study. In the Kenya experiment the players were the only beneficiaries of the contribution—which could lead to a total value of less than US$3—whereas in Liberia the entire village (median population of about 500) could in principle benefit from a project worth up to US$400. Also critical is the fact that the Kenya experiment did not allow for pre-play communication whereas the Liberia experiment did. In total, Greig and Bohnet (2009) studied 270 subjects in 68 groups of four whereas we studied 1979 subjects in 83 groups of 24 (with limited attrition).

Table 2: Comparison of designs used in Kenya and Liberia

	Kenya	**Liberia**
Communities studied	One community (slum)	83 communities
Unit of randomization	Subjects	Community
Beneficiaries of public good	Subjects	Community
Group-level experimental arms	3: men, women, mixed	2: women-only, mixed
Group size	4	24
Number of groups	68	83
Number of subjects	270	1979

[5] In Greig and Bohnet, for every x contributed, all four members received $x/2$.

Pre-play communication	No	Yes
Endowments	$0.64	$5
Multiplier	2	2 or 5
Private return on investment	0.5	< 0.083 / < 0.21
Max value of public good	$3	$400
Other games played	Investment game	No other games
Additional treatments	None	Multiplier variation

Note: The number of subjects is not exactly equal to the number of groups times the group size in either study because of small doubling of roles in Kenya and a small amount of attrition in one village in Liberia.

Source: Greig and Bohnet 2009 (Kenya) and Fearon et al. 2015 (Liberia).

A key feature of the Kenya analysis is the use of data on player expectations about the play of others, generated by asking subjects how much on average they expected other game players to play. We have similar data for Liberia, reported in Table 3, including a measure of the number of others that players expected to contribute nothing; the number of others they expected to contribute the full amount; an estimate of the expected average contribution using these numbers; and a measure of whether they expected men or women to contribute more, or about the same (recorded here as 0 for men give more, 0.5 for the same, and 1 for women give more).

We see from Table 3 that expectations broadly tracked the compositional condition: women did not have higher expectations about the contributions of others than men did in the mixed condition, but did have higher expectations in the women-only condition. Nevertheless, women in both conditions reported believing that women would contribute more than men in the mixed condition, while men on average said they expected equal contributions by gender. These facts could on their own give rise to a belief in overall higher levels of contribution by others in the women-only condition—a feature we return to below.

Table 3: Expectations given different treatments

	Women-homog.	**Women-mixed**	**Men-mixed**
Expected share giving 0	0.11	0.13	0.14
Expected share giving 300	0.85	0.80	0.81
Expected average amount given by others	273.45	258.66	254.52
Actual average given by others	245.93	222.92	222.92
Predict women give strictly more	0.83	0.73	0.48

Source: Authors' own construction.

Table 4 uses essentially the same specification as in Greig and Bohnet (2009, table 1); there is no interaction term in our model because we do not have an all-male group, and we cluster standard errors at the group level (the level of randomization). For ease of interpretation we use women in mixed groups as the base condition and then look at the differences for men, and for women in all-female games ('homogeneous'). The results in Column 2 use list-wise deletion for the often-missing expectations variable; in Column 3 we impute average community values for missing

expectations, while controlling for missingness and allowing for different effects for units with missing data. We note that in this specification and later specifications that employ expectations, the model includes a post-treatment variable (expectations), and the other coefficients can only be interpreted as direct effects conditional on strong assumptions outlined in Baron and Kenny (1986).

Table 4: (External) replication of table 1 (cols 1 and 2) in Greig and Bohnet (2009)

	1	2	3	4
Constant	220.437 (8.392)**	106.968 (20.505)**	109.919 (20.492)**	86.707 (32.737)**
Homogeneous	25.496 (9.653)**	19.541 (10.536)	15.464 (8.68)	80.63 (4024)**
Male	4.96 (6.916)	10.988 (9.334)	6.297 (6.88)	10.842 (35.21)
Expectations		0.378 (0.07)**	0.379 (0.07)**	0.467 (0.117)**
Homog. * expectations				−0.24 (0.145)
Male * expectations				−0.017 (0.13)
N	1979	1093	1979	1979

Note: ** indicates $p < .01$.

Source: Authors' own construction; Greig and Bohnet 2009.

We see some features here that are consistent with Greig and Bohnet (2009) and some differences. First, our constant is considerably larger (as a share of endowments) than those found in Kenya and more in line with, if not greater than, those found elsewhere: even the group that contributed the least, women in the mixed condition, gave about 220/300 = 73 per cent of their endowment. Thus the evidence of weak social capital from the Kenya study is not replicated here. This could be related to the facts that our participants were on average members of relatively small rural communities rather than a large urban slum, and that the communities also had a week to inform community members about the game and the projects. On the other hand, as noted above, personal returns from contributing were arguably much smaller in our setting (because diffused over the broader community).

Second, as in the Kenya study we find that women give considerably less in the mixed condition than in the women-only condition. In Greig and Bohnet (2009) the marginal effect for women in the mixed condition was a drop equal to about 10 per cent of the endowment (it is not clear whether this effect is significant or not). We see a drop of 26 Liberian dollars (or 30 from the non-parametric estimation in Table 1), which is also close to 10 per cent of the endowment. In addition we see a strong relationship between expectations and contributions, though our estimated coefficient is considerably lower, at close to one third compared to one half in Kenya.

An important point of difference between the results, however, is that we do not see strong evidence that women contribute less than men in the mixed condition. (We are unable to assess whether the effects of the mixed condition are different for men and for women.) And so the explanation provided in Greig and Bohnet (2009) for the effect of homogeneous groups on women does not find clear support here. Greig and Bohnet (2009) argue that women have overly pessimistic expectations in the mixed conditions while men have overly optimistic expectations. In contrast (see Table 3), women and men have similar expectations in mixed groups—they are both overly optimistic—and although women expect contributions to be about 14 LD higher in homogeneous communities, this is nowhere near large enough to account for differences in play,

if the effects of expectations in Table 4 are to be believed.[6] Perhaps as importantly, as seen in the final column, the estimated effects of expectations appear to be weaker for women (by about 50 per cent) in homogeneous groups.

5 Model and implications for interpreting the evidence

If (inaccurate) differences in expectations cannot account for the effect of composition, what can?

In Fearon et al. (2015) we found evidence that contributions were related to greater levels of mobilization activity in the week before the game in CDR-treated communities, although only in communities where both men and women could be selected to play. Mobilization does not, however, appear to explain why women contributed more in the women-only villages.

In the survey, we asked if the respondent had been contacted by anyone about how to play the game, about the community project, or about staying home on game day, and also about whether they knew of other community meetings to discuss the game and whether they knew the community representatives' names. Table 5 shows that the average of yes or no responses (yes equals 1) for these five questions is slightly greater for women in the homogeneous groups versus women in the mixed groups, although this difference is not statistically significant. The rates are about the same for women in the women-only condition as for men in mixed. Dividing the sample by whether the community received the CDR programme (which does appear to have affected mobilization activity in the mixed communities), we see that in the no-CDR, mixed-group villages women were markedly less mobilized than women in the no-CDR, all-women villages. However, in the CDR-treated villages, if anything, there is more mobilization of both genders in the mixed groups. So overall this factor does not seem likely to explain why contributions were higher in the women-only communities.

Table 5: Reports of mobilization activity by condition and CDR treatment status, 0–1 scale

	Women in all-women	Women in mixed	Men in mixed
All	0.46	0.41	0.47
No-CDR	0.47	0.33	0.42
CDR	0.45	0.48	0.51

Source: Authors' own construction.

Experiments are excellent for drawing inferences about what causes what, but, by itself, finding that X causes Y does not explain why this is. An explanation is arguably more important in social science settings than in, say, drug testing, since causal effects are more likely to differ across contexts and settings. This puts a greater premium on learning about mechanisms (which is arguably where the 'science' is in biomedical research as well).

In what follows we develop a simple structural model of the decision to contribute that has parameters of interest corresponding to four different possible motivations for contributing. We then use a Bayesian hierarchical model to estimate these structural parameters, and finally compare the estimates and implications for women in the women-only villages to men and women in the

[6] In contrast, the differences in beliefs are close to large enough in Greig and Bohnet (2009) to account for the differences in play, at least using their model estimates. There, the homogeneous condition was associated with a 13.5-percentage-point increase in expectations, which translates into a $0.57 \times 13.5 = 7.7$-point difference in behaviour, not very different to the observed 9.5-point difference.

mixed villages. The idea is to use the structural model to help with the problem of drawing inferences about mechanisms from a diverse set of survey responses and the results of several experimental manipulations that are themselves implemented at different levels.

We highlight that our analysis in this section is exploratory; although we set out to measure effects of composition on contributions we only focused on parameter estimation after seeing the core results.

5.1 Model of decision making

We assume that when deciding what contribution $x_i \in \{0, 1, 2, 3\}$ to make, players seek to maximize:

$$u_i(x_i) = \alpha_i x_i - \gamma_i(x_i - \rho_i \hat{x}_i)^2 + r_i x_i + \phi_i q_i x_i$$

where $\hat{x}_i$ is *i*'s expectation regarding the average contribution of others in the same village, r_i is *i*'s multiplier, and q_i is the *i*'s perceived probability of having his or her action discovered by others. For simplicity we ignore the expected benefits from contributions by others (these drop out under the assumption of linear gains in the public good).

Note that we assume quasilinear utility and treat the valuation of the public goods as the numeraire. This is captured by the absence of a coefficient on $r_i x_i$. Recall that r_i is either 2 or 5 and was randomly assigned to *i* by us; q_i and $\hat{x}_i$ are empirical measures derived from the survey data. Our structural parameters of interest are α_i, γ_i, ρ_i, and ϕ_i.

- Parameter α_i reflects the intrinsic value of contributing to the public good. This is often assumed to be negative, reflecting the opportunity cost of not having the money for own spending, although we allow for the possibility that it is positive on net. For example, it can be positive if individuals feel sufficiently good about doing what they see as the right thing, or they see making a contribution as a signal to the community, to the foreign donors running the project, or even to themselves that they are 'good types' who are community-spirited.[7]
- $\gamma_i \geq 0$ and $\rho_i \geq 0$ reflect conformity concerns: with $\rho_i = 1$ a player values contributions equal to those of others contributing; with $\rho_i < 1$ a player seeks to contribute less than others; and with $\rho_i > 1$ she seeks to contribute more than others. Parameter γ_i measures the weight that *i* puts on matching what he or she believes others are doing (or, more precisely, ρ_i times this).
- Parameter ϕ_i reflects *i*'s valuation of incentives arising from social rewards or punishments from contributions. Note that we assume that players can be concerned about sanctioning from the village for less-than-complete contributions and do not assume that sanctioning relates only to deviations from equilibrium play.

Individuals choose between options $x_i \in \{0, 1, 2, 3\}$ (in hundreds of Liberian dollars). For purposes of estimation, we will assume that the most consequential variation across individuals in a community is in α_i, their marginal value for contributing independent of use of the funds, or

[7] Individual contributions were anonymous—not seen by the community—although game players were told that their individual contributions could be observed by the researchers at a later date.

desire to match or avoid punishment, relative to the other motivations. Fixing the other parameters, the optimal choice for individual *i* is increasing in α_i and can be characterized by three cut-points in α_i of the form $c_i^j = (2j-1)\gamma_i - (2\gamma_i\rho_i\hat{x}_i + r_i + \phi_i q_i)$ for $j = 1, 2, 3$.

If $\alpha_i < c_i^1$ then *i* optimally would choose $x_i = 0$. If $\alpha_i \in (c_i^1, c_i^2)$ then $x_i = 1$ is preferred; $\alpha_i \in (c_i^2, c_i^3)$ implies $x_i = 2$, and $x_i = 3$ if $\alpha_i > c_i^3$.

In particular, suppose that in a village the α_i are distributed by the cdf F, whereas the other parameters (γ, ρ, ϕ) depend only on gender and village-level features, including whether the group composition in the village is mixed or homogeneous. Then for each *i* we have probabilities for choices of $x_i \in \{0, 1, 2, 3\}, w_i^x$, where

$$w_i^0 = F(c_i^1)$$

$$w_i^1 = F(c_i^2) - F(c_i^1)$$

$$w_i^2 = F(c_i^3) - F(c_i^2)$$

$$w_i^3 = 1 - F(c_i^3).$$

It follows that if we specify a distribution F for the α_i—say, normal with mean α and standard deviation σ—then we can compute the likelihood of the observed choices of x_i for the 24 game players in a particular community. The five unknown parameters entering the likelihood function—and characterizing motivations to contribute—are $\Theta = \{\alpha, \sigma, \gamma, \rho, \phi\}$.

For ease of reference Table 6 summarizes the meaning of each motivation or preference parameter, along with some estimation assumptions discussed in the next section.

Table 6: Summary of motivation/preference parameters and estimation assumptions

Parameter	Motivation/preference	Estimation assumptions
α_i	*i*'s marginal value for contributing independent of use of funds, matching, and sanctioning concerns	Varies across individuals within gender groups in villages; α_i is not estimated
α	Mean of the distribution from which α_i is drawn	Varies by community and potential condition for each gender
σ	Standard deviation of the distribution from which α_i is drawn	Varies by potential condition
γ	Weight put on matching target contribution $\rho\hat{x}_i$	Varies by potential condition for each gender
ρ	Share of reported expectation $\hat{x}_i$ that *i* would ideally match if no other motivations	Varies by potential condition for each gender
ϕ	Weight on contributing to avoid sanctioning/discomfort if revealed to have given less than 300 LD	Varies by community and potential condition for each gender

Notes: The composition treatment conditions are women-only and mixed. By 'varies by potential condition' we mean that there are distinct parameters for women for each of these two conditions.

Source: Authors' own construction.

5.2 Empirical model

We use a multilevel Bayesian model to estimate the key parameters, letting these be a function of respondent gender and treatment condition. Our model estimates the parameters in the first equation in Section 5.1, as a function of treatment condition, taking q and $\hat{x}$ as given.[8]

Let $\theta[i]$ denote the value of the parameter $\theta \in \Theta$ that we use to calculate the likelihood that person *i* chose the observed x_i.

Under the assumption that α_i is not observed, the likelihood is calculated using the probability that an individual takes action $x_i \in \{0, 1, 2, 3\}$, given by the categorical distribution with event probabilities w as defined above.

$$L(x_i) = f_{categorical}(x_i | w(r_i, q_i, \hat{x}_i, \alpha[i], \phi[i], \gamma[i], \rho[i], \sigma[i])$$

These individual-level parameters are generated from condition- and group-level parameters as follows. Let W_i, M_i, and H_i denote indicator variables for being a woman player in a mixed village, male in a mixed village, or female in a homogeneous village, respectively, and let $\theta_j, j \in \{H, W, M\}$ denote condition-level parameters, described below.[9]

We let α and $\emptyset$ vary by condition and village as combinations of village-level and condition-level features.

$$\alpha[i] = M_i\alpha_M + W_i\alpha_W + H_i\alpha_H + v^{\alpha}_{v[i]}$$

$$\phi[i] = M_i\phi_M + W_i\phi_W + H_i\phi_H + v^{\phi}_{v[i]}$$

where $v[i]$ denotes the village to which *i* belongs. We assume the village-level random effects have distributions given by: $v^{\alpha}_{v} \sim N(\mu_\alpha, \sigma_\alpha)$ and $v^{\phi}_{v} \sim N(\mu_\phi, \sigma_\phi)$.

Other parameters we let vary by condition only; thus for $\theta \in \{\gamma, \rho, \sigma\}$:

$$\theta[i] = M_i\theta_M + W_i\theta_W + H_i\theta_H$$

Each of these last three is constrained to be positive. In addition we constrain $\sigma_M = \sigma_W$. We employ diffuse priors on all parameters and hyperparameters, given by normal (or half normal) distributions with mean 0 and standard deviation of 5.

Thus we estimate different parameters for men and for women for each treatment condition: that is, we use the model not simply to measure parameters but also to infer the counterfactual parameters that would arise were women in different treatment conditions. We thus seek to assess whether gains due to composition may be attributed to differences in instrumental motivations

---

[8] Thus in the model presented here we do not model $\hat{x}$ as a function of treatment, though this can in principle be added as an additional component of the likelihood function. We note that since $\hat{x}$ is post-treatment, treating it as fixed, as done here and also in Greig and Bohnet (2009), could in principle introduce post-treatment bias.

[9] We note a slight abuse of notation here since we let α_i denote person *i*'s value for α as defined above,

and $\alpha_M, \alpha_F, \alpha_H$ denote condition-level parameters.

deriving from expectations about the actions of others, concerns around sanctioning, or differences in valuations of the public good or in terms of intrinsic motivations. In addition we allow for considerable village-level heterogeneity, at least for parameters α and ϕ.

5.3 Effects on structural parameters

Estimated posterior means for our five structural parameters in three conditions (women-only, women in mixed communities, and men in mixed communities) are shown in Table 7. There are two sorts of comparisons of interest. First—and most relevant to our question of why higher contributions were made by women in the women-only groups—is the treatment effect of the group's gender composition on women players. For parameter θ this is given by $\theta_H - \theta_W$ in the fifth column. Second, within each condition we can compare the relevant weight the average respondent is estimated to have put on different motivations. Recall that these numbers are relative to the value put on money raised from the game for the community, which has been normalized to 1.

Table 7: Parameter estimates

Parameter	Men (mixed)	Women (mixed)	Women-only	Composition effect	Pr > 0
α	1.03	−0.06	5.31	5.37	0.92
ϕ	−0.16	−0.01	0.14	0.15	0.51
γ	2.97	2.86	2.02	−0.84	0.04
ρ	0.93	0.97	0.81	−0.16	0.30
σ	12.05	12.05	11.66	−0.39	0.42

Notes: σ is constrained to be the same for men and women in the mixed condition; the final column shows the posterior probability that the difference between women-only and mixed conditions (for women) is positive.

Source: Authors' own construction.

Consider the treatment effects on women players first. The most striking positive effect we see here is the difference in α for women between conditions. From these estimates, α—the village-/condition-level mean of the distribution of α_i—is five points higher for women in the homogeneous community, and is, on average, positive. This means that for the typical woman in these groups, every 100 LD invested gives the equivalent of an intrinsic 500 LD direct positive return in units of the value of the total funds raised, independent of what anyone else does. With σ around 12, approximately one third face marginal costs of contributing and two thirds marginal benefits (more generally, these shares depend on the village-level intercept also). In contrast, in the mixed group the mean for women is negative and close to 0, meaning that the typical woman gains no intrinsic benefits from contributing, and half face costs.

In the typical case we observe little responsiveness to fears of sanctioning and see little difference across conditions.

We see relatively large *negative* effects on γ, however, meaning that in the gender-homogeneous groups women put *less* weight on matching what they expected others to contribute than they did when they knew they were playing with men. Note that in most cases ρ is close to 1, though possibly lower in the homogeneous group condition; thus if there were no other considerations, both men and women in the mixed groups would have wanted to match what they thought others were doing, whereas women in the homogeneous condition would have wanted to undercut others a bit. Overall, the matching incentives pulled offers upwards whenever players were optimistic

about the contributions of others.[10] The result of this is that women in homogeneous groups had a substantially *weaker* motivation to contribute in order to match what others were expected to do, but this negative effect was more than offset by the increase deriving from the greater intrinsic motivation to contribute that worked through α.[11]

The estimates also provide a sense of the relative importance of different motivations in game players' contribution decisions. For instance, in the mixed groups, on average both men and women have γ values around 3, which means they put about three times the weight on the discomfort of deviating from the expected contribution of others by 100 LD as they did on increasing the total contributed by 100 LD. Women in the homogeneous condition similarly put more weight on conformity, though not as much as in the mixed groups. The sanctioning concern seems to matter hardly at all relative to the other motivations in any condition. As noted, the big difference is in the weight put on contributing, independent of the total raised or the conformity concerns. The suggestion is that especially for many women in the homogeneous condition, but also to a smaller degree for men in the mixed groups, contributing was typically not seen as a net cost that they would not want to pay unless they thought that what the money raised would be spent on would compensate them. Rather, the results suggest that for many game players, the game did not have the structure either of a classical public good problem, or of a simple coordination game in which contributing is costly but one might do it for conformity reasons. Instead, for quite a few players, contributing was a dominant strategy independent of hopes for what the money raised would be spent on.

6 Conclusion

Employing a public goods game in 83 villages in northern Liberia, we use random assignment of gender composition of the groups of individuals making contribution decisions to estimate the causal effect of playing with a mixed-gender group versus a group comprising only women. We find that the all-women groups contributed substantially more to a community project than did either men or women in the communities where both genders played. This is therefore a group composition effect rather than a 'women are unconditionally more community-minded' effect. Women contributed more when they knew they were playing with other women, but not more than men in the mixed groups.

Our main result thus supports the logic of practitioners who seek to engage communities through women-only groups. It does not provide clear evidence in favour of arguments or the assumption that women are per se more community-minded when asked to make decisions between private and social goods, although this might be the case for particular public goods in particular settings.

The basic finding is similar to that of Greig and Bohnet's (2009) lab experiment conducted in a Nairobi slum, where groups of four played a public goods game in different gender compositions.

[10] In particular, we can see from the utility function that the marginal gains in contributions are increasing in γ (that is, $\frac{\delta^2 u}{\delta x_i \delta \gamma} > 0$) when $-2(x_i - \rho\hat{x}) > 0$ or $x_i < \rho\hat{x}$. Note that here we consider best responses, treating expectations as exogenous.

[11] The negative effect arising from the drop in γ is partially compounded by the drop in ρ, which lowers women's targets for matching.

They also found that women contributed more in all-women groups than in mixed sets. The similarity is interesting given the many differences in the set-up and context.

Our account of why women contribute more in women-only groups differs from the explanation given in the Kenya experiment. In Kenya, differences appeared to result from women holding overly pessimistic views about the behaviour of players in mixed groups (whereas men held overly optimistic views). In contrast, in Liberia differences in the beliefs of men and women were slight and differences in expectations across conditions did not appear large enough to account for differences in contributions, at least when our estimated model was employed.

Rather, our model and the data suggest that women placed great intrinsic value on contributing to the production of public goods when they knew that they were working collectively with other women. We cannot assess systematically whether this is because mobilization undertaken by women in advance of play was more effective at clarifying appropriate norms of behaviour, or extracting promises around behaviour, or whether the intrinsic motivations stemmed from greater solidarity within the gender groups.

We close, however, with some speculation motivated by the model results and field observations. In the initial community meetings to introduce the project and game, our local collaborators[12] explained that we had two main purposes: to provide some funds for a small development project that the community could choose, and also to conduct research to understand better the people and their lives in these communities. The second reason was given as a way of explaining why we were employing the unusual game procedure, which was referred to both as a process of raising matching funds (an idea that many communities were familiar with), and as a 'game'. (Using a local analogy, participants and enumerators sometimes spoke of a 'lucky ticket', meaning that getting picked to play was like winning a lottery.)

Attendees at the initial meetings where the game was explained immediately and audibly grasped the conflict between private and social good posed by the decision of how much of the endowment to keep and how much to put in the envelope. On one occasion (at least), an attendee was heard to say 'They are testing us', meaning that his interpretation was that we wanted to learn how community-spirited people in their village were. Our introductory scripts emphasized that the decision to keep some or all of the endowment was private and that it could be justifiable to do so—we avoided any language clearly identifying contribution as the right choice. Nonetheless, a possible inference by community members would have been that if the community contributed a lot, they would be more likely to receive more development assistance in the future.

If so, then game players may have had a rationale for contributing largely independent of interest rates, expectations about others' contributions, or value for the community project. The act of contributing the whole endowment may have been understood by some or many participants as a signal of community spirit to outsiders thought to have access to more resources. It was evident that residents of these impoverished communities were desperate for 'development', a term they frequently used. This rationale could explain some part of our estimates of powerful intrinsic motivations to contribute.[13]

---

[12] The Liberian non-governmental organisation NEPI (Network for Empowerment and Progressive Initiative).

[13] We note that this motivation might also bias our expectations measure upwards, contributing to the apparent 'over-optimism' we observe. That is, when asked how much they thought others in the community would contribute in the game, some respondents may have been saying, in effect, 'A lot, because this is a good community you should bring development to'.

But how, if at all, might such a rationale explain the greater contributions of women when they knew they were playing only with other women? This is speculative, but a possibility suggested by some field observations is that the motivation to signal was greater the more a player identified with the defined set of other game players. Understanding yourself as a representative of 'the women' of the community rather than as a random community member may increase the desire to signal—to the outsiders, to the rest of the community, and to yourself—that you and your group are made of the right material. On hearing that only women could be chosen to play the game, the women in the audience sometimes seemed to feel pleased and important, perhaps as if proud or excited to be chosen as representatives of their community. (By contrast, occasionally one or two men would lose interest and walk away when it was announced that only women could be chosen.)

If correct, this interpretation also sheds light on the effects of the CDR programme itself. We note that these effects of the composition treatment are quite different to the effects on parameter values we estimate for the CDR treatment, using an analogous model. In models that include parameters for CDR effects (not reported here) we see that although both CDR and gender homogeneity are associated with greater contributions, parameter estimates suggest that CDR is associated with a *drop* in α which is offset by changes in γ and ρ. This is consistent with an interpretation in which group homogeneity induces identity concerns, whereas CDR facilitates collective action through enhanced coordination. Though CDR is sometimes promoted as a way of fostering greater group identity, these results, consistent with our analysis in Fearon et al. (2015), suggest that it was organization and not identity that mattered.

The speculation above is in line with a large literature and tradition, not much known or drawn on in the design of development interventions, of motivating collective action by appeals to subgroup solidarity or even competition.[14] It is understandable that development practitioners, who put a high premium on inclusion and avoidance of conflict, would not consider trying to generate collective action by these means. But perhaps creative thinking might be able to harness the power of group identification and competition in order to generate collective action in support of development projects, without, or with minimal, downside risks.

[14] It is unfortunate that with only 83 communities under study we did not have the power to include a third treatment of villages where only men could play the game. This might have shed light on whether we observed an 'all-women' effect or a single-gender group effect.

References

Baron, Reuben M., and David A. Kenny (1986). 'The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations'. *Journal of Personality and Social Psychology*, 51(6): 1173–82.

Duflo, Esther (2012). 'Women's Empowerment and Economic Development'. *Journal of Economic Literature*, 50(4): 1051–79.

Fearon, James D., Macartan Humphreys, and Jeremy M. Weinstein (2015). 'How Does Development Assistance Affect Collective Action Capacity? Results from a Field Experiment in Post-Conflict Liberia'. *American Political Science Review*, 109(3): 450–69.

Fearon, James, Macartan Humphreys, and Jeremy Weinstein (2014). 'Replication Data for: How Does Development Assistance Affect Collective Action Capacity? Results from a Field Experiment in Post-Conflict Liberia'. Dataverse. Available at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/28006 (accessed 8 May 2017).

Gates, Melinda French (2014). 'Putting Women and Girls at the Center of Development'. *Science*, 345: 1273–75.

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 & Organization*, 70(1): 1–9.

Mansuri, Ghazala, and Vijayendra Rao (2012). *Localizing Development: Does Participation Work?* Washington, DC: The World Bank.

World Bank (2012). *World Development Report: Gender Equality and Development*. Washington, DC: World Bank.