I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11323

Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams

Gordon B. Dahl
Andreas Kotsadam
Dan-Olof Rooth

FEBRUARY 2018

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11323

Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams

Gordon B. Dahl
UC San Diego and IZA

Andreas Kotsadam
Ragnar Frisch Centre for Economic Research

Dan-Olof Rooth
SOFI, Stockholm University and IZA

FEBRUARY 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams*

We examine whether exposure of men to women in a traditionally male-dominated environment can change attitudes about mixed-gender productivity, gender roles and gender identity. Our context is the military in Norway, where we randomly assigned female recruits to some squads but not others during boot camp. We find that living and working with women for 8 weeks causes men to adopt more egalitarian attitudes. There is a 14 percentage point increase in the fraction of men who think mixed-gender teams perform as well or better than same-gender teams, an 8 percentage point increase in men who think household work should be shared equally and a 14 percentage point increase in men who do not completely disavow feminine traits. Contrary to the predictions of many policymakers, we find no evidence that integrating women into squads hurt male recruits' satisfaction with boot camp or their plans to continue in the military. These findings provide evidence that even in a highly gender-skewed environment, gender stereotypes are malleable and can be altered by integrating members of the opposite sex.

JEL Classification: J16, J24

Keywords: gender attitudes, occupational segregation, contact theory

Corresponding author:
Gordon B. Dahl
Department of Economics
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508
USA
E-mail: gdahl@ucsd.edu

* We thank seminar participants at several universities and conferences for useful comments and sug-gestions, and especially Henning Finseraas. Special thanks to the Norwegian Defense Research Estab-lishment (FFI), and in particular to Frank Steder and Torbjørn Hanson. This study could not have been conducted without their help and the support of FFI's Age Cohort Research project team. Thanks also to the soldiers and staff at the North Brigade who participated in the study. Ada Fuglset, Eirik Strømland, and Wiktoria Szczesna provided excellent research assistance. The research presented in this paper has IRB approval from The Norwegian Center of Research Data (Number 39028).

1 Introduction

Despite women making up almost half of the labor force in most developed countries, occupational segregation remains high.[1] Integration may partly have stalled due to preferences of men and women for different types of work environments, including women's desire for family friendly jobs with more flexibility (Goldin, 2014b; Goldin and Katz, 2016). On top of this, however, it could also be the case that further integration is hindered by exclusionary gender stereotypes and norms, especially in traditionally masculine work environments. Historically, women were denied access to universities, fired from their jobs if they married or had children and barred from certain occupations within firms (Goldin, 2014a). Removing these explicit restrictions resulted in an inflow of women into a variety of fields in the decades after World War II. While few formal constraints remain today, implicit restrictions stemming from gendered attitudes could remain a barrier for female employment and advancement in certain jobs (Bertrand et al., 2016; Moss-Racusin et al., 2012; Stamarski and Son Hing, 2015), especially when stereotypes cause belief distortions (Bordalo et al. 2016).

In particular, men's attitudes about mixed-gender productivity, gender roles and gender identity could play a critical role in explaining the persistence of occupational segregation. Employers may not hire women in male-dominated fields because they believe doing so will lower worker morale, group cohesiveness or productivity (Harrell and Miller, 1997). Stereotypical attitudes about gender roles related to home production versus paid work could reduce women's willingness to invest in certain careers, and also influence which tasks women are assigned within the workplace (Alesina et al., 2013; Goldin and Katz, 2002; Goldin, 2004). Likewise, gender identity concerns such as those discussed in Akerlof and Kranton (2000), Bertrand et al. (2015) and Goldin (2014b) could contribute to occupational segregation. In this paper, we examine whether exposure of men to women in a traditionally male-dominated environment can change these types of gender attitudes.

Understanding the link between gendered attitudes and occupational segregation

[1] In the second half of the 20th century, occupational sex segregation steadily declined in both Europe and the U.S., but stalled starting in the 1990s (Bettio and Verashchagina, 2009; Blau et al., 2013; Olivetti and Petrongolo, 2016). Around the same time, female labor force participation reached its peak and the convergence in wage earnings slowed (Blau and Kahn, 2003, 2017). Consistent with these trends, studies find that occupational sorting by gender accounts for a sizable portion (between 22 and 42 percent) of the gender wage gap (Goldin, 2014a). For additional evidence on and explanations for the slowing of women's progress, see Bailey and DiPrete (2016).

is important for designing policies which aim to better integrate the workplace.[2] Yet estimating the causal link between the two is difficult due to reverse causality, self-selection and unobserved heterogeneity. Working in a male-dominated occupation could cause men to develop less egalitarian attitudes, but it is also possible that men with less egalitarian attitudes select into these types of occupations in the first place. Dynamically, it is difficult to know if some occupations become more integrated because attitudes are changing, or if increased integration is responsible for a shift in attitudes.

To overcome these empirical challenges, we set up a field experiment which randomly assigned women to teams which were traditionally all male. Our context is the military in Norway, where women make up less than 15 percent of the professional military force (a fraction which is similar to the U.S.). In cooperation with the Norwegian Defense Research Establishment, we randomized female recruits to some squads but not others during the eight weeks of boot camp. The experiment provided intensive exposure of treated males to females, as squads are typically comprised of six members who train as a team and live together in the same room.[3] We conducted a baseline survey prior to the start of boot camp, as well as an endline survey near the completion of boot camp, both of which measured attitudes towards a variety of gender related questions.

Empirically, we find that living and working with women for eight weeks causes men to adopt more egalitarian attitudes. Our first result is that men who have women randomly assigned to their squads are 14 percentage points more likely to think mixed-gender teams perform as well or better than same-gender teams. This is a 24 percent increase relative to the overall mean of 55 percent. Second, exposed men are 8 percentage points more likely to believe that men and women should share household work equally, relative to an average of 67 percent. Third, we find that men exposed to women are 14 percentage points more likely to not completely disavow their feminine side, relative to an overall mean of 65 percent. These are sizable swings which move men closer to the attitudes of women: depending on the outcome, treatment reduces the gap in mean attitudes between men and women by between 31 and 46 percent.

Our results are robust to a variety of alternative specifications, including models based on first differences versus one-period lags, nonlinear estimation and varying squad sizes. We verify that observable male covariates are uncorrelated with the assignment

[2]For example, see Bayer and Rouse (2016), which argues that under representation of women in the economics profession could be due to implicit attitudes and institutional practices, and proposes several programs to address these problems.

[3]Throughout the paper, we interchangeably use the terms squad and room, since assignment to a squad is identical to assignment to a room.

of females to a team, and find no evidence that prior attitudes are affected by future exposure in a placebo test. These exercises provide empirical support for randomization.

We also asked questions about which sex makes the best leaders at various levels of command. We asked whether men or women make better troop leaders (a troop contains several squads), better leaders at the highest level in the military and better leaders for foreign operations. We do not find that exposure to a female squad member changes these attitudes. To put this in context, note that we did not randomize the gender of these higher-level leaders, and that most of them are male. This suggests that exposure to rank and file women does not have a spillover effect on attitudes about higher up female leadership. What may be required is exposure to female leaders at the relevant level (Beaman et al., 2009; Bertrand et al., 2014). Indeed, using similar Norwegian military data, Finseraas et al. (2016) conducted a vignette experiment which asked about attitudes towards hypothetical females becoming squad leaders, and found that exposure at the squad level did make a difference.

Beyond the general lessons learned about how integration can alter gender attitudes, these results are relevant for countries which are considering further integration of women into the military. Many policymakers have worried that allowing women to serve in the military would ruin its esprit de corps, causing lower performance and dissatisfaction within the ranks.[4] However, we find no evidence that integrating women into squads hurt male recruits' satisfaction with either boot camp or with their assigned room. Further, we find that male recruits exposed to women remain just as willing to complete their one-year military service, feel equally qualified for further military service and are equally likely to consider applying for a job in the military after their initial assignment is over.

Our paper is related to Fortin (2005), which finds that gender role attitudes across 25 OECD countries are associated with female employment rates and the gender wage gap, and Fortin (2015) which links changes in gender attitudes to the leveling off of female labor force participation in the U.S.

Our study is also related to the literature on contact theory which explores how biases and beliefs of a dominant group are affected by exposure to members of a minority group. This theory predicts that mixing groups will break down stereotypes and encourage understanding, especially if the two groups interact at a personal level

[4]For examples, see "Marine Commander's Firing Stirs Debate on Integration of Women in Corps," New York Times July 12, 2015 and "Gender Integration of Marines Brings Out Unusually Public Discord," New York Times, September 18, 2015.

and are given equal status and common goals as is the case during boot camp (Allport, 1954; Pettigrew and Tropp, 2006).[5] Field experiments which randomly assign a college roommate of another race or ethnicity find a reduction in prejudice (Boisjoly et al., 2006; Burns et al., 2016; Van Laar et al., 2005), and exposure to black students within a military setting results in less racial bias and a higher likelihood of rooming with a black in the following year (Carrell et al., 2015). Similarly, Finseraas and Kotsadam (2017) and Finseraas et al. (2017) find that contact with immigrants in a military setting changes attitudes and increases trust in immigrants.[6]

Our paper contributes to the literature on contact theory by being the first field experiment to randomly assign men to work and live together with women for an extended period and study gender attitudes, similar to what has been done in the race and ethnicity studies.[7] Unlike the race and ethnicity studies, however, our treated individuals have already likely had many repeat and personal interactions with members of the opposite group. This is because boys attend school with girls, grow up with sisters or have other female relatives, and live in the same neighborhoods. What is new about boot camp is the experience of working side-by-side with women in a traditionally male-dominated setting.

The remainder of the paper proceeds as follows. Section 2 describes our setting and the implementation of our field experiment. Section 3 introduces our survey questions and describes the data. In Section 4, we present our main experimental results for gender attitudes. We then present results for attitudes on female leadership, recruits' satisfaction with boot camp and plans to continue in the military. Section 6 concludes.

[5] Pollution theory, on the other hand, suggests that female employees will reduce the prestige and wages of previously male-dominated occupations (Goldin, 2014b). Stigma could increase if the rise in female employment is driven by a gender quota, although laboratory hiring experiments find that stigma is reduced when merit-based criteria are introduced (Evans, 2003; Resendez, 2002).

[6] These studies are related to a broader literature on the peer effects of randomly assigned roommates (e.g., Sacerdote, 2011; Stinebrickner and Stinebrickner, 2006; Zimmerman, 2003). They are also related to the observed reductions in racial discrimination in criminal trials when whites serve with black jurors (Anwar et al., 2012), and to the finding that randomly exposing wealthy students to poor students affects pro-social behavior and discrimination in Indian schools (Rao, 2014).

[7] The U.S. Marine Corps has examined the performance of mixed-gender teams, but only in the context of rotating physical assignments which lasted a few hours, such as a two hour hike or rifle shooting (see "Marine Corps Study: All-Male Combat Units Performed Better than Mixed Units," NPR, September 10, 2015).

2 Setting and Field Experiment

2.1 Occupational Segregation and Attitudes in Norway versus other Countries

As background, we first document the amount of gender segregation and gender attitudes in Norway compared to other countries. Norway is often viewed as a progressive country when it comes to gender issues, but like the rest of Europe and the U.S., its workforce is highly sex segregated. In both the U.S. and Norway, 47 percent of the labor force was female as of 2014. Table 1 provides some examples from 2014 of both male-dominated and female-dominated occupations. In both countries, stereotypically female jobs such as kindergarten teachers, nurses and social workers are primarily held by women, while stereotypically male jobs such as firefighters, pilots and computer programmers are primarily held by males.[8] An especially relevant comparison is that women comprise 13% of the military in Norway, compared to 15% in the U.S. (not including civilian employees).

To provide a more holistic view of the amount of segregation, we calculated Duncan segregation indices using 4 digit occupational codes from Norwegian census data (Duncan and Duncan, 1955). This commonly used measure is calculated as the absolute difference in the fraction of men and the fraction of women in an occupation, summed over all occupations and multiplied by one half. In 2010 in Norway, we calculate the index to be 0.53 (based on 483 categories), which means that 53% of women would have to change occupations to make the occupational distributions of men and women identical. Norway's index is slightly higher compared to other EU countries when using common occupational classifications (Bettio and Verashchagina, 2009). Using U.S. census data, we calculate a Duncan index of 0.51 for the U.S. (based on 491 categories); while the U.S. occupational categories do not map directly to those of the EU countries and so cannot be directly compared, segregation is clearly high in the U.S. as well.

We next provide a comparison of gender role attitudes in Norway versus other countries, based on common questions asked in OECD countries. Using responses from the World Values Survey (WVS), Fortin (2005) documents that 80% of women and 64% of men in Norway believe that working mothers can have a warm relationship with their children. The U.S. has somewhat more positive views about maternal labor force participation, with 83% of women and 73% of men agreeing with the statement.

[8]The fraction female in these gendered-occupations are remarkably similar across the two countries, with a few exceptions. For example, pharmacists are heavily female in Norway but not so in the U.S., while architects are heavily male in the U.S. but not so in Norway.

In both countries, the gender gap in attitudes is substantial. Another question in the WVS asks if being a housewife can be fulfilling. Here, Norway diverges from the U.S., with 54% of women and 56% of men agreeing with the statement, compared to 76% of women and 76% of men in the U.S. For both of these attitude measures, Norway is in the upper middle of the distribution (i.e., somewhat more gender equal) compared to other OECD countries participating in the survey.

2.2 Military Conscription in Norway

Norway has a selective mandatory military draft. Since 2010, all 17 year old men and women are required to register and be screened for service by the military. The first step in the screening process is internet based and involves answering questions related to health, school, personality, motivation and desire to serve. Around 60,000 males and females complete this online form each year. In the second step of the screening process, approximately 20,000 18 and 19 year olds are selected for physical and cognitive testing, and an interview with a recruitment officer. Based on this screening, the military then selects a subset of individuals for military service. Around 8,000 to 10,000 individuals are chosen to serve and sent to boot camp for 8 weeks of training. This is followed by approximately 10 months of further required service, with the exact length depending on an individual's assignment. At the end of this initial assignment, individuals can choose whether to apply for further service in the military.

Service is mandatory for men if chosen. But the screening process generally prioritizes candidates who express a motivation and willingness to be in the military, so de facto, most men are not coerced into service. Around one in six males end up being chosen for service. During the period of our study in 2014, service was voluntary for women, even though they had to participate in the screening process. Around 14 percent of the women sent to boot camp for training are women. Both men and women who end up serving in the military are therefore self-selected to be highly motivated. This type of selection is similar to what would be observed in a regular workplace, where individuals apply for and are hired for jobs for which they are a good fit.

As is true in many countries, Norway has been trying to increase female representation at all levels in the military. The goal is to have one out of five positions in the military staffed by women by the year 2020 (Ström-Erichsen 2013). To help achieve this goal, in 2013 the Norwegian Parliament passed a law extending the mandatory draft to women, making it the first European country to do so.[9] Mandatory conscription of women took effect in 2015 (a year after our study period). The Norwegian Chief of Defense, Admiral Haakon Bruun-Hanssen, heralded the change, saying “The new law means equal rights and duties for men and women... Now we have twice as many people to choose from. This will make it easier to direct motivated personnel and the right expertise to our different tasks and positions” (Norwegian Armed Forces, 2015).

2.3 Field Experiment and Survey Administration

Gender-mixed squads are increasingly being used throughout the Norwegian military in order to facilitate the integration of female soldiers into the military. In part to assess the effectiveness of such integration, the Norwegian Armed Forces and the Norwegian Defense Research Establishment have conducted a series of surveys starting as early as 2008 (see Hanson et al. (2016) for a summary of findings). The surveys and accompanying reports focus on recruitment and motivation to serve, but have also probed attitudes towards gender integration among soldiers. The results of these surveys suggest the integration of women was largely successful, although the analysis is based on non-random assignment of women to squads.[10]

The Army’s North Brigade has been at the forefront of integration: since 2010 all eight of its battalions have had mixed-gender sleeping quarters. In order to evaluate the effects of mixed-gender rooms on men’s attitudes, we convinced three battalions of the North Brigade, or about half of the contingent, to randomize soldiers into rooms during boot camp in 2014. These three battalions are the Second Battalion of Northern Norway (Andre Bataljon Nord-Norge), the Artillery Battalion (Artilleribataljonen), and the Armored Battalion (Panserbataljonen). In these battalions, the room also constitutes the squad. A squad is a team which both lives and works together and completes training tasks and assignments as a unit. Between 5 and 10 squads form a troop, and troops with a common function make up a battalion.

We emphasize that our field experiment did not introduce the use of mixed-gender squads, as male and female recruits had already been integrated into squads in the North Brigade for several years before our study. By the time we implement our field

[9] See “Norway becomes first NATO country to draft women into military,” Reuters, June 14, 2013.

[10] Anthropological studies without random assignment by Hellum (2014) and Hellum (2017) find that gender mixed rooms increase feelings of sameness across gender and reduce gender essentialist notions, while Lilleaas and Ellingsen (2014) find that mixed rooms promote mutual understanding, de-sexualization and reduced sexual harassment. For a summary of qualitative studies, see Ellingsen et al. (2016).

experiment in 2014, the military had several years to iron out any logistical issues and assignment to a mixed-gender squad was not anything out of the ordinary. Moreover, recruits were not told that assignment to a squad with or without women was part of an experiment.

Our baseline survey was conducted at a military base near Oslo, where recruits are given a battery of final qualification tests before service starts. Recruits are divided into groups of 20-30 members, and rotate among testing stations throughout the day. These testing stations assess a recruit's physical fitness, mental ability and psychological profile. We managed one of the stations, having recruits fill out an online survey questionnaire that we developed in consultation with the Norwegian Defense Research Establishment. In addition to our gender attitude questions, the survey asked questions about demographics, personality traits, leadership potential and military service. Recruits were told the survey was for research purposes only, and would not be a part of their official record or used by the military for screening or assignment purposes. The survey included over 100 questions and took on average 18 minutes to complete.[11]

The recruits were subsequently flown to northern Norway for an eight week boot camp. Before their arrival at boot camp, recruits were randomly assigned to rooms/squads, most of which have six individuals in them. Randomization occurred at the troop level, with officers using a template Excel spreadsheet which was programmed in advance to randomize name lists. While the military wanted at least two women per room if possible, officers had some discretion to override this rule.

The randomizer worked as follows. First, officers entered the number of rooms, which normally hold 6 individuals each. Then females were assigned randomly to rooms, with two females per room unless there was an odd number, in which case 3 women were assigned to a room. Finally, each room was randomly assigned male soldiers up to the specified room size. Since the number of individuals in a troop does not necessarily equal a multiple of 6, the troop officer could make manual adjustments to even out room sizes. At this stage, an officer could manually override the rule of at least two females per room; for example, he could split three women among multiple rooms. Note that randomization is preserved with these manual adjustments, as room assignments

[11]The survey was conducted in a classroom, with recruits being given an internet link to an online survey. Most soldiers used their own mobile phones to complete the survey, but tablets and paper versions of the survey were also available for use. While participation was voluntary, recruits had to stay in the room while the survey was being conducted and were not allowed to talk to each other.

are completed prior to the arrival of recruits at boot camp when the only information available is gender and name. During boot camp, it is possible that individuals are moved to different rooms, which is something we do not measure. When we asked the military, they said this happened rarely, if at all. A conservative interpretation of our estimates is that they capture intention to treat (ITT) effects, based on a recruit's initial assignment.

Boot camp, which is also known as basic training, lasts eight weeks. Its purpose is to prepare recruits physically, mentally and emotionally for service. It is an intense period of training, comprised of both field exercises and classroom time. Many tasks are completed as a group, with squads learning to function effectively as a team. Since squad members train as a unit and share living quarters, and are not allowed to leave base during boot camp, they spend most of their daytime and nighttime hours together.

Near the completion of basic training, an endline survey was conducted at each battalion's base in Northern Norway. Soldiers were gathered and completed a slightly modified version of the baseline survey, using phones or tablets as in the baseline survey.

3 Data and Survey Questions

3.1 Descriptive Statistics

In the three battalions which randomly assigned women to squads, there are a total of 20 troops. The smallest troop has 5 squads, while the largest has 10 squads. Three-fourths of squads in our data have a standard room size of 6 members, although other room sizes arise. The most common reason for a nonstandard room size is that the number of members within a troop is not equally divisible by 6. For our main sample, we focus on the 95% of squads which have between 5 and 7 members in them.[12] In our main sample, we have a total of 153 rooms (where a room constitutes a squad). Fifty-seven of these rooms have a female in them, with two thirds having exactly two women, 19 percent having 1 woman, and the remaining having 3 or 4 women.

The baseline survey was asked of all recruits in Oslo, including the battalions which did not participate in the randomization of women to rooms. We used an anonymous id number for each soldier to merge our survey data with administrative data from the military on muscle strength and cognitive skills. The cognitive skill measure is a general

[12] Our empirical results are robust to including all room sizes, as well as using only the standard room size of 6.

ability index (GAI) based on verbal comprehension and perceptual reasoning tests which have been shown to be correlated with a full-scale IQ test in other contexts. Summary statistics collected at baseline for both the entire sample and our main randomized sample are found in Appendix Table A1.

Our main sample for the three battalions which agreed to randomization includes 781 men and 119 women. The appendix table lists several background characteristics for these recruits. Two of these variables, muscle strength and general ability test scores, are of particular note, as the military uses this information to choose recruits. Men in our sample have above average values for both of these variables. Over half of men report having above average or far above average muscle strength relative to their male peers. Likewise, 51 percent of men score at the 6th stanine or above on the general ability test administered by the military (by construction, 40% of the population will score between the 6th and 9th stanine). Women in our sample have above average muscle strength, but lower general ability scores (only 25 percent score at the 6th stanine or above).

Basic demographic variables are also summarized in the tables, as are the fraction of missing values. For variables collected by the military, there are between 1 and 8 percent with missing values, depending on the particular subsample. Our survey questionnaire has somewhat higher missing values, with between 10 and 16 percent missing for most questions.[13]

3.2 Gender Attitude Questions

The main purpose of our experiment is to assess how integrating women into rooms affects male recruits' attitudes towards mixed-gender productivity. We are also interested in the broader issues of how female integration affects perceptions of gender roles and gender identity. We ask three gender related questions, both at baseline and after boot camp is over, to see how men's attitudes change. The distribution of attitudes for these questions are found in Figure 1. For our analysis, we will dichotomize the possible answers to create indicators to use as our main outcome variables. In each case, a value of 1 corresponds to a more gender equal viewpoint.

Our first question relates to whether mixed-gender teams underperform all-male

[13]For the questions about sisters and brothers, 25 to 30 percent have missing values. This appears to be because some respondents without a sister (or without a brother) skipped the question if they had 0 sisters (or 0 brothers), even though "none" was explicitly an option. We include a dummy variable for missing values in our regressions so as to be able to include as many observations as possible.

teams in our military setting. The question asks respondents to give their view on the statement: “Teams perform better when made up of the same sex.” Respondents were given a scale, beginning with 1 for “Completely disagree” up to 7 for “Completely agree.” The distribution of answers for males at baseline for our main, randomized estimation sample is found in the top panel of Figure 1. We create a dummy variable equaling 1 if the respondent disagrees with this statement, which we define as an answer of 1, 2 or 3 on the scale. Using this categorization, 63% of men disagree with the statement at baseline.

Our second question relates to gender roles. A commonly used method to assess attitudes regarding traditional gender roles is to ask about the division of household chores. We ask individuals their opinion on the statement: “It is important for men and women to share household work equally.” Respondents could answer “Strongly agree,” “Agree,” “Neither agree nor disagree,” “Disagree” or “Strongly disagree.” The distribution of answers is plotted in the second panel of Figure 1. We define a dummy variable for agreement which equals 1 if the individual either agrees or strongly agrees with the statement. Based on this dummy variable, two thirds of men agree with the statement.

Our final gender question concerns a somewhat different concept, namely an individual’s self-perception of femininity. Instead of studying stereotypical attitudes about others, this question asks people about themselves. A series of questions regarding personality traits was asked in the survey, with respondents being asked how well certain statements described them. In this list was the statement “I am feminine.” Respondents could answer “Does not fit at all,” “Does not fit well,” “Reasonable fit,” “Fits well,” or “Fits completely.” We view this question as being related to gender identity, and the self-perception of being different from women, similar in spirit to Akerlof and Kranton (2000) and Bertrand et al. (2015). A similar idea, but related to racial identity, is discussed in Austen-Smith and Fryer (2005).

Figure 1 reveals that virtually no men think this statement is an apt description of them. The relevant distinction is whether the respondent thinks the statement does not fit them at all versus whether it is merely a poor fit. We define a dummy variable which equals 0 for an answer of “Does not fit at all” and a 1 for all other answers. When the dummy variable is 0, we interpret this as a complete disavowal of femininity; roughly 58% of male respondents do not completely disavow their femininity by this measure in the baseline survey. It is important to keep in mind that a value of 1 does not mean the respondent feels they are feminine, but rather corresponds to an answer of "does not fit well" (versus "does not fit at all") for most individuals.

It is informative to compare men's answers to those of women. Appendix Figure A1 plots the distribution of female's attitudes to the same three questions at baseline. In contrast to men, only 10% of women think same gender teams outperform mixed-gender teams. Women are also more likely to believe household work should be shared equally among the sexes (88%). And finally, no women completely disavow their femininity, with most women acknowledging it but not subscribing to the description completely.

It is also informative to compare attitudes in our sample of military recruits to those in the general population. This is only possible for the question regarding household work, since that is the only issue for which we could find a comparable question. Surveys conducted by Jakobsson and Kotsadam (2010; 2011; 2014) in conjunction with Gallup asked a random sample of Norwegians "Is it important that women and men share responsibility for the household?" Respondents could answer on a scale, with 0 labeled as "No, not at all" and 10 as "Yes, of course." In Appendix Figure A2, we collapse responses from the Gallup surveys into 5 categories to enable an easier comparison to the household work question in our survey. While the questions are somewhat different, it appears men in our military sample have less gender egalitarian attitudes compared to the general population. In contrast, there is little difference in the distribution of attitudes for women in the military versus the general population. We conclude that males with less gender-equal attitudes select into military service, while female recruits are not a selected sample on this dimension.

The survey also asked questions about whether men or women make better leaders at various levels in the military, as well as questions about satisfaction with boot camp and plans to continue in the military. We discuss these questions when we analyze them later in the paper.

4 Female Exposure and Gender Attitudes

4.1 OLS

We begin by presenting OLS regressions of gender attitudes on background characteristics in Table 2. These regressions are based on the entire sample of baseline survey respondents, regardless of whether they are in a battalion which participated in the field experiment. We have two measures of prior exposure to females: whether the male recruit has a large share of female friends, and whether he has a sister. We also include controls for muscle strength, general ability test scores and parental characteristics.

Column 1 regresses a dummy variable for whether same gender teams perform better, with a value of 1 for disagreement. Recruits who respond that more than 40% of their friends are females are 8 percentage points more likely to think mixed-gender teams perform better. In contrast, men with high muscle strength are 11 percentage points less likely to think mixed-gender teams have better performance. The other variables are not statistically significant. In column 2, the dependent variable is whether respondents agree it is important for men and women to share household work equally. In this regression, having a sister increases the probability of agreement by 6 percentage points. Men with above average general ability test scores are less likely to believe household work should be shared equally, although it is not obvious why this pattern holds. Finally, in column 3, the outcome is whether the respondent does not completely disavow their femininity. Men with high muscle strength are more likely to reject their feminine side completely, while men with above average general ability test scores are less likely. Having parents who are divorced also significantly decreases the recruit's disavowal of a feminine side.

It is tempting to conclude that having female friends causes individuals to believe same gender teams perform better, but it is also possible that those who think mixed-gender environments are better are those who choose to have female friends. Having a sister, which predicts less stereotypical gender roles for household work, is arguably more exogenous, although endogeneous fertility could create a bias.[14]

4.2 Experimental Results

We are interested in the causal relationship between female integration in the squad and men's gender attitudes. Attitudes are measured both before and after treatment. We model attitudes for individual *i,* in squad *j,* in troop *k,* in period *2* (the endline survey, after treatment), as:

$$y_{ij2} = \alpha_k + \beta y_{i1} + \theta F_{ij} + \gamma x_{i1} + \epsilon_{ij2} \tag{1}$$

where α_k is a set of troop fixed effects, y_{i1} measures baseline attitudes in period 1 (before treatment and assignment to a squad), F_{ij} is a dummy variable for whether a

[14]For example, fertility stopping rules as a function of child gender could influence the presence and gender composition of siblings, and hence reflect underlying gender attitudes of parents (Dahl and Moretti, 2008).

female is assigned to individual i's squad j and x_{i1} is a set of pre-determined control variables measured in period 1.

In our regressions, we use binary indicators for attitudes in period 2, with a value of 1 indicating a more gender equal attitude for our main outcomes. To control for initial, pre-treatment attitudes, we include a full set of dummy variables for the possible answers to the baseline questions.[15] Troop fixed effects are included since randomization occurs within troops. The regressions also include all of the control variables used in Table 2. Since treatment is at the room level, in all of our experimental regressions we report standard errors clustered by room.

Table 3 presents our main experimental estimates for the gender attitude questions. For each question, we report two sets of estimates. The first includes troop fixed effects and dummies for initial gender attitudes in period 1.[16] The second adds in the covariates appearing in Table 2. As expected given randomization, adding in these additional covariates does not materially affect the estimates. Therefore, we focus our discussion on the columns in the table which include these pre-determined covariates.

Column 2 reports how having a female assigned to your team affects attitudes about mixed-gender productivity. Men exposed to female team members are 13 percentage points more likely to disagree with the claim that same gender teams perform better. Relative to the endline mean of 55 percent, this is a 24 percent increase.

In column 4, the dependent variable is whether recruits agree that it is important for men and women to share household work equally. There is an 8 percentage point increase in egalitarian attitudes for this outcome after working and living with females for the 8 weeks of boot camp. This is a 12 percent increase relative to the endline mean of 67 percent.

Finally, column 6 shows how exposure changes attitudes related to gender identity. There is a 14 percentage point increase in the number of men who do not completely disavow their femininity, relative to an overall mean of 65 percent. It is important to remember that this change in attitudes does not represent a full embracement of being feminine; rather the relevant comparison is that most men are switching from a complete disavowal of femininity to a weaker disavowal of femininity.

These are large swings in gender attitudes. One way to see this is to compare the size

[15] We include a dummy variable for missings and combine adjacent response categories when creating the set of dummy variables if there are less than 10 responses in a category.

[16] Including initial attitudes substantially increases the explanatory power of the regressions; relative to only including troop fixed effects, the R-squares rise from between .02 and .07 to between .19 and .36, depending on the outcome.

of these estimates to the coefficients on various background characteristics in Table 2. The magnitude of the effects are as large or larger compared to most of the coefficients, including having female friends or a sister. Another way to gauge just how large these effects are is to consider the gap which exists between men and women in these binary attitude variables. For the mixed-gender productivity question, there is a 31 percentage point gap in male versus female attitudes at baseline for our estimation sample. Over 40 percent of this gap is erased if men are exposed to women. Similarly, there is an 18 percentage point gap in attitudes towards sharing household work. Exposure to women erases almost half of this gap. As expected, the gap in the gender identity variable is the largest, with a 44 percentage point difference. Exposure to women reduces this gap by 31 percent.

Based on our experimental results, we conclude that men's gender attitudes can be changed via the intensive exposure to women provided by our field experiment. Experimenter demand effects or social desirability bias are unlikely to explain our results, for several reasons. First, treated recruits are not primed to be thinking about gender integration, as they are not made aware that women are randomly being assigned to some squads and not others as part of an experiment. Related, the survey was conducted in a large group setting, and not at the squad level. Moreover, integration of women into the military was not a new phenomenon, as it had already been going on for several years by the time of our field experiment. Finally, the survey questionnaire was not focused on gender issues, but instead asked a few gender questions interspersed among a long list of unrelated questions.

To provide some context, we document the changes in attitudes stemming from the experience of boot camp itself. Even in the absence of being exposed to women, boot camp could change men's attitudes about gender issues because it is a new, masculine environment. Focusing on men in all-male squads and comparing attitudes before and after boot camp, we find that 22 percent become less gender equal, 67 percent do not change and 11 percent become more gender equal in their attitudes towards mixed-gender teams. Likewise, for attitudes on sharing household work, 14 percent become less gender equal, 76 percent do not change and 10 percent become more gender equal. Finally, for the gender identity question, 7 percent become less gender equal, 80 percent do not change and 13 percent become more gender equal. It is interesting to note the largest asymmetric changes occur for the question on mixed-gender teams. Apparently, working intensively in an all male team during boot camp results in more men thinking this type of arrangement is better than a mixed-gender team.

4.3 Robustness and Heterogeneity

Table 4 presents a series of robustness and hetereogeneity results. For ease of comparison, the first panel reports our baseline estimates from Table 3. Specification B estimates a probit instead of a linear model. The marginal effects are similar to the baseline OLS estimates.

In panel C, we estimate the model in first differences. This changes the baseline specification in two ways. First, it dichotomizes lagged attitudes, instead of allowing for a set of dummy variables for the possible categorical responses in period 1 as control variables. Second, it constrains the coefficient on lagged attitudes to be 1. The estimates from this first difference model are positive and significant, and somewhat larger for the first two outcomes. In panel D, we exclude observations whose lagged attitude is missing. These observations are included in the baseline regressions, since we include a dummy variable for missings, but excluded from the first difference regressions in panel C. This constrained sample yields broadly similar estimates compared to baseline. The next specification excludes the controls for baseline attitudes entirely. The standard errors go up between 2 and 26 percent, depending on the outcome, but the results are qualitatively similar.

The next three robustness checks consider samples with different numbers of females in the room and different room sizes. It would be interesting to explore the differential impacts of having 1 versus 2 or more females in the room. However, the experiment was not set up to do this, as the military had a preference for 2 females in a room. There are only 11 rooms with 1 women in them, and even fewer with 3 or more women. With this caveat in mind, in specification F, we use two treatment variables: one female in the room and two or more females in a room. While less precise, the estimates are not statistically different from each other for any of the outcomes. Specification G includes all room sizes, rather than our baseline restriction of rooms with 5-7 recruits in them. This makes little difference. Finally, specification H restricts the sample to rooms with exactly 6 recruits and either 0 females or 2 females in the room. While this reduces the number of observations in the baseline sample by almost a quarter, the estimates are similar to baseline and remain statistically significant.

The last specification in the table interacts the presence of a female in the squad with the fraction of women in the troop. Between 5 and 10 squads make up a troop, and there are 20 troops in the Northern Brigade which participated in our experiment. Troop members interact with each other throughout the day, although not at the same intensity as within a squad. The fraction of women at the troop level varies substantially, ranging from 0 to 35%, with the median across troops being 13%. As reported in the table, there is no evidence for an interaction effect. However, we note the interpretation of this interaction term is not straightforward, as individuals are not randomly assigned to troops, and different troops perform different tasks (e.g., light infantry versus artillery troops).

As a final robustness check, we explore different codings for the outcome variables. As a reminder, we create a binary variable for each question which equals 1 if an individual has a pro-gender equality attitude. For example, for the household work question, we code an answer of strongly agree or agree as a 1 and neither agree nor disagree, disagree, or strongly disagree as a 0. If we recode the neutral category of neither agree nor disagree as a 1 instead of a 0, the coefficient falls and is no longer statistically significant (coeff. = 0.044, s.e. = 0.028). A similar result holds if we likewise code the "neutral" category of 4 for the mixed-gender teams question as a 1 instead of a 0 (coeff. = 0.048, s.e. = 0.048). These alternative codings suggest that at least some of the changes in attitudes captured in the baseline estimates are for individuals moving from a neutral attitude to a pro-gender equality attitude. For the gender identity question, there is no meaningful recoding, as most individuals give just one of two answers.

In Appendix Table A2, we report heterogeneous effects based on pre-determined characteristics of men. For example, for heterogeneous effects by men's muscle strength, we interact treatment with a dummy for whether the man has high versus low muscle strength. We find no statistical evidence of heterogeneity by men's muscle strength, or by men's general ability test scores. Likewise, we find no significant heterogeneity by whether the man has a high share of female friends, a sister or a highly educated mother. The same is true for heterogeneous interactions based on the man's initial gender attitude for the relevant outcome. We are quick to point out that heterogeneity could exist, but that our standard errors are too large to detect it. While it would also be interesting to explore heterogeneity by female's characteristics, we have even less precision to explore such effects.

4.4 Internal Validity

Male recruits were randomly assigned to squads/rooms, some of which had females in them and some of which did not. This assignment occurred before recruits arrived at boot camp. As a test of random assignment, in Table 5, we explore whether pre-determined covariates can predict treatment. We regress the treatment dummy, i.e., having a female in the room, on all of the covariates used in Table 2. While those covariates have predictive power for gender attitudes (see Table 2), they are not statistically significant predictors for treatment status. The coefficient estimates are close to zero and not statistically significant. Moreover, the joint test for all the variables is insignificant, with a p-value of 0.73.[17] This test for covariate balance across treatment and controls corroborates random assignment.

A related test for random assignment is to add pre-determined covariates into the regression. If assignment is random, then the estimated treatment effect should not change appreciably, which is what we found in Table 3.

To further test for random assignment, we conducted a set of placebo tests. The idea of the tests are that prior attitudes should not be affected by future exposure to females if individuals are randomly assigned. Indeed, this is what we find in Table 6. Initial attitudes measured before boot camp in period 1 are not affected by future treatment status for any of the three outcome measures.

Finally, we tested for differential attrition. Not everyone finishes boot camp, with about 15 percent of recruits not completing the full 8 weeks according to the military. In addition, some recruits will not be available to take our endline survey for other reasons (e.g., sickness or another conflicting assignment) or could choose to skip the gender attitude questions when taking the survey. Together, these reasons contribute to slightly less than one third of our baseline sample with missing values for the endline gender attitude questions.

If attrition is random, this should not affect the causality of our estimates. But if women cause men to exit boot camp or not complete the survey, this could create a bias. In Appendix Table A3 we test whether treatment predicts attrition. For each of our gender attitude questions, we regress a dummy variable for whether it has a missing outcome in the endline survey. We find no evidence that being assigned to a

[17] Although not reported, we also tried adding baseline attitude variables into this type of regression. The joint test for all of the variables, including both the variables listed in Table 5 and the baseline attitude variables, has a p-value of 0.85. Moreover, none of the estimated coefficients in this regression are individually significant and the group of baseline attitude variables is not jointly significant either.

room with a female in it increases attrition. If anything, the estimates are negative. While we present estimates from three separate regressions, it should be noted that having a missing value is highly correlated across the three attitude variables. It is therefore not surprising the estimates are similar.

5 Other Outcomes

5.1 Female Leadership

As part of the survey, we also asked questions about which sex makes the best leaders at various levels of command. Our first question on female leadership asks respondents: "Which sex do you believe is the best at leading a troop?" As a reminder, troops are groups of 5 to 10 squads. Possible responses are "Men," "Equally good," and "Women." While 48 percent of men in our analysis sample think men and women make equally good troop leaders, 52 percent think men are best. This contrasts with women, 82 percent of which think men and women are equally good or that women are best (of these, only 3 percent think women are best).

Our second question is: "Which sex do you believe are the best leaders at the highest level?" The possible responses remain the same, and the distribution of attitudes reveals a similar pattern. Sixty-five percent of men believe men and women are equally good or that women are best, compared to 84 percent of women. The third question is: "Which sex do you believe is the best at leading foreign operations?" Forty-six percent of men and 57 percent of women believe the genders are equally good or that women are best.

We create dummy variables which equal one for an answer of "Equally good" or "Women" for each of the questions. It is important to recognize that a value of 1 largely represents "Equally good," as only a handful of either male or female respondents think that women are the best leaders. Table 7 estimates similar regressions as for our main gender attitude questions in Table 3.

For all three female leadership questions, the estimates are close to zero and statistically insignificant. Apparently, exposure to women at the squad level does not affect these attitudes. To put this in context, note that we did not randomize the gender of these higher-level leaders, and that most of the leaders at these levels are male. This suggests that exposure to rank and file women does not have a spillover effect on attitudes about higher up female leadership.

What may be required is exposure to female leaders at the relevant level. Indeed, using a similar Norwegian military setting, Finseraas et al. (2016) conducted a vignette experiment which asked about attitudes towards hypothetical females becoming squad leaders, and finds that exposure to females at the squad level does make a difference. Related work by Beaman et al. (2009) finds that prior randomized exposure to a female leader through gender quotas results in changes in voter attitudes regarding the effectiveness of female leaders and weakens gender stereotypes. In a similar vein, Bertrand et al. (2014) find that while an increase in female board members may have improved the representation of female employees at the very top, there is no evidence these gains trickled down.

5.2 Satisfaction with Boot Camp and Plans for Future Service

While we are not able to measure the performance of squads during boot camp, we do have a set of questions related to the issues policymakers have focused on. Leaders within and outside the military have debated whether allowing women to serve in the military would ruin its esprit de corps, causing lower performance and dissatisfaction within the ranks. Many arguments have centered around whether women will "ruin" the military experience, and result in a less cohesive and less prepared fighting unit. From a recruiting perspective, a related worry is that as women enter military occupations, men will choose to leave them.

To test these hypotheses, the survey asked a series of questions about recruits' enthusiasm, self-assessed preparedness and plans to continue in the military. Table 8 reports estimates from regressions which parallel those appearing in Table 3. Our first question is "Do you want to do military service?" We find no evidence that having a female in the squad affected men's enthusiasm for the military, with a point estimate near zero. Likewise we find no effect on a respondent's agreement with the question "I feel qualified for further military service." Our third question is: "Do you plan to continue in the military once your initial assignment is completed?" Exposure to women during boot camp has no significant effect on this outcome either.

We then asked two questions about recruits satisfaction with their experience. We asked: "Overall, how satisfied were you with military service?" and "How satisfied were you with your room?" As opposed to the other questions asked in Table 8, there is no corresponding baseline response for these questions, as they are asking about the boot camp experience. So we cannot control for lagged responses in these regressions. For both questions, the point estimates are close to zero, and not statistically significant.

We conclude that, contrary to the predictions of many policymakers, integrating females into squads during boot camp did not result in measurable losses in male recruits desire to serve, preparedness, plans to continue in the military or satisfaction with boot camp.

6 Conclusion

This paper provides novel evidence on whether the integration of women into a traditionally male-dominated environment can change stereotypical gender attitudes of males. We overcome the difficulties of reverse causation, self-selection and unobserved heterogeneity by implementing a field experiment which randomly assigned some men to live and work with women during boot camp in the military. We find that this intensive interaction with women for 8 weeks causes men to become more egalitarian in their attitudes towards mixed-gender productivity, gender roles and gender identity. Moreover, the integration of women did not reduce male recruits' satisfaction with boot camp or their plans to serve in the military.

These findings demonstrate that men's gendered attitudes are not fixed, but can change through interaction with women. Our field experiment was a strong intervention, changing both living arrangements and the working environment. In this setting, men and women were equal in rank and had to complete a similar set of tasks. Moreover, men and women were placed into teams which required cooperation to reach common goals, such as the completion of a training exercise. These features combine to create exactly the type of setting predicted by contact theory to result in a change in attitudes.

Our findings have important implications for policies aimed at integrating the workplace as well as for societal norms more broadly. Our results suggest that exposure is an important lever which can be used to overcome pre-existing priors regarding a woman's suitability for a job, as well as stereotypical attitudes which affect life outside the workplace. And contrary to the concerns of some policymakers, it appears that efforts to integrate women into the military can be achieved without destroying the camaraderie of service.

Our study is the first to randomly assign men and women to live and work together for an extended period of time and study how attitudes change. This adds to prior work with has explored how extended exposure can change racial and ethnic attitudes. While the results of our field experiment provide new empirical evidence in support of contact theory in the gender domain, several questions remain. Do these results transfer to other settings, such as the integration of police forces, which have also been historically male-dominated? Can the attitudes of older individuals be changed via exposure, or are only young people's attitudes malleable? And finally, while the observed changes in attitudes are large, will they persist into the future? These are interesting questions for future research.

References

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Allport, G. W. (1954). *The Nature of Prejudice*. Reading: Addison-Wesley.

Anwar, S., P. Bayer, and R. Hjalmarsson (2012). The impact of jury race in criminal trials. *The Quarterly Journal of Economics 127*(2), 1017–1055.

Austen-Smith, D. and R. Fryer (2005). An economic analysis of "acting white". *Quarterly Journal of Economics 120*(2), 551–583.

Bailey, M. J. and T. A. DiPrete (2016). Five decades of remarkable but slowing change in U.S. women's economic and social status and political participation. *The Russell Sage Foundation Journal of the Social Sciences 2*(4), 1–32.

Bayer, A. and C. E. Rouse (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives 30*(4), 221–242.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: Does exposure reduce bias? *The Quarterly Journal of Economics 124*(4), 1497–1540.

Bertrand, M., S. E. Black, S. Jensen, and A. Lleras-Muney (2014). Breaking the glass ceiling? The effect of board quotas on female labor market outcomes in Norway. *National Bureau of Economic Research Working Paper 20256.*

Bertrand, M., P. Cortes, C. Olivetti, and J. Pan (2016). Social norms, labor market opportunities, and the marriage gaps for skilled women. *National Bureau of Economic Research Working Paper 22015*.

Bertrand, M., E. Kamenica, and J. Pan (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics 130*(2), 571–614.

Bettio, F. and A. Verashchagina (2009). Gender segregation in the labour market: Root causes, implications and policy responses. European Commission's Expert Group on Gender and Employment.

Blau, F. D., P. Brummund, and A. Y.-H. Liu (2013). Trends in occupational segregation by gender 1970-2009: Adjusting for the impact of changes in the occupational coding system. *Demography 50*(2), 471.

Blau, F. D. and L. M. Kahn (2003). Understanding international differences in the gender pay gap. *Journal of Labor Economics 21*(1), 106–144.

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Boisjoly, J., G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles (2006). Empathy or antipathy? The impact of diversity. *American Economic Review 96*(5), 1890–1905.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016). Stereotypes. *The Quarterly Journal of Economics 131*(4), 1753–1794.

Burns, J., L. Corno, and E. La Ferrara (2016). Interaction, prejudice and performance. Evidence from South Africa.

Carrell, S. E., M. Hoekstra, and J. E. West (2015). The impact of intergroup contact on racial attitudes and revealed preferences. *National Bureau of Economic Research Working Paper No. 20940*.

Dahl, G. B. and E. Moretti (2008). The demand for sons. *The Review of Economic Studies 75*(4), 1085–1120.

Duncan, O. D. and B. Duncan (1955). A methodological analysis of segregation indexes. *American Sociological Review 20*(2), 210–217.

Ellingsen, D., U.-B. Lilleaas, and M. Kimmel (2016). Something is working–but why? Mixed rooms in the Norwegian army. *NORA-Nordic Journal of Feminist and Gender Research 24*(3), 151–164.

Evans, D. C. (2003). A comparison of the other-directed stigmatization produced by legal and illegal forms of affirmative action. *Journal of Applied Psychology 88*(1), 121.

Finseraas, H., T. Hanson, Å. A. Johnsen, A. Kotsadam, and G. Torsvik (2017). Trust, ethnic diversity, and personal contact: Experimental field evidence. *Oslo University Mimeo*.

Finseraas, H., Å. A. Johnsen, A. Kotsadam, and G. Torsvik (2016). Exposure to female colleagues breaks the glass ceiling: Evidence from a combined vignette and field experiment. *European Economic Review 90*, 363–374.

Finseraas, H. and A. Kotsadam (2017). Does personal contact with ethnic minorities affect anti-immigrant sentiments? Evidence from a field experiment. *European Journal of Political Research*.

Fortin, N. M. (2005). Gender role attitudes and the labour-market outcomes of women across OECD countries. *Oxford Review of Economic Policy 21*(3), 416–438.

Fortin, N. M. (2015). Gender role attitudes and women's labor market participation: Opting-out, aids, and the persistent appeal of housewifery. *Annals of Economics and Statistics/Annales d'Économie et de Statistique* (117/118), 379–401.

Goldin, C. (2004). The long road to the fast track: Career and family. *The Annals of the American Academy of Political and Social Science 596*(1), 20–35.

Goldin, C. (2014a). A grand gender convergence: Its last chapter. *The American Economic Review 104*(4), 1091–1119.

Goldin, C. (2014b). A pollution theory of discrimination: Male and female differences in occupations and earnings. In *Human capital in history: The American record*, pp. 313–348. University of Chicago Press.

Goldin, C. and L. F. Katz (2002). The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of Political Economy 110*(4), 730–770.

Goldin, C. and L. F. Katz (2016). A most egalitarian profession: Pharmacy and the evolution of a family-friendly occupation. *Journal of Labor Economics 34*(3), 705–746.

Hanson, T., B. S. Frank, and K. Sverre (2016). Hva motiverer til tjeneste i forsvaret? En innledende kvantitativ analyse av holdninger og adferd i brigade nord. Technical report, FFI.

Harrell, M. C. and L. L. Miller (1997). New opportunities for military women effects upon readiness, cohesion, and morale. Technical report, RAND Corporation.

Hellum, N. (2014). Sminkedritt over hele vasken – en kvalitativ feltstudie av kjønnsblandede rom og maskulinitetskultur i Forsvaret. Technical Report 2156, FFI.

Hellum, N. (2017). Not focusing on whether it's a spout or a handle: An anthropological study on even gender balance among conscripts in a Norwegian Air Force battalion. Technical Report 01196, FFI.

Jakobsson, N. and A. Kotsadam (2010). Do attitudes toward gender equality really differ between Norway and Sweden? *Journal of European Social Policy 20*(2), 142–159.

Kotsadam, A. and N. Jakobsson (2011). Do laws affect attitudes? An assessment of the Norwegian prostitution law using longitudinal data. *International Review of Law and Economics 31*(2), 103–115.

Kotsadam, A. and N. Jakobsson (2014). Shame on you, John! Laws, stigmatization, and the demand for sex. *European Journal of Law and Economics 37*(3), 393–404.

Lilleaas, U.-B. and D. Ellingsen (2014). *Likestilling i Forsvaret Fortropp, baktropp og kamparena, (Gender equality in the Armed Forces: Battleground, rear guard and vanguard)*. Cappelen Damm Akademisk.

Moss-Racusin, C., J. Dovidio, V. Brescoll, M. Graham, and J. Handelsman (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences 109*(41), 16474–16479.

Olivetti, C. and B. Petrongolo (2016). The evolution of gender gaps in industrialized countries. *Annual Review of Economics 8*, 405–434.

Pettigrew, T. F. and L. R. Tropp (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology 90*(5), 751–783.

Rao, G. (2014). Familiarity does not breed contempt: Generosity, discrimination and diversity in delhi schools. Working paper.

Resendez, M. G. (2002). The stigmatizing effects of affirmative action: An examination of moderating variables. *Journal of Applied Social Psychology 32*(1), 185–206.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*. Elsevier.

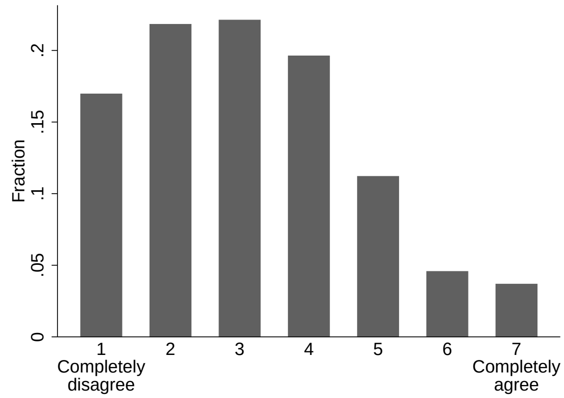
Stamarski, C. and L. Son Hing (2015). Gender inequalities in the workplace: The effects of organizational structures, processes, practices, and decision makers' sexism. *Frontiers in Psychology 6*, 1400.

Stinebrickner, R. and T. R. Stinebrickner (2006). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics 90*(8), 1435–1454.

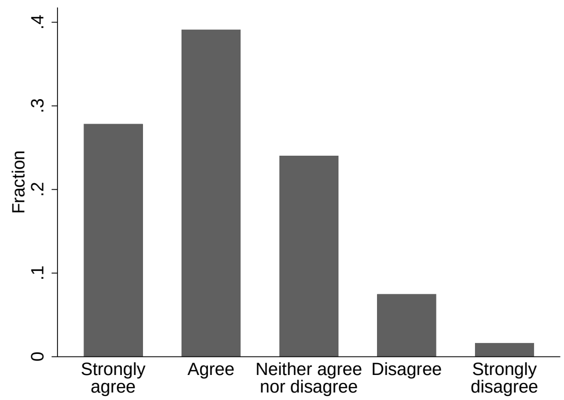
Van Laar, C., S. Levin, S. Sinclair, and J. Sidanius (2005). The effect of university roommate contact on ethnic attitudes and behavior. *Journal of Experimental Social Psychology 41*(4), 329–345.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *The Review of Economics and statistics 85*(1), 9–23.

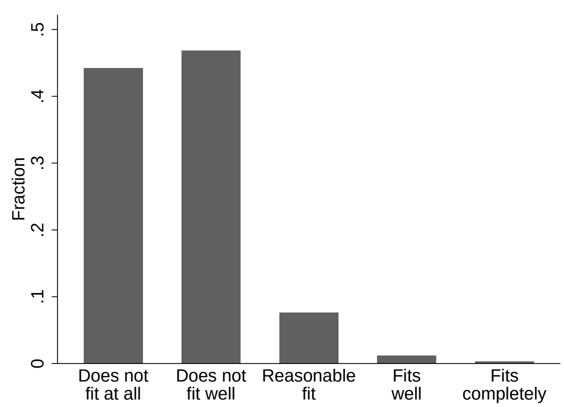
Figure 1. Distribution of gender attitudes for male recruits at baseline



A. Mixed-gender productivity: "A team performs better when it is made up of the same gender"



B. Gender roles: "It is important that men and women share household work equally"



C. Gender identity: "How well does the following statement describe you: I am feminine"

Notes: Responses to baseline survey conducted prior to boot camp, excluding missings. There are 678, 683 and 686 observations in panels A, B and C, respectively.

Table 1. Examples of gender segregated occupations in Norway and the U.S. in 2014

	% Female			% Female	
	Norway	U.S.		Norway	U.S.
Male dominated	(1)	(2)	Female dominated	(3)	(4)
Plumbers	2	2	Dental assistants	97	97
Firefighters	6	6	Pre-K / K teachers	92	97
Aircraft pilots	6	7	Hair dressers	90	94
Truck drivers	10	6	Registered nurses	89	90
Military	**13**	**15**	Social workers	84	82
Computer programmers	20	20	Pharmacists	76	56
Civil engineers	22	17	Primary school teacher	74	81
Geoscientists	31	25	Physical therapists	73	70
Architects	46	25	Librarians	71	84

Notes: Fractions for Norway from authors' tabulations of register data. Fractions for the U.S. from the U.S. Bureau of Labor Statistics, BLS Reports: Women in the labor force: A databook (2015).

Table 2. OLS estimates of gender attitudes

	"Same gender teams perform better" *Disagree=1* (1)	**"Important to share HH work equally"** *Agree=1* (2)	**"I am feminine"** *Statement does not fit me at all=0* (3)
Female friends	.0845**	.0185	.0310
	(.0263)	(.0258)	(.0271)
Has a sister	.0141	.0636**	-.0025
	(.0320)	(.0314)	(.0330)
Has a brother	-.0308	.0081	.0037
	(.0335)	(.0328)	(.0344)
High muscle strength	-.1088**	.0273	-.0529**
	(.0261)	(.0256)	(.0268)
High GAI test score	.0463	-.0844**	.0882**
	(.0259)	(.0254)	(.0267)
Mother higher education	.0000	.0167	.0347
	(.0271)	(.0266)	(.0279)
Father higher education	.0076	-.0364	.0219
	(.0275)	(.0271)	(.0284)
Mother works	-.0666	-.0205	.0215
	(.0455)	(.0448)	(.0472)
Parents divorced or separated	-.0122	-.0126	.0476*
	(.0277)	(.0272)	(.0285)
R-square	.031	.019	.018
[p-value]	[.000]	[.049]	[.066]
N	1,430	1,439	1,442

Notes: Sample includes all male recruits from the baseline survey, including recruits in battalions which both did and did not participate in the experiment. Dummy variables for missing values of control variables are also included in the regressions. Female friends is a dummy for 40% or more of one's friends being female. High muscle strength is a dummy for above average or far above average muscle strength. GAI stands for general ability index, and is based on verbal comprehension and perceptual reasoning tests; high GAI test score is a dummy for scoring at or above the 6th stanine.
***p-value<.05, *p-value<.10.*

Table 3. Experimental estimates of gender attitudes

	"Same gender teams perform better" *Disagree=1*		**"Important to share HH work equally"** *Agree=1*		**"I am feminine"** *Statement does not fit me at all=0*	
	(1)	(2)	(3)	(4)	(5)	(6)
Female on team	.1487**	.1333**	.0770*	.0821**	.1359**	.1382**
	(.0503)	(.0525)	(.0396)	(.0407)	(.0508)	(.0507)
Controls	No	Yes	No	Yes	No	Yes
R-square	.193	.221	.220	.260	.355	.386
Dependent mean	.546	.546	.667	.667	.649	.649
N	522	522	526	526	538	538

Notes: Sample includes all male recruits in battalions which participated in the experiment. The outcome variables come from the endline survey, which is taken at the end of boot camp. All regressions include indicators for troop, since randomization of men and women to squads (i.e., rooms) occurs within troops, as well as the corresponding lagged attitude measured in the baseline survey (as a fully saturated set of dummy variables for the possible answers). Controls are listed in Appendix Table A1. Standard errors clustered by room.

***p-value<.05, *p-value<.10*

Table 4. Robustness and heterogeneity

	"Same gender teams perform better" *Disagree=1* (1)	**"Important to share HH work equally"** *Agree=1* (2)	**"I am feminine"** *Statement does not fit me at all=0* (3)
A. Baseline model	.1333**	.0821**	.1382**
	(.0525)	(.0407)	(.0507)
N	522	526	538
B. Probit, marginal effect	.1581**	.0968*	.1638**
	(.0600)	(.0474)	(.0543)
N	522	520	538
C. First difference model	.1953**	.1231**	.1209**
	(.0715)	(.0491)	(.0527)
N	472	478	492
D. Nonmissing baseline attitude	.1410**	.0747*	.0980*
	(.0585)	(.0417)	(.0519)
N	472	478	492
E. No controls for baseline attitude	.1180**	.0517	.1076*
	(.0535)	(.0476)	(.0638)
N	522	526	538
F. One female	.1340	.1296	.1621**
	(.0914)	(.1262)	(.0799)
Two or more females	.1390**	.0606	.1155**
	(.0561)	(.0426)	(.0535)
N	522	526	538
G. All room sizes	.1144**	.0706*	.1304**
	(.0515)	.0405	(.0485)
N	539	543	556
H. Room size=6 & females=0 or 2	.1530**	.1029*	.1619**
	(.0776)	(.0533)	(.0627)
N	396	399	406
I. Female on team	.1307**	.0800**	.1393**
	(.0535)	(.0397)	(.0541)
FOT * Share of women in troop	.0176	.0130	-.0066
	(.0680)	(.0511)	(.0799)
N	522	526	538

*Notes: Regressions mirror those in Table 3. Panel F estimates separate coefficients for 1 versus 2 or more females in the room. Panel G adds in rooms with less than 5 or more than 7 members in them. Panel H limits the sample to rooms with 6 members and either 0 or 2 females. Standard errors clustered by room. **p-value<.05, *p-value<.10*

Table 5. Test of random assignment

	Female on team
	(1)
Female friends	-.0120
	(.0313)
Has a sister	-.0450
	(.0383)
Has a brother	.0464
	(.0400)
High muscle strength	-.0130
	.(0289)
High GAI test score	.0007
	(.0283)
Mother higher education	.0282
	(.0324)
Father higher education	-.0058
	(.0331)
Mother works	.0330
	(.0552)
Parents divorced or separated	.0048
	(.0337)
Joint F-statistic	.77
[p-value]	[.726]
Dependent mean	.282
N	781

Notes: Sample includes male recruits in battalions which participated in the experiment and who were assigned to rooms with between 5 and 7 members. The regression includes indicators for troop, since randomization of men and women to squads (i.e., rooms) occurs within troops. Dummy variables for missing values of the control variables are also included. The joint F-statistic has 17 degrees of freedom and is based on the controls shown in the table plus the dummies for missing values (there are 17 instead of 18 degrees of freedom, because a missing value for mother higher education is perfectly correlated with a missing value for father higher education).

***p-value<.05, *p-value<.10*

Table 6. Placebo tests: Are prior attitudes affected by future exposure?

	"Same gender teams perform better" *Disagree=1* (1)	**"Important to share HH work equally"** *Agree=1* (2)	**"I am feminine"** *Statement does not fit me at all=0* (3)
Female on team	-.0621	-.0056	-.0497
	(.0525)	(.0460)	(.0532)
Controls	Yes	Yes	Yes
R-square	.064	.087	.058
Dependent mean	.609	.669	.558
N	678	683	686

Notes: Regressions mirror those in Table 3, but use attitudes measured in the baseline survey (before bootcamp began). Since attitudes are not measured prior to the baseline, these regressions do not control for lagged attitudes. Standard errors clustered by room.
***p-value<.05, *p-value<.10.*

Table 7. Experimental estimates of female leadership attitudes

	"Which sex makes the best troop leaders?"	**"Which sex makes the best leaders at the highest level?"**	**"Which sex makes the best leaders for foreign operations?"**
		Equally good or *Females=1*	
	(1)	(2)	(3)
Female on team	-.0097	-.0061	-.0115
	(.0493)	(.0479)	(.0434)
Controls	Yes	Yes	Yes
R-square	.281	.259	.333
Dependent mean	.477	.649	.459
N	528	524	519

Notes: Regressions mirror those in Table 3. Standard errors clustered by room.
***p-value<.05, *p-value<.10.*

Table 8. Experimental estimates of military recruiting, preparedness and satisfaction outcomes

	"Want to do military service"	**"Feel qualified for further service"**	**"Plan to continue in the military"**	**"Satisfaction with boot camp"**	**"Satisfaction with room"**
	Strongly agree=1	*Strongly agree=1*	*Yes=1*	*Good=1*	*Good=1*
	(1)	(2)	(3)	(4)	(5)
Female on team	.0028	-.0512	.0213	.0007	.0004
	(.0425)	(.0546)	(.0356)	(.0310)	(.0439)
Controls	Yes	Yes	Yes	Yes	Yes
R-square	.273	.254	.515	.078	.070
Dependent mean	.603	.521	.223	.888	.842
N	521	528	534	534	533

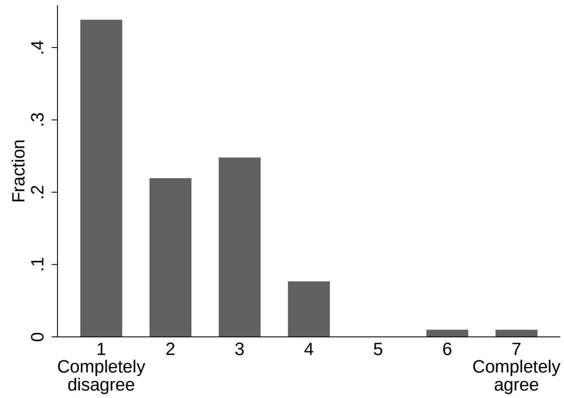
Notes: Regressions mirror those in Table 3. Outcomes in columns 4 and 5 are only measured in the endline survey, and so these regressions do not control for lagged attitudes. Standard errors clustered by room.
***p-value<.05, *p-value<.10.*

Appendix Figures and Tables

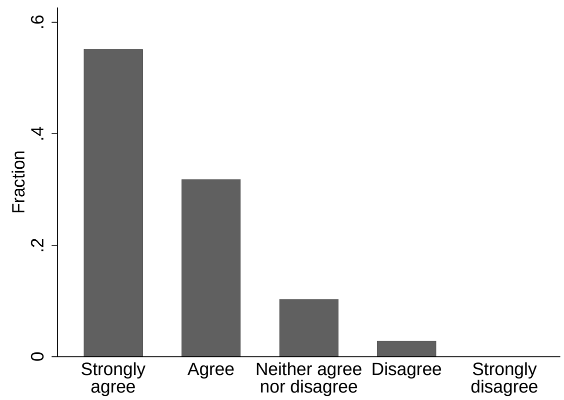
Does Integration Change Gender Attitudes?
The Effect of Randomly Assigning Women to Traditionally Male Teams

Gordon B. Dahl, Andreas Kotsadam and Dan-Olof Rooth

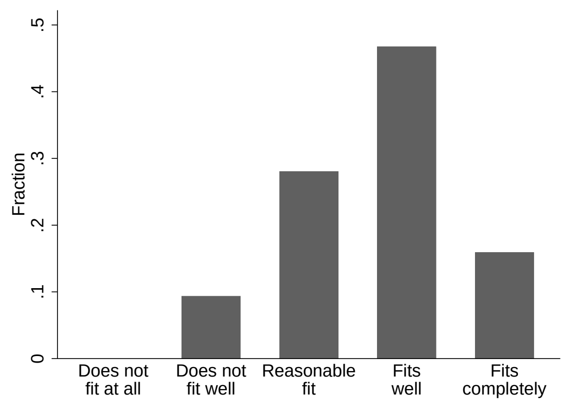
Appendix Figure A1. Distribution of gender attitudes for female recruits at baseline



A. Mixed-gender productivity: "A team performs better when it is made up of the same gender"



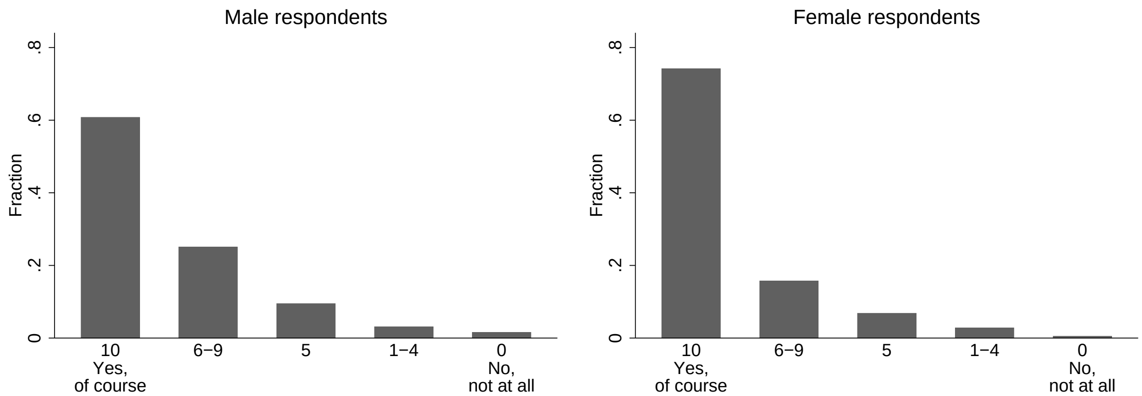
B. Gender roles: "It is important that men and women share household work equally"



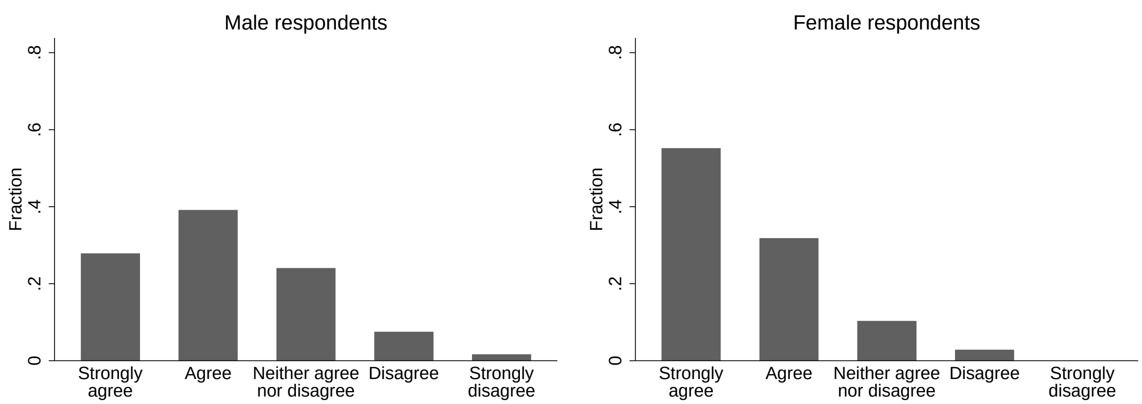
C. Gender identity: "How well does the following statement describe you: I am feminine"

Notes: Responses to baseline survey conducted prior to boot camp, excluding missings. There are 105, 107 and 107 observations in panels A, B and C, respectively.

Appendix Figure A2. Comparison of attitudes in the general population versus in the military on the importance of sharing household work



A. Gallup question on gender roles: "It is important that women and men share responsibility for the household"



B. Military question on gender roles: "It is important that men and women share household work equally"

Notes: Panel A presents aggregated results from nationwide surveys conducted by Jakobsson and Kotsadam in conjunction with Gallup in 2008, 2009 and 2010. Panel B presents results from our sample of all military recruits, based on their attitudes at the start of bootcamp in 2014. There are 2,889 male and 3,211 female respondents in the Gallup data and 1,439 male and 234 female respondents in our military data.

Appendix Table A1. Summary statistics for men and women at baseline

	Full sample		Experimental sample	
	Men (1)	Women (2)	Men (3)	Women (4)
Opposite sex friends	.321	.408	.330	.370
Share missing	*.158*	*.111*	*.128*	*.101*
Has a sister	.505	.531	.521	.571
Share missing	*.299*	*.252*	*.279*	*.252*
Has a brother	.533	.573	.561	.538
Share missing	*.296*	*.263*	*.270*	*.311*
High muscle strength	.496	.485	.540	.496
Share missing	*.042*	*.080*	*.045*	*.084*
High GAI test score	.484	.290	.510	.252
Share missing	*.010*	*.031*	*.012*	*.042*
Mother higher education	.450	.447	.515	.412
Share missing	*.159*	*.115*	*.129*	*.101*
Father higher education	.527	.546	.575	.521
Share missing	*.159*	*.115*	*.129*	*.101*
Mother works	.767	.813	.796	.815
Share missing	*.157*	*.111*	*.129*	*.101*
Parents divorced or separated	.276	.286	.273	.286
Share missing	*.160*	*.118*	*.129*	*.109*
N	1,697	262	781	119

Notes: The samples in columns 1 and 2 include recruits in battalions which both did an did not participate in the experiment. The samples in columns 3 and 4 only include recruits in battalions which participated in the experiment. Opposite sex friends is a dummy for 40% or more of one's friends being of the opposite sex. High muscle strength is a dummy for above average or far above average muscle strength. GAI stands for general ability index, and is based on verbal comprehension and perceptual reasoning tests; high GAI test score is a dummy for scoring at or above the 6th stanine.

Appendix Table A2. Heterogeneity by characteristics of males

FOT = female on team	**"Same gender teams perform better"** *Disagree=1* (1)	**"Important to share HH work equally"** *Agree=1* (2)	**"I am feminine"** *Statement does not fit me at all=0* (3)
A. Female friends			
Female on team	.0708	.0724	.1363**
	(.0671)	(.0495)	(.0644)
FOT * Female friends	.1284	.0062	-.0991
	(.1055)	(.0907)	(.0828)
B. Has a sister			
Female on team	.0129	.0611	.1185
	(.1026)	(.0851)	(.0925)
FOT * Has a sister	.1499	.0163	-.0138
	(.1294)	(.1019)	(.0995)
C. Mother higher education			
Female on team	.2014**	.0567	.1333*
	(.0869)	(.0739)	(.0683)
FOT * Mother higher education	-.1321	.0296	-.0553
	(.1077)	(.0941)	(.0788)
D. Muscle strength			
Female on team	.1611**	.0997*	.1637**
	(.0695)	(.0531)	(.0643)
FOT * High strength	-.0263	-.0268	-.0580
	(.0996)	(.0969)	(.0936)
E. GAI test score			
Female on team	.1811**	.0801	.1756**
	(.0644)	(.0663)	(.0682)
FOT * High GAI test score	-.0751	.0117	-.0699
	(.0897)	(.0898)	(.0792)
F. Baseline gender stereotype			
Female on team	.0939	.0080	.1137**
	(.0693)	(.0452)	(.0491)
FOT * Negative stereotype	.0641	.1733	-.0305
	(.1163)	(.0964)	(.0811)
Dependent mean	.546	.667	.649
N	522	526	538

*Notes: Regressions mirror those in Table 3, with the addition of interacting the treatment variable (female on team) with various dummy variables. In Panel F, negative baseline stereotype is defined using the relevant lagged gender attitude variable (i.e., the baseline survey responses). Standard errors clustered by room. **p-value<.05, *p-value<.10.*

Appendix Table A3. Testing for differential attrition

	"Same gender teams perform better"	**"Important to share HH work equally"**	**"I am feminine"**
	Missing outcome in endline survey=1		
	(1)	(2)	(3)
Female on team	-.0420	-.0360	-.0518
	(.0405)	(.0412)	(.0417)
Controls	Yes	Yes	Yes
R-square	.143	.138	.134
Dependent mean	.332	.327	.311
N	781	781	781

Notes: Regressions mirror those in Table 3. Standard errors clustered by room.
***p-value<.05, *p-value<.10.*