International Gender Balancing Reforms in Post-Conflict Countries: Lab-in-the-Field Evidence from the Liberian National Police

Sabrina Karim, * Michael J. Gilligan, # Robert Blair † and Kyle Beardsley ‡

Abstract

In the aftermath of civil conflict, war-torn states often require government institutions to be reformed. Gender balancing, or the inclusion of more women in security-sector institutions, is an increasingly common reform incorporated into state building processes. Our theoretical priors suggest that gender balancing may have implications in terms of unit cohesion, operational effectiveness with respect to sexual and gender based violence (SGBV), and organizational gender norms. We study these propositions using laboratory experiments with police officers of the Liberian National Police (LNP). We randomly assigned the proportions of women and men in 102 groups of six LNP officers to observe their deliberative processes and group choices. In our experiment, adding more women increased unit cohesion, but we found no evidence to suggest that simply adding more women would increase group (or individual) sensitivity to SGBV. We also found that although there was an increase in participation and influence by women, male beliefs about women’s role in policing did not improve with the inclusion of women. As one of the first experimental studies to assess the effects of gender composition within the actual population of interest, our results shed light on how international interventions to address gender equality in post-conflict countries affect important outcomes related to security.

1 We would like to thank our enumerators from the Center for Applied Research and Training (CART), Kou Gbaintor-Johnson, Joshua Riggings, Abraham S. Kromah, then Deputy Inspector General of Police Operations, William K. Mulbah, then Deputy Commissioner of Police and Chief of Personnel, Chief Superintendent Karson Zubah, Superintendent G. William Forkpa, and Patrolman Adolphus H. Yah Jr. Additionally, we would like to thank the Folke Bernadotte Working Group on UNSC 1325 for their generous support and working group members for their feedback.

* Assistant Professor, Cornell University, Department of Government
# Professor, New York University, Department of Politics
† Assistant Professor, Brown University, Department of Political Science
‡ Associate Professor, Duke University, Department of Political Science
1. Introduction

Over the last fifteen years, issues related to gender power imbalances and the representation of women have risen to the forefront of both academic and policy debates about peacebuilding, state building and post-conflict reconciliation. Through the Women, Peace and Security (WPS) agenda, embodied in United Nations Security Council Resolution (UNSCR) 1325 and subsequent resolutions, the UN and other multinational institutions have formally recognized the differential impact of armed conflict on men and women, and have called for greater participation of women in political institutions and domestic security forces as part of the reconstruction process (Aoláin and Haynes 2011; Olsson and Gizelis 2015; Anderson 2016). The United Nations and the United States have repeatedly identified progress in addressing gender inequalities as fundamental for state building success (Hudson and Leidl 2015). Avoiding manifestations of gender inequality in post-conflict societies misses a key source of perceived injustice and insecurity across the population. Moreover, misapplying reforms to address gender power imbalances can exacerbate social tensions.

Of the reforms that have proliferated through the WPS agenda, perhaps none is so common—or so controversial—as gender balancing. Broadly defined, gender balancing is a set of policies designed to increase the number of women in traditionally masculine institutions through quotas or other policy changes (Tripp and Kang 2008). Gender balancing in security institutions relates to a broader global norm stressing women’s inclusion in both domestic and international organizations (Paxton, Hughes, and Green

---

2 For example, the passing of nine U.N Security Council Resolutions on WPS, as well as the “Hillary Doctrine” and the Swedish Feminist Foreign Policy, among others (Hudson and Leidl 2015).
While the normative basis for gender balancing in the security sector is well developed in the literature, the specific implications for individual and group behavior within security institutions are underexplored.

We consider arguments positing that gender balancing in security organizations improves certain processes including: unit cohesion, operational effectiveness with respect to sexual and gender based violence (SGBV), and organizational receptiveness to empower marginalized contributors. Proponents of gender balancing reforms argue that they can help groups become more consensus-based and collaborative; that they can improve the provision of security for women, who tend to be underserved by the security sector and who often face unique challenges in dealing with violence, especially sexual violence; and that they can improve the receptiveness of male group members to the goal of gender equality.

Whether gender balancing fulfills these three objectives remains an open question. Critics have argued that gender balancing policies erode unit cohesion and morale, especially within security institutions,\(^3\) leading to backlash against women; that they are merely window dressing, drawing attention away from the deeper, more urgent challenges that women face in war-torn societies; or, that the very intuition underlying these policies springs from an essentialist understanding of women’s roles, thus entrenching \textit{de facto} gender inequality for the sake of superficial \textit{de jure} reform.

Does the inclusion of women in historically masculine institutions improve unit cohesion, enhance operational effectiveness with regards to “gendered”\(^4\) security, and/or

\(^3\) Witness, for example, the recent debate over integrating women into combat roles in the American military (Schaefer et al. 2015, 17-27).

\(^4\) We affirm that gender is a social construction and that gender is not synonymous with women. We refer to gender balancing as the UN uses it: as an effort to include more women in institutions that are dominated by
increase receptiveness to change in the organizational culture, as proponents argue it should? We test these possibilities using a novel laboratory experiment with a sample of 612 officers from the Liberian National Police (LNP). The officers were grouped into 102 teams that randomly varied in their gender composition (no women, 1/3 women, 2/3 women or all women) and then were asked to complete a series of cooperative tasks. We assess the effects of sex composition on the collegiality and inclusiveness of team decision-making processes, sensitivity to sexual and gender-based-violence (SGBV), and attitudes about women’s roles on the police force.

To our knowledge, this is one of the first experimental studies to assess the effects of sex composition within the actual population of interest.\(^5\) Over thirty years of work in social-psychology has sought to understand how gender balancing affects group dynamics in a lab setting (e.g., Chrisler and McCreary 2010; Dindia and Allen 1992; Garrison et al. 1992; Smith-Lovin and Brody 1989; Vugt, Cremer, and Janssen 2007; Kushell and Newton 1986; Carli 2001; Carli and Eagly 1999; Patterson and Schaeffer 1977), and more recent lab experiments such as Karpowitz et al. (2012) have shown that sex composition does affect interpersonal interactions. These studies, however, have typically relied on samples of undergraduates or randomly-selected citizens. Extrapolating from the lab to the real world is especially problematic when the populations of interest (e.g., police, politicians) are so different from the population at large. Our use of actual Liberian police officers minimizes this problem of external validity.

---

\(^5\) One exception is Finseraaas et al. (2016), the findings for which are mentioned below. However, they do not look at the effect of gender balancing on unit cohesion and operational effectiveness.
Additionally, many lab and lab-in-the-field experiments rely on activities, such as public goods games and ultimatum games,\(^6\) that may be too stylized to generate lessons about real-world behaviors. In addition to some these games, which are useful as points of comparison, we use new activities designed in collaboration with UN Police trainers to mimic the types of tasks that LNP officers might actually be instructed to complete as part of their training. Moreover, in Liberia, as in many countries with centralized police forces, officers are temporarily assigned to different units in different locations throughout the country. Observing short-term interactions among ad hoc groupings of police officers with varying sex compositions is thus directly relevant to the realities of policing in these settings.

In what follows, we locate the paper’s contribution within the existing and emerging scholarship on the international WPS agenda and gender balancing. We then highlight theoretical priors about how gender balancing might affect unit cohesion, operational effectiveness with regards to SGBV, and organizational gender norms. We then describe our lab-in-the-field experiment with the LNP and present our findings. We conclude with a discussion of implications.

2. THE WOMEN, PEACE AND SECURITY AGENDA AND GENDER BALANCING IN THE SECURITY SECTOR

Feminist scholars have long argued that gender is an important factor in international relations, yet gendered institutions and the role of women have been ignored in

\(^{6}\) These games typically involve the possibility of a participant to gain a small amount of money, depending on their constrained choices in responding to artificial scenarios constructed by the researchers.
international relations theorizing (Enloe 2000; Tickner 1992; Pettman 1996).\(^7\) One key observation from this work is the notable absence of women in the security sector (Goldstein 2003).

Since the beginning of the WPS agenda in 2000, however, gender balancing has become a central component of peacekeeping and peacebuilding operations (DCAF 2011), which have increasingly adopted gendered approaches to rebuilding domestic institutions, especially in the security sector (Kronsell 2012; Mobekk 2010, Bastick 2008).\(^8\) These reforms are intended to contribute to a number of objectives such as attenuating sexual violence, preventing civilian casualties, increasing trust in institutions, and generally increasing the chances for a durable peace (Karim and Beardsley 2017). Advocates especially have called for reforming male-dominated institutions like the security sector such that they become more representative of the population as a means to promote good governance and instill trust in the population (Mobekk 2010; Bastick 2008; Sedra 2010; Ball et al. 2002). Drawing an analogy to discussions related to nurturing peace in ethnically divided societies, inclusive representation in the security sector may help minority groups feel more comfortable accessing the state’s security services (Hasisi and Weitzer 2007; Weitzer and Hasisi 2008). The same may be true for gender balancing reforms, which some scholars believe can increase civilian trust in state institutions by making them more representative of society at large and accessible to the marginalized (Keiser et al. 2002; Meier and Nicholson-Crotty 2006, Karim 2017a).\(^9\)

---

\(^7\) For an overview of one strand of the gender and conflict literature and debate, see Reiter (2015) and Sjoberg et al. (2016).

\(^8\) Efforts to improve the representation of women in the security sector parallel efforts to improve the gender balance in politics more generally (Tripp and Kang 2008).

\(^9\) For example, the deployment of all female-formed police units to peacekeeping missions is an innovation designed to help legitimize the role of peacekeepers in the eyes of the public, particularly women (Pruitt 2016).
The UN specifically has tied increased women’s representation to better outcomes related to peace and security.\textsuperscript{10} Consistent with this claim, the UN has been instrumental in promoting gender balancing reforms as part of peacebuilding efforts. For example, the UN Security Council has endorsed gender balancing (among other approaches to women’s empowerment) in nine resolutions. Huber and Karim (2017) estimate that the presence of a UN peacekeeping operation increases the probability of a post-conflict country adopting gender balancing reforms in the security sector by 22%. These trends, combined with the push to include more women in politics, have led some to suggest that a new global norm on women’s inclusion has emerged (Paxton, Hughes, and Green 2006; Bush 2011).

Gender balancing has led to a marked change in the composition of security sector institutions around the world. For example, the Sierra Leone police developed family support units and recruited women to fill these posts; the Policía Nacional de Timor-Leste (PNTL) required a 20% quota for women, and the Kosovo Police Service (KPS) targeted women and ethnic minorities in recruitment (DCAF 2011).\textsuperscript{11} These changes are particularly relevant for police organizations, as research has found that gender is an extremely salient external status characteristic in policing, one that is noticeable to all members of the force (Martin 1990; 1999; Gerber 1996).

We chose to study gender balancing in the LNP because it has been such a focus of international peacebuilding efforts and because it reflects common challenges that are likely to arise during gender reforms in post-conflict countries. The Second Liberian Civil

\textsuperscript{10} See, for example, the UN’s webpage on “Women in Peacekeeping,” http://www.un.org/en/peacekeeping/issues/women/womeninpk.shtml (September 11, 2017).

\textsuperscript{11} Other countries in the report where gender has been a major part of SSR include Hungary, South Africa, the UK, Central African Republic, Indonesia, Peru, Somalia, Afghanistan, Russia, Tajikistan, Rwanda, Ivory Coast, Israel, Jamaica, Brazil, and Nepal (DCAF 2011).
War ended in 2003, and Liberia became an early test case for the WPS agenda. The UN Mission in Liberia (UNMIL), a typical multidimensional peacekeeping force, deployed in 2003 and has placed a heavy emphasis on security sector reform and national police training. The gender reforms enacted by the LNP are consistent with similar reforms being pursued in many security forces worldwide with assistance from peace operations. Our results specifically relate to other weak, war-torn states where gender reforms are occurring with international assistance, such as the DRC, Haiti, Kosovo, Bosnia, and others.

The evolution of gender reforms in the LNP offers a glimpse into the challenges likely facing similar security institutions worldwide. Initial implementation of gender balancing policies in the LNP struggled due to the number of women who did not have the necessary qualifications, such as secondary-school completion. As such, specialized training programs were put in place to speed up the process. At first, this led to some backlash against the policy as men within the police force believed that women were less competent than they were. Another stark challenge relates to the retention and promotion of the women that were recruited. Like in other police forces and more broadly in other security organizations, women in the LNP are more likely to serve in an administrative capacity, as demonstrated by the breakdown of men and women in different positions in the appendix.

The international community has played a major role in the LNP’s gender reforms. In 2005, the Norwegian and Danish governments and international organizations, including UNMIL, helped create the Women and Children Protection Unit in the LNP (Bacon 2011). This unit was established to handle crimes committed against women and children, such as

---

12 Our results discussed below, however, do not find that women are less competent than men.
child abuse and abandonment, sexual assault, corruption of minors, and other criminal
offences. UNMIL and UN Women also helped the LNP develop a special Gender Unit,
which is responsible for gender sensitization of LNP personnel through training at the LNP
academy and additional gender training sessions (Bacon 2011). One of UNMIL’s first tasks
was to develop a Gender Policy for the LNP, which promoted gender balancing and was
the first such policy in any UN Mission. In 2008, the UN helped the LNP introduce a 15-
percent gender-balance target, followed by a 20-percent target, and then a 30-percent
target. By 2015, the LNP included about 19 percent female officers, up from just two
percent in 2005 (Bacon 2012).

3. GENDER BALANCING AND IMPROVEMENTS IN PROCESSES
RELATED TO STATE BUILDING

In the aftermath of civil conflict, war-torn states often require government institutions to
be (re)-built or at the very least reformed (Lake 2016). To the extent that rebuilt and
reformed state institutions are perceived as effective and legitimate, underlying grievances
against and mistrust of the state diminish, and enduring peace becomes more likely. If
state institutions are widely perceived in a positive way, then not only might peace hold,
but the quality of peace strengthens (Wallensteen 2015). We assume that improving
personnel, improving relationships among personnel, and improving the culture of
organizations are important for the efficacy and legitimacy of state institutions (Peters

13 See “The Liberian National Police’s female recruitment programme,” Geneva Centre for the
Democratic Control of Armed Forces (DCAF),
14 UNMIL, Office of Gender Adviser, 2010. Gender Mainstreaming in Peacekeeping Operations Liberia
In security organizations, these transformations become particularly salient, as the provision of security is dependent on how well groups work together to achieve a task (*unit cohesion*) (MacCoun et al. 2006), on how groups achieve tactical objectives (*operational effectiveness*) (Millett and Murray 2010), and on norms that welcome contributions from all able contributors (*egalitarian norms*) (Karim and Beardsley 2017).

### 3.1 Gender Balancing and Unit Cohesion

Unit cohesion is imperative for the operational effectiveness of security forces of all kinds (Castillo 2014, Staniland 2014, MacCoun et al. 2006, Shils and Janowitz 1948, Jordan et al 2002). There has been a long debate about conceptualizing and measuring unit cohesion (Siebold 1990). Shils and Janowitz (1948) define it as “primary group solidarity,” and Castillo (2014: xi) argues that it “depends on the bonds of loyalty that soldiers form within small units, rather than a grand political cause.” MacCoun et al. (2006: 647) suggest that unit cohesion is composed of task cohesion and social cohesion, where “social cohesion refers to whether group members like each other and task cohesion refers to whether they share the same goals.” They argue that task cohesion is more important for effectiveness.

Borrowing from these different conceptions of cohesion, we focus on the degree to which groups with more women share the same goals as ones with fewer or no women. In particular, this means assessing *whether individual preferences align with group preferences as a measure of task cohesion*. In addition, we consider how group gender balance shapes the *level of collegiality within groups, as an indicator of social cohesion*. 
Starting with task cohesion, preference alignment is important to prevent dissent within groups. Some perspectives suggest that the gender balance of groups can improve task cohesion due to how women make decisions and influence decision-making. Research in American politics has concluded that women in political settings are more consensus-oriented (Hawkesworth 2005: 145), which may influence how decisions are made for both women and men in the group (Jewell and Whicker 1994; Rosenthal 2002; Thomas 1994). Other work posits that women possess superior interpersonal skills that enable them to be more understanding of others, more patient, and more skilled at effecting compromise (Flammang 1985; Reingold 2000). These skills would be especially important if there are divergent beliefs, and could help build consensus when intra-group conflicts arise.

The gender balance of groups might also improve social cohesion. In the policing literature, some evidence suggests that women behave in a more “humanistic way,” are better communicators, are more empathetic, and have a calming presence in groups (Garcia 2003; Schuck 2014; Schuck and Rabe-Hemp 2007; McDowell 1992). These interpersonal skills and communicative abilities are key to making groups more socially cohesive (Maznevski 1994). Both task cohesion (preference alignment) and social cohesion (collegiality) thus might improve as the representation of women increases. According to these arguments, women can introduce collaborative methods to group decision-making,

---

15 For example, the principal-agent literature on the military emphasizes the importance of preference alignment as a way to ensure that soldiers do not shirk (Feaver 2001, Brehm and Gates 1993).

16 It is important to recognize that a focus on correlations between gender and behavioral characteristics risks essentializing women so that women become means to ends and surrogates for those behavioral characteristics even though there is tremendous heterogeneity among women. While we caution that such essentialist claims risk, among other things, entrenching gender stereotypes that are the basis for gender power imbalances, we subject some of these claims to empirical testing because of the reality that they are the basis for many gender balancing and mainstreaming efforts.
improve collegiality, bring more people’s beliefs to the table, and help ensure that group decisions are reflective of most people’s preferences.

**HYPOTHESIS 1**: Groups that contain more women will be more cohesive as a unit in terms of preference alignment and collegiality.

Failure to confirm this hypothesis would support countervailing arguments suggesting that the inclusion of women into new police units could have a negative effect on cohesion. At face value, women’s integration could mean that new preferences are brought into a rather homogenous organization. Indeed, some scholarship posits that ascriptive differences are associated with divergences in preferences (Huddy and Terkildsen 1993, Byrnes et al 1999). Diversity could cause preference heterogeneity and therefore more group discord. Opponents to the integration of soldiers from different ethnicities, race, and gender have made arguments along these lines (Goldich and Webb 2010, Brown and Ayres 2004). Thus, it is possible that preferences become more divergent if women are incorporated into groups (Van Creveld 2001 Fenner and deYoung 2001; Maginnis 2013; Gutmann 2000).

Regarding social cohesion, some have posited that as women are integrated into groups, men may become even more antagonistic toward them (Carli and Eagly 1999; Remmington 1983; Shackelford, Wood, and Worchel 1996; Pugh and Wahrman 1983). When men are outnumbered by women they may redouble their efforts to assert dominance because they feel alienated or believe the presence of women will undermine the

---

17 See also Haslett et al (1999). Related literature has focused on assessing the effect of ethnic integration in the military (Samii 2013).
cohesiveness of their traditionally male environment (Yoder 1991). Some studies also suggest that the presence of women causes men to assert their masculinity more forcefully (Smith-Lovin and Brody 1989). This disrupts group dynamics as participants become fixated on gender roles (Swan and Wyer Jr. 1997).

3.2 Gender Balancing and Operational Effectiveness in Responding to SGBV

While operational effectiveness typically focuses on the achievement of “strategic objectives within the theatre of war (Millett and Murray 2010: 12),” we relax the definition and assess the success groups have in achieving a particular outcome, in this case, detecting sexual violence. With the unanimous passage of UNSCR 1325 and the development of the international WPS agenda, security organizations have received new mandates to make gender sensitivity—especially sensitivity to sexual and gender-based violence (SGBV)—a priority area for civilian protection and security. While some studies have considered how women’s integration affects overall policing efficacy (Karim 2017a), earlier studies have not assessed how women might respond to gendered crimes, a particularly pervasive source of human insecurity in many environments with recent armed conflict.

Some advocates of gender balancing argue that women are more sensitive than men to women’s needs, especially SGBV, and that their presence will improve responsiveness to issues that disproportionately affect women. The UN itself has made such claims in order to motivate gender balancing in peacekeeping operations: women’s presence “makes male peacekeepers more reflective and responsible; and it broadens the repertoire of skills and styles available within the mission” (Rehn and Sirleaf 2002: 63). Additionally, UNSC 2106

---

18 Other scholars have found that sex composition does not affect team performance (Myaskovsky et al 2005).
(2013) emphasizes that women can exert “influence over parties to armed conflict with respect to addressing sexual violence.”

Gender balancing aims to address SGBV through several proposed mechanisms: female victims may feel more comfortable reporting gender-based crimes to female officers; female officers may be better equipped to interact with female victims; or female officers may be more likely to detect gender-based crimes in the first place, especially when evidence is ambiguous. Studies in the American and bureaucratic politics literatures argue that women pay particular attention to women’s issues (Keiser et al. 2002; Meier and Nicholson-Crotty 2006). Consciously or subconsciously, bureaucrats translate shared demographics (such as sex) into shared values, and shared values into greater commitment to improving the condition of members of their own groups (Meier and Nicholson-Crotty 2006). In psychology, gender scheme theory predicts that men and women process information differently—a phenomenon known as “sex typing” (Bem 1981). In the context of policing, the above logic points to the expectation that female officers will have a greater awareness of SGBV.

It is also possible that the presence of women will prime all members of the group—both men and women—to be more cognizant of SGBV. In the psychology literature, scholars have found that when individuals are exposed to non-traditional gender stereotypes (for example, female police officers or female mechanics), they tend to display greater acceptance of various social groups and increased cognitive flexibility (Prati et al. 2015). Individual group members may be influenced by the presence of women (Barnello and Bratton 2007), making them more sensitive to gendered issues, and encouraging them to think about gender when they might not have otherwise (Williams and Heikes 1993).
HYPOTHESIS 2: Groups and members of groups that contain more women will improve operational effectiveness with regards to sensitivity to SGBV.

There are, however, alternative reasons to believe that women will not necessarily be more sensitive to women’s needs, even SGBV. Young (1994) and Reingold (2000) have noted that there is great heterogeneity in the category “women,” and that women may not be inclined to consciously or unconsciously support outcomes that favor women. Evidence also suggests that gender scheme theory is inaccurate in predicting differences in how men and women interpret information. Research in psychology finds that gender scheme theory too narrowly conceptualizes how individuals draw conclusions from information, without considering social contexts and other factors (Spence 1993; Deaux and Major 1987). Research in marketing also finds that gender does not correlate strongly with the way individuals process and interpret information (Schmitt, Leclerc, and Dubé-Riou 1988; Deaux 1984). If women are no more likely than men to process information such that they see certain environments, issues, or contexts as gendered, then we would not expect much difference in whether women or men are more sensitive to SGBV.

3.3 Gender Balancing and Organizational Gender Norms

Finally, women’s integration may change organizational gender norms in traditionally masculine environments. Military and police organizations can be highly discriminatory towards women, and may prevent women from participating to the full range of their
capabilities and from making influential decisions (Eagly and Makhijani 1992, Karim and Beardsley 2017, Karim 2017b). Increasing women’s presence in these organizations may catalyze larger changes in institutional culture and norms. Recent research has found that interaction with “outsider groups,” in this case women, has the potential to change perceptions of those groups in the long term (Brockman and Kalla 2016). In this version of the “contact hypothesis,” interactions with women could instill more equitable beliefs about women or even change norms about women’s participation and influence in groups. A recent study in Norway found that male soldiers who were randomly assigned to be roommates and work in squads with female soldiers during the recruitment period were less likely to discriminate against women (Finseraas et al. 2016).

In addition to changing beliefs about women, reforms that improve the representation of women might make the environment friendlier for women to participate and influence decision making. Many studies have found that women are quieter and less likely to participate in group activities when they are outnumbered by men (Dindia and Allen 1992; Karpowitz and Mendelberg 2007; McCarrick, Manderscheid, and Silbergeld 1981; Niederle and Vesterlund 2007; Smith-Lovin and Brody 1989). As a result, women may feel more comfortable or more valued in the context of all-female or majority-female groups (Eagly and Johnson 1990; Garcia 2003). Kanter (1977) argues that once a threshold (usually 30 percent) of women’s participation is achieved, women have an opportunity to make a difference in policy formation.

Indeed, gender balancing policies are predicated on the assumption that women will be able to influence group decisions and that women’s views will be better represented when more women are included in the decision making processes (Olsson and Truggestad
With the presence of more women in a group, the group’s decisions may become more reflective of women’s views.

**HYPOTHESIS 3:** Members of groups that contain more women will see a shift in gender norms with regards to beliefs about women and women’s participation and influence.

Other perspectives call Hypothesis 3 into question. It is possible that gender norms will become more rigid as a result of women’s integration into a traditionally masculine space. Yoder (1991) argues that when male-dominated organizations become more gender-diverse, an “intrusiveness effect” occurs, whereby discrimination becomes more marked, and the organizations become more resistant to the inclusion of more women. Backlash against women can come in different forms, including subtle punishments, assertion of dominance, discrimination, and even violence (Burk and Major 2014, Beaton and Tougas 1997; Rudman and Glick 2001). This suggests that gender norms might move in the other direction—making policemen less likely to accept gender reforms, more likely to discriminate, and more likely to overpower women.

Moreover, some studies suggest that mixed-gender groups do not do well in incorporating the perspectives of women (Carli 2001; Remmington 1983; Shackelford, Wood, and Worchel 1996; Pugh and Wahrman 1983). Some scholars have argued that external status characteristics such as gender may affect perceptions of competency in mixed-gender groups—when a man and woman work together, the man is assumed to have higher status and greater competence (Berger, Rosenholtz, and Zelditch 1980; Lockheed
and Hall 1976; Gerber 1996), and therefore his preferences are adopted by the group. Thus, it is possible that women will continue to have little influence in groups that include men, even as the numbers of women increase.

4. **Research Design**

To test these hypotheses, we conducted a lab-in-the-field experiment with LNP officers in Monrovia, Liberia in January of 2013. The experiment consisted of a set of team activities designed to assess unit cohesion, operational effectiveness with regards to SGBV, and changes in organizational gender norms. We randomly selected officers from available personnel in Montserrado County (where Monrovia is located), oversampling women to compensate for the smaller number of women in the LNP overall. (See the appendix for a fuller description of our sampling procedures.) We conducted the activities in groups of six officers, with the treatment being the sex composition of the group: groups were randomly assigned to include zero, two, four or six women. There were 33 all-male groups, 29 two-woman groups, 21 four-woman groups, and 19 all-woman groups. We implemented the activities with six groups of six officers simultaneously. Each of these six groups was placed in a separate room to complete the activities. We randomly assigned teams of trained Liberian enumerators to the groups. The enumerators facilitated the implementation of the team activities and recorded various aspects of the team decision-making processes (described in more detail below). In addition to the team activities, the enumerators also

---

19 We presented the research design at the” Experiments in IR Conference,” Park City, Utah, September 21-22, 2012. With theoretical priors that point to possible relationships in competing directions, we use two-tailed tests of statistical significance.
administered background and exit questionnaires, including a survey experiment designed to assess discrimination against women in training with firearms.

Table 1 compares the women in our study to the men and assesses covariate balance across a number of characteristics. *Age range* is an ordinal scale of the age of the participant; *Education* is an ordinal scale of the level of education; *Tenure* is the length of time in the LNP in months; *Rank* is an ordinal scale of rank; *Friends* is a count of the number of co-group members that the respondent considers to be a friend; and *Cognitive Score* is the proportion of questions correctly answered on a test of basic reasoning and mathematical skills. Columns two and three display the mean values for women and men, respectively, and the fourth column gives the difference in means and *p*-value from a *t*-test. The next three columns disaggregate the women into those in homogenous groups and those in mixed groups, with *p*-values from *t*-tests of the difference of means between those two groups. The last three columns similarly disaggregate the men. There are several differences between women and men in their underlying characteristics, but, for the most part, the randomization of the individuals to the groups worked. On all variables except for *Friends*, women in the homogenous groups are similar to women in the mixed groups.20 The men in the homogenous groups are similar to the men in the mixed groups on all variables. We still include these variables as controls in our regression models to correct for any incidental unbalance.

---

20 The difference in the *Friends* variable may be a product of unfortunate randomization, but, also plausible is that women are more likely to be friends with other women and then placed in groups with more women. The effect on balance is stronger for women than men because we oversampled women.
Table 1: Mean Characteristics of LNP Participants

<table>
<thead>
<tr>
<th></th>
<th>All Women</th>
<th>All Men</th>
<th>p-value</th>
<th>Women in all-women groups</th>
<th>Women in mixed groups</th>
<th>p-value</th>
<th>Men in all-men groups</th>
<th>Men in mixed groups</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age range</td>
<td>2.48</td>
<td>2.89</td>
<td>&lt;0.001</td>
<td>2.47</td>
<td>2.49</td>
<td>0.871</td>
<td>2.91</td>
<td>2.87</td>
<td>0.625</td>
</tr>
<tr>
<td>Education</td>
<td>5.33</td>
<td>5.63</td>
<td>&lt;0.001</td>
<td>5.40</td>
<td>5.26</td>
<td>0.196</td>
<td>5.65</td>
<td>5.60</td>
<td>0.541</td>
</tr>
<tr>
<td>Tenure</td>
<td>59.1</td>
<td>82.3</td>
<td>&lt;0.001</td>
<td>56.5</td>
<td>61.2</td>
<td>0.288</td>
<td>82.92</td>
<td>81.59</td>
<td>0.846</td>
</tr>
<tr>
<td>Rank</td>
<td>1.51</td>
<td>1.90</td>
<td>0.001</td>
<td>1.39</td>
<td>1.62</td>
<td>0.156</td>
<td>2.01</td>
<td>1.77</td>
<td>0.161</td>
</tr>
<tr>
<td>Friends</td>
<td>0.84</td>
<td>0.90</td>
<td>0.53</td>
<td>1.04</td>
<td>0.681</td>
<td>0.010</td>
<td>0.980</td>
<td>0.799</td>
<td>0.155</td>
</tr>
<tr>
<td>Cog. Score</td>
<td>0.4</td>
<td>0.51</td>
<td>0.016</td>
<td>0.377</td>
<td>0.418</td>
<td>0.548</td>
<td>0.530</td>
<td>0.484</td>
<td>0.440</td>
</tr>
</tbody>
</table>

Note: Values are from difference-of-means *t*-tests with equal variances.

4.1 Experiments

In the lab, teams of six LNP officers completed five team tasks. Some of the tasks were designed to measure deliberation in general, an integral part of teamwork, while others were more specific to policing. In the first stylized activity, each group was given 300 Liberian dollars (LD)\(^{21}\) and asked to decide collectively how much, if anything, to place at risk in a lottery that was decided by the roll of a die.\(^{22}\) The group was instructed that the total pot, including any winnings and minus any losses, would be divided equally among the six of them. Before deliberation, each officer secretly wrote his or her preferred bet on a slip of paper, which we recorded anonymously. Groups were given five minutes to discuss the group bet. At the end of the five minutes the group had to come to a decision using whatever means they chose (e.g., consensus, voting, individual fiat). The goal of the activity was to observe how the officers made a group decision when they had different and conflicting individual preferences over risk, and not necessarily to learn about their attitudes toward risk per se.

---

\(^{21}\) This is approximately $3.81. We scaled the payoffs based on the daily pay for the LNP. The officers on average make 130LD per day, so they could make at least a day’s salary in the time it took them to complete all the exercises.

\(^{22}\) The outcome of the lottery was decided as follows: if a one was rolled the group would lose all its bet; if a two was rolled the group would lose half its bet; with a three the group would lose nothing; a roll of four won the group 1½ time its bet; a roll of five twice its bet; and a roll of 6, 2 ½ times its bet.
In the stylized second activity, which we call the “pay-it-forward” activity, the team was again given 300 LD and asked to decide as a group how much they wanted to give to one of the other five groups in the session at the same time with which they were anonymously paired. We doubled the amount that they donated to the other team. We gave the teams five minutes to deliberate and make their decision, but before deliberations began each officer recorded his or her preferred donation on a slip of paper, which we anonymously recorded. Again, we were less interested in how much the groups contributed, and more in how they deliberated despite conflicting preferences over the donation.

We also conducted two cooperative activities in which teams pursued a common goal. The first of these activities required that each group build a free-standing tower using a single sheet of newsprint and a meter of masking tape. The group that built the largest tower would receive a payout of 600 LD, to be split evenly among the members of the group. In the second cooperative activity, officers were given a photograph of a hypothetical crime scene and instructed to memorize as many details about the scene as they could. They then secretly and anonymously answered a set of questions about the photo, and were individually awarded 10 LD for each correct answer. After answering the questions individually, the group reconvened to reach consensus answers to the same questions. The group was awarded 60 LD to be split evenly among them for each correct answer.

For the fifth activity, we tested how attuned the group was to the possibility of SGBV. The photograph used in the hypothetical crime scene activity contained a number of ambiguous clues suggestive of several possible crimes: burglary, physical assault,
murder, domestic violence or rape. We tested the conjecture that women should be more likely than men to see SGBV by asking the subjects a series of questions about what crime(s) they thought the photo depicted. We also asked them to report what evidence led them to their conclusion, and what they would do upon arrival at the crime scene. Again, we first asked the officers to answer these questions privately, then to conduct a group discussion to come up with a consensus “crime report” based on the evidence in the photo. We collected these reports and entered them into a spreadsheet.

During the group deliberations in each activity, three enumerators recorded who spoke, who argued, how aggressive people were, and, in the case of the tower activity, who physically touched the materials in an attempt to build the tower. Finally, we also used an exit questionnaire to survey how the participants felt about the team activities and to assess their views on women’s role in the LNP. Some questions asked the participants to report their comfort and satisfaction levels within the group environment. Other questions asked them to evaluate whether women and men are equally capable and valuable assets in the LNP.

4.2 Outcomes

We estimate average treatment effects across a variety of indicators, listed in Tables 2 and 3. Different indicators are used to measure the outcomes for the above-mentioned hypotheses, and we use an average-effects estimator to assess averaged and standardized treatment effects across the different measures for each outcome. The exception is for Hypothesis 2, the test for operational effectiveness, where we use a single dichotomous

---

23 An UNMIL peacekeeper and our Liberian research assistants staged the crime scene photographs.
24 We use the same estimator as Casey et. al. 2013. Also see Kling et al. (2004) Clingingsmith et al. (2009).
measure of whether the officers reported a gender-based crime as possibly committed in the photo.

The outcomes pertinent to Hypothesis 1, which assesses the effects of women’s integration on unit cohesion, are indicators of preference alignment and collegiality. We use six measures. The first two pertain to the differences between the individual’s stated preferences on the group bet and pay-it-forward activities, respectively, divided by the maximum possible differences. In other words, we calculated the difference between each individual’s preferred amount and the group amount, then divided that number by the maximum possible amount the group’s decision could have been different from the individual’s preference, and then subtracted that ratio from 1.\textsuperscript{25} The resulting number ranges between 0 and 1 and is increasing in the similarity of the group outcome to each individual’s preferred amount. The next three measures are the proportions of group responses to the memory, suspected crime, and evidence questions that match the individual’s responses. These vary from 0—no individual’s responses were reflected in the group responses—to 1—all of the individual’s responses were reflected. The final measure is a dichotomous variable of whether the individual’s response to what procedure should be followed upon arrival at the crime scene matched that of the group.

As a part of unit cohesion, we also measure the level of collegiality. To do this, we used three measures, each of them coded by the enumerators: the number of times each person talked and was not heard, the number of times each person argued, and an indicator for whether the group’s final decision was made by consensus. We also used five responses

\textsuperscript{25} Specifically, the similarity score is \(1 - \frac{|\text{group decision} - \text{preferred amount}|}{\text{300} - \text{preferred amount}}\) if the preferred amount was less than or equal to 150 LD and \(1 - \frac{|\text{group decision} - \text{preferred amount}|}{\text{preferred amount} - 6}\) if the preferred amount was greater than 150 LD.
to questions in the post-experiment survey that captured how the individual felt about their group decision-making experience. These were dichotomous variables indicating (1) whether the person did not feel comfortable participating in the group, (2) whether the person felt other group members did not listen to them, (3) whether the person felt unhappy with the group discussion, (4) whether the person felt unhappy with the group decision, and (5) whether the person would not want to be part of the same group again.

To test Hypothesis 2, about operational effectiveness with respect to awareness of SGBV, we assess the teams’ and individuals’ likelihood of suspecting that the crime depicted in the crime-scene photo was a case of SGBV. We use a single variable for both outcomes—a dichotomous variable, indicating whether teams or individuals believed the picture depicted rape or domestic violence.

Finally, to test Hypothesis 3 on changes in gender norms, we first assess attitudes regarding gender equality, followed by women’s participation and influence. To measure attitudes, we use six measures taken from the post-experiment survey. These are all dichotomous variables indicating whether respondents believed that the LNP should have a gender quota, that women make good leaders, that women are as good as (or better than) men in their jobs as police officers, that women are as good as (or better than) men in handling rape cases, that women are as good as (or better than) men in handling political violence and riots, and that the public trusts women as much as (or more than) men.

In addition to changes in beliefs, we also assess participation in group deliberations and activities. We measure participation using two sets of observations by the enumerators during the group discussions. The first is the number of times the subject talked, summed across all activities and averaged across the enumerators. The second is the number of
times the subject attempted to use the building materials in the tower building activity, again averaged across the enumerators.

We also assess the degree to which women were able to influence the group’s decisions. We use the same measures as those described for preference alignment, but we only look at the mixed groups. Specifically, we assess whether the preference of the women in the group better match the group’s overall decision when there were four women in the group versus when there were only two.

Table 2 provides a summary of the hypotheses and indicators associated with each hypothesis. Table 3 provides the descriptive statistics for these measures.
<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Indicators</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYPOTHESIS 1: Groups that contain more women will be more cohesive as a unit in terms of preference alignment and collegiality.</td>
<td>Preference alignment</td>
<td>Similarity of group bet to individual preferences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Similarity of group amount given in pay-it-forward activity to individual preferences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proportion of responses that match group’s memory activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proportion of suspected crimes that match the group’s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whether the response of proper procedure at the crime scene matches the group’s</td>
</tr>
<tr>
<td></td>
<td>Collegiality</td>
<td>Number of times talked and not heard (avg. across enumerators)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of times argued (avg. across enumerators)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group decision not made by consensus (group level)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Did not feel comfortable participating in group</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Felt people in group did not listen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not happy with group discussion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not happy with group decisions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Would not want to be in the same group again</td>
</tr>
<tr>
<td>HYPOTHESIS 2: Groups and members of groups that contain more women will improve operational effectiveness with regards to sensitivity to SGBV.</td>
<td>Sensitivity to SGBV</td>
<td>Suspect SGBV (rape or domestic violence) at crime scene</td>
</tr>
<tr>
<td>HYPOTHESIS 3: Members of groups that contain more women will see a shift in gender norms with regards to beliefs about women and women’s participation and influence</td>
<td>Beliefs about women</td>
<td>Agree with “Should the LNP have a required amount of women?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agree with “Are female LNP as good as males?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Answered “women” or “both” to “Who is better at handling rape and domestic violence?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Answered “women” or “both” to “Who is better at handling political violence and riots?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Answered “women” or “both” to “Who do you think the public trusts?”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agree with “Do female LNP make good unit leaders?”</td>
</tr>
<tr>
<td></td>
<td>Participation</td>
<td>Total number of times talked across all activities (avg. across enumerators)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total number of times touched building materials (avg. across enumerators)</td>
</tr>
<tr>
<td></td>
<td>Influence</td>
<td>Women’s similarity of group bet to individual preferences in mixed groups</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women’s similarity of group amount given in pay-it-forward activity to individual preferences in mixed groups</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women’s proportion of responses that match group’s memory activity in mixed groups</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women’s proportion of suspected crimes that match the group’s in mixed groups</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women’s proportion of suspected crimes’ evidence that match the group’s in mixed groups</td>
</tr>
<tr>
<td></td>
<td></td>
<td>For women, whether the response of proper procedure at the crime scene matches the group’s in mixed groups</td>
</tr>
</tbody>
</table>
Table 3: Outcomes for hypotheses and their measures.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unit Cohesion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarity of group bet to individual preferences</td>
<td>0.6665</td>
<td>0.3095</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Similarity of group amount given in pay-it-forward activity to individual preferences</td>
<td>0.7346</td>
<td>0.2791</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Proportion of responses that match group’s memory activity</td>
<td>0.6002</td>
<td>0.2141</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Proportion of suspected crimes that match the group’s</td>
<td>0.4111</td>
<td>0.4592</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Proportion of suspected crimes’ evidence that match the group’s</td>
<td>0.4406</td>
<td>0.3056</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Whether the response of proper procedure at the crime scene matches the group’s</td>
<td>0.4314</td>
<td>0.4957</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Number of times talked and not heard (avg. across enumerators)</td>
<td>2.0948</td>
<td>2.4228</td>
<td>0.0</td>
<td>21.0</td>
</tr>
<tr>
<td>Number of time argued (avg. across enumerators)</td>
<td>0.8268</td>
<td>1.3693</td>
<td>0.0</td>
<td>9.5</td>
</tr>
<tr>
<td>Group decision not made by consensus (group level)</td>
<td>0.1650</td>
<td>0.3715</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Did not feel comfortable participating in group</td>
<td>0.0147</td>
<td>0.1205</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Felt people in group did not listen</td>
<td>0.0441</td>
<td>0.2055</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Not happy with group discussion</td>
<td>0.0408</td>
<td>0.1981</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Not happy with group decisions</td>
<td>0.0474</td>
<td>0.2126</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Would not want to be in the same group again</td>
<td>0.1324</td>
<td>0.3392</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Operational Effectiveness with Respect to SGBV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspect SGBV at crime scene</td>
<td>0.4461</td>
<td>0.4975</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Changes in Organizational Gender Norms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of times talked across all activities (avg. across enumerators)</td>
<td>34.2500</td>
<td>19.7900</td>
<td>1.5</td>
<td>158.0</td>
</tr>
<tr>
<td>Total number of times touched building materials (avg. across enumerators)</td>
<td>9.9477</td>
<td>8.3507</td>
<td>0.0</td>
<td>65.0</td>
</tr>
<tr>
<td>Agree with “Should the LNP have a required amount of women?”</td>
<td>0.7451</td>
<td>0.4362</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Agree with “Do female LNP make good unit leaders?”</td>
<td>0.9052</td>
<td>0.2931</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Agree with “Are female LNP as good as males?”</td>
<td>0.8480</td>
<td>0.3593</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Answered “women” or “both” to “Who is better at handling rape and domestic violence?”</td>
<td>0.9379</td>
<td>0.2415</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Answered “women” or “both” to “Who is better at handling political violence and riots?”</td>
<td>0.4918</td>
<td>0.5003</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Answered “women” or “both” to “Who do you think the public trusts?”</td>
<td>0.8546</td>
<td>0.3528</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
4.3 Treatments and Controls

The treatment at the individual and group levels is the number of women on the team. The all-male teams comprise the control category. We randomized the assignment of LNP officers to each of these types of groups, but there was imbalance between men and women, and therefore between groups with more women and those with fewer. To address this imbalance, all individual-level models control for the participants’ sex, length of service in the LNP (in months), the number of friends they identified in the group, and a series of categorical variables for age range, education, rank, ethnicity, religion, and birthplace.26

The subjects’ competency as police officers is a potentially important confounding variable as well. The competence measure we use to distinguish high-competence from low-competence groups is a latent-variable index generated using a Bayesian scaling model, based on individual responses to the cognitive score questions, the memory test, and answers to the crime scene activity.27 The advantage of this measure is that it is based on observed assessments and not on characteristics such as rank and education, which may or may not capture how skilled or well-trained the individuals are, and may be susceptible to gender biases within the society and institution.

5. RESULTS

26 It is not uncommon for people in Liberia to not know their exact age so we use the age range variable.
27 We used three cognitive questions, six memory questions, and crime scene questions to create a latent competency score. We asked an UNMIL Police peacekeeper to post-code the crime scene questions based on the training the LNP receive in the academy. Crimes that matched a statutory crime were coded as correct. When the evidence matched the crime, each piece was coded as correct. If they correctly listed the immediate action upon arrival (per their training), their answer was coded as correct. Correct answers were always coded as a one, ensuring that the direction of the scale remained consistent. The specific latent scale model can be found in the appendix.
Finding 1: Unit Cohesion

Our results, presented in Table 4, provide some support for Hypothesis 1. From Model 1, we observe higher levels of preference alignment as the number of women in the group increases. The average effect for the all-female teams is statistically significant at the five-percent level. The all-women groups exhibited seventeen percent of a standard error deviation more similarity between ex ante individual views and their group’s final decisions. In teams with two women there was eight percent of a standard deviation more similarity between ex ante individual views and the group’s final decisions, and in teams with four women there was eight percent of a standard deviation more similarity between ex ante individual views and the group’s final decisions, but neither of these results is statistically significant. Note that the average effects are monotonically increasing in the number of women in the group: the more women in the group, the more task cohesive the group is, although not always at statistically significant levels.

The men in all-male teams are the residual category so the fact that all average effects estimates are positive confirms that preserving a homogeneous group of men is not necessary to maximize preference alignment. Our experiment suggests that including women may be more important than group homogeneity for preference alignment. In comparing the two homogenous groups to one another (all-men to all-women), the all-female group was more aligned in their preferences than the all-male group. Moreover, in comparing the mixed groups to the all-men groups, we do not observe a decrease in preference alignment.

Women were in general more risk averse and less willing to contribute to other groups, so groups that contained more women possessed lower variance in the ex-ante
preferences on these variables. To address potential confounding, we include in this specification the group standard deviation of individual risk preferences, the individual desired amount to give in the pay-it-forward activity, and the number of correct memory activity answers. Only the standard deviation of the members’ ex ante risk preferences was a significant predictor of cohesion.

Table 4: Effects of Number of Women on Group Processes and Decisions

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preference Alignment</td>
<td>Non-Collegiality</td>
<td>Group suspected SGBV</td>
<td>Members suspected SGBV</td>
<td>Women’s role in the LNP</td>
<td>Participation</td>
</tr>
<tr>
<td>2 women</td>
<td>0.0786</td>
<td>-0.0656</td>
<td>0.328</td>
<td>-0.141</td>
<td>-0.0441</td>
<td>0.0797</td>
</tr>
<tr>
<td></td>
<td>0.0469</td>
<td>0.0564</td>
<td>0.815</td>
<td>0.386</td>
<td>0.0663</td>
<td>0.155</td>
</tr>
<tr>
<td>4 women</td>
<td>0.0809</td>
<td>0.0799</td>
<td>-0.943</td>
<td>0.327</td>
<td>0.0228</td>
<td>0.385</td>
</tr>
<tr>
<td></td>
<td>0.059</td>
<td>0.0773</td>
<td>0.776</td>
<td>0.411</td>
<td>0.0766</td>
<td>0.208</td>
</tr>
<tr>
<td>6 women</td>
<td>0.171*</td>
<td>0.132</td>
<td>0.839</td>
<td>0.303</td>
<td>0.00581</td>
<td>0.478*</td>
</tr>
<tr>
<td></td>
<td>0.0775</td>
<td>0.0767</td>
<td>1.331</td>
<td>0.432</td>
<td>0.082</td>
<td>0.197</td>
</tr>
<tr>
<td>Female=1</td>
<td>-0.025</td>
<td>-0.149**</td>
<td>-0.0317</td>
<td>0.278***</td>
<td>-0.322**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0556</td>
<td>0.0563</td>
<td>0.373</td>
<td>0.0633</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td>0.249***</td>
<td>-0.0327</td>
<td>3.238***</td>
<td>2.155***</td>
<td>0.0147</td>
<td>0.224***</td>
</tr>
<tr>
<td></td>
<td>0.0257</td>
<td>0.0241</td>
<td>0.958</td>
<td>0.204</td>
<td>0.0251</td>
<td>0.0473</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.212</td>
<td>1.207</td>
<td>0.760</td>
<td>-0.575</td>
<td>0.012</td>
<td>-1.612***</td>
</tr>
<tr>
<td></td>
<td>0.226</td>
<td>0.525</td>
<td>7.049</td>
<td>2.062</td>
<td>0.327</td>
<td>0.455</td>
</tr>
<tr>
<td>N</td>
<td>608</td>
<td>608</td>
<td>100</td>
<td>600</td>
<td>608</td>
<td>608</td>
</tr>
</tbody>
</table>

Note: Reported statistics are average-effect coefficients (top; * p<0.05; ** p<0.01; *** p<0.001) and standard errors (bottom).

In addition to those listed above, specifications in the first five columns include the following control variables: tenure in LNP and categorical variables for age range, education, number of friends, rank, ethnicity, religion and birthplace.

* This specification also includes the following control variables: group standard deviation of risk preferences, group standard deviation of the desired amount to give in the pay-it-forward activity, and the group standard deviation of number of correct memory activity answers.
“Competence” in this specification refers to the group median value of competence. The specification also included the following variables: group mean tenure in LNP, group mean level of education, group mean age range, group mean number of friends, and categorical variables for the maximum rank in the group.

While there was more alignment in preferences (high task cohesion) with more women in groups, there is weak evidence that group deliberations became less collegial (low social cohesion), as coded by our enumerators and as reported by the group participants themselves, as women were added. Model 2 of Table 4 shows these results, and positive effects indicate more frustration and thus less collegiality. Individuals in the four-women teams were about eight percent of a standard deviation less collegial than those in the all-male groups, and individuals in the all-female groups were about thirteen percent of a standard deviation less collegial, but neither result is statistically significant. Regardless of the collegiality in how decisions were made, the outcomes of those decisions still demonstrate that individual preferences were more likely to be represented in the group decisions in the all-women groups. This is potentially good news because the research on unit cohesion shows that task cohesion is actually more important than social cohesion when it comes to overall operational effectiveness (MacCoun et al. 2006).

**FINDING 2: Operational Effectiveness with Respect to SGBV**

We find, contrary to Hypothesis 2, no evidence that the integration of women into groups affects sensitivity to SGBV. Model 3 of Table 4 presents results from a logistic model in which the dependent variable is an indicator for whether the group, after its deliberations, listed SGBV as a suspected crime in the crime-scene photos. In addition to the lack of statistical significance, there is a non-monotonic relationship between the number of
women in the group and the probability that the group would suspect a gender-based crime. Importantly, groups of four women were actually less likely to suspect SGBV than were groups of all men (our residual category), although not significantly so.

We also, in Model 4, look at the individual subject’s (not the group’s) suspected crimes from the crime-scene photo. As in Model 3, these are logistic model estimates on an indicator variable that is equal to one if the subject suspected SGBV in the crime-scene photo. The hypothesis is that serving on a team with more women makes team members more attuned or primed to suspect SGBV. The estimates offer no support for Hypothesis 2.

Several points in particular are worth noting. First, we do not find that women are more likely, ceteris paribus, to suspect SGBV than men. The coefficient on the female indicator variable is actually negative and very close to zero. This finding contradicts the supposed mechanism through which including more women might increase awareness of SGBV. Second, as in the group result, the estimates are not monotonically increasing: the effect of being in a six-woman group was actually a bit smaller than the effect of being in a four-woman group. This non-monotonicity defies the expectations of Hypothesis 2.

Finally, it is worth noting that there is more to being effective in responding to SGBV than just having a sensitivity to the issue. It is still possible that women might be more effective in responding to SGBV even if they do not have a proclivity to see crimes as gendered, by, for example, being able to better interview female survivors. Or, women may be more likely to confide in female police officers.  

---

28 There is some evidence that suggests this may be the case (Karim 2017a).
**FINDING 3: Organizational Gender Norms**

Finally, we assess whether or not organizational gender norms might change based on shifts in attitudes, group participation, and influence. In Model 5 of Table 4 we test whether, after serving on a team with more women, team members were more likely to recognize the value of women in the LNP. Our results, based on questions in the exit questionnaire, do not provide evidence for a change. The coefficients in this case are extremely small—even if the results were statistically significant they would be substantively insignificant. The results do, however, show very strong differences between men and women in their answers to these questions. This is perhaps best illustrated in the last graph in Figure 1. The lines are fairly flat, indicating no effect of the treatment, but the line for women is substantially higher than the line for men, indicating that women were substantially more likely than men to affirm the value of women in the LNP. The coefficient in Model 5 indicates that women were about 28 percent of a standard deviation more positive about women’s roles in the LNP than were men. Though this is not an experimental finding, the data show that women’s beliefs about their role differ from those of men.

Model 6 of Table 4 indicates that while beliefs about women may not have changed, both male and female officers participated more in group discussions and in the tower building activity the more women there were in the group. The average effect sizes for groups with four and six women are quite large, although only the latter is statistically

---

29 We were concerned about the possibility of social desirability bias—subjects giving the “right” answers to questions according to their training, rather than what they truly believed. Intuitively we would suspect that social desirability bias, if it existed, should be more serious among more competent officers: in other words, we suspect that more competent officers should have been more likely to know the politically correct answers to the questions we asked. Not observing that pattern buoys our confidence in the findings in this column and the others.
significant. Compared to the control category (no women), participants in groups with four women exhibited 38.5 percent of a standard deviation more participation, and participants in groups of six women showed roughly 48 percent of a standard deviation more participation. In other words, members felt more comfortable participating in groups that contained more women.

Figure 1, which shows heterogeneous treatment effects on men and women, hints at two possible mechanisms for increased participation. The focus in these graphs should be on the magnitude of changes rather than their statistical significance, given the high multicollinearity between gender, the treatment and the interaction between them. The first possible mechanism is that men became much more active in groups in which they were outnumbered by women four to two. The increase in the average effect size on participation for men in groups of two women was roughly zero, but it rose to over forty percent of a standard deviation in groups with four women. This result provides some evidence for the “intrusiveness” effect, whereby men become more active when they feel challenged by women in domains where they have been traditionally dominant. An intrusiveness effect would also comport well with the increase in male non-collegiality in the groups with four women.

A second effect that is clear from Figure 1 is that women became more active in groups that contained more women. We measure participation using the raw number of interventions in group deliberations and during the tower building activity (rather than the proportion of total interventions), so these results capture actual increases in participation by women. Women on average participated less than men even in groups in which they were in the majority, but in all-female groups they participated just as much as men did in
all-male groups. The results suggest that women may feel inhibited to speak out and participate in groups with male majorities, but reach their full participation level when in all-female groups.

Finally, women appear to be able to influence groups when they are in the majority. Looking again at the private versus group preferences in heterogeneous groups, women’s preferences were better represented in the four-women groups than in the two-women groups. We see this effect in the top-left graph in Figure 1, as the alignment of the group’s decisions with the individual women increases and the alignment of the group’s decisions with the individual men decreases when moving from the two-women group to the four-women group. As women were included in groups, they were better able to influence the final group decision to be more in line with women’s preferences.

<INSER FIGURE 1 HERE>
6. Conclusion

Gender balancing in the security sector is an increasingly common reform in post-conflict countries, especially if peacekeeping missions are present. The UN has repeatedly stressed that increased representation of women in security forces (and other traditionally male-dominated institutions) helps improve overall peace and security for all. Our theoretical priors suggested that gender balancing may have a number of implications in terms of addressing the processes that mediate larger state building goals including unit cohesion, operational effectiveness with respect to SGBV, and organizational norms of inclusivity. In our experiment, adding more women increased unit task cohesion and increased the participation and influence of women. There was no statistically significant evidence of less collegiality (social cohesion) as more women were added to teams, and the all-women teams had the greatest amount of similarity between individual preferences and group decisions (task cohesion).

We found no evidence to suggest, however, that simply adding more women would increase group (or individual) sensitivity to women’s issues. The LNP women in our study were no more sensitive to SGBV issues than men, ceteris paribus. We also found that although there was an increase in participation and influence by women, male beliefs about women’s roles in policing did not improve with the inclusion of women. Moreover, we found some evidence of an intrusiveness effect, as the participation and non-collegiality of men were highest when the men were outnumbered.

Interestingly, one the main determinants of task cohesion, participation, and sensitivity to SGBV was not individual-level gender or group-level gender composition,
but rather overall *competence*.\(^{30}\) This suggests that security sector reform, at least in its goal of improving police responsiveness to SGBV, should focus not only on gender balancing, but also on training police officers to increase competence, as this appears to improve overall operational effectiveness more broadly.

One concern of this study may be that the outcomes take longer to manifest. It is possible that with more prolonged interactions with women, these outcomes might improve even more, as suggested by Finseraas et al. (2016). Regardless, the experiments in this study provide important insight into group dynamics in security organizations in which groups are created ad-hoc to work on operations, as is often the case in police forces.

There may also be concern that these results will not generalize to other contexts. Do members of the LNP have enough in common with members of police forces in other countries? Many weak, post-conflict countries are implementing reforms very similar to those adopted in Liberia, and the initial problems that affected the implementation of these policies in the LNP—gender inequality in the country, low availability of women, societal backlash, etc.—are likely to be universal. This suggests that the context in which these reforms occurred is generalizable.

The results from our experiments are important for proponents of gender balancing policies, and for proponents of security sector reform more broadly. Our findings suggest that gender balancing policies may be able to provide employment and influence in the security sector without undue disruptions—disruptions that critics worry might arise from adding women to a highly male-dominated profession. Many of the findings were especially strong among all-female groups, which were more cohesive and participatory.

---

\(^{30}\) See Karim and Gorman (2016) for details of LNP competence.
Given that there is a trend in moving toward all-female units for security (Pruitt 2016), this analysis is particularly important as it demonstrates that these units might perform differently, if not better than all-male groups. As such, the study could help bolster the case for gender balancing policies in the security sector, and coupled with other work that shows the positive effect that these policies can have on communities, is one that could be replicated in other post-conflict countries.

WORKS CITED


31 The U.S. also used all-female “Engagement Teams” in Iraq and Afghanistan wars.


