Exposure to female colleagues breaks the glass ceiling - Evidence from a combined vignette and field experiment*

Henning Finseraas†, Åshild A. Johnsen‡, Andreas Kotsadam§, and Gaute Torsvik¶

11th May 2015

Abstract

In a vignette experiment among recruits in the Norwegian Armed Forces, female candidates are valued as less fit to be squad leaders than their identical male counterparts. Adding positive information leads to higher evaluations of the candidates, but does not reduce the amount of discrimination. Male soldiers randomly assigned to share room and work in a squad with female soldiers during the recruit period, do not discriminate. We argue that the boot camp is an ideal setting for the contact theory of majority-minority relations, and that intense collaborative exposure to female colleagues reduces discriminatory attitudes.

*Acknowledgments: The paper has benefited from comments from Sara Cools, Raquel Fernandez, Arnfinn H. Midtbøen, and Øyvind Skorge as well as seminar participants at the University of Bergen, University of Oslo, Norwegian Social Research, Linnaeus University, and NOVA. We thank Ada Fuglset, Eirik Strømland, and Wiktoria Szczesna for excellent research assistance. Thanks also to the soldiers and staff at the North Brigade and FFI. The project is part of the research activities at the Centre for the Study of Equality, Social Organization, and Performance (ESOP) at the Department of Economics, University of Oslo. ESOP is supported by the Research Council of Norway.

†Institute for Social Research, henning.finseraas@samfunnsforskning.no

‡University of Stavanger Business School, ashild.a.johnsen@uis.no

§University of Oslo, andreas.kotsadam@econ.uio.no

¶University of Oslo, gaute.torsvik@econ.uio.no

1 Introduction

There are significant gender differences in labor market outcomes across the world. In particular, there is a large under-representation of women at higher levels of the corporate ladder. In Europe, women account for less that 12 percent of board directors, despite accounting for 45 percent of the labor force (Pande and Ford, 2011), and less than 15 percent of executive officers in US Fortune 500 companies are women (Bertrand et al., 2014). This vertical segregation is commonly referred to as *the glass ceiling*. Despite being considered a country with a high degree of gender equality, the glass ceiling is clearly evident also in Norway (Schwab et al., 2014).[1]

To increase our understanding of the role discrimination play for the glass ceiling, we ran a vignette experiment on recruits in the Norwegian Armed Forces. We let male soldiers, at the end of booth camp, evaluate a candidate for a squad leader position. A random sample of the soldiers evaluated a female candidate. Except for the gender specific name, the male and the female candidate had identical qualifications. As female soldiers (around 10 % of the soldiers are women) are randomly allocated to rooms during this period, a random sample of the male soldiers were "treated" with female soldiers as room mates and colleagues (a room defines a squad in the boot camp).

We find that the female candidate is ranked lower than the male candidate. Adding positive information (physical strength and leadership experience) to their resume improves the evaluations of the candidates, but does not reduce gender discrimination. However, living and working together with a female recruit in a squad, has a strong causal impact on the male soldiers perception of female leadership. Those who are treated with female soldiers in their squad, do not discriminate when they evaluate the squad leader candidate.

The primary contribution of the paper is to the discrimination literature (Guryan and Charles (2013), Midtbøen (2014), Pager (2007), Riach and Rich (2002), and Rich (2014) provide extensive surveys of the literature). Most importantly, we move beyond merely identifying discrimination. We examine how exposure to those discrimination is targeted against affects discrimination. We contribute to the literature on peer effects by testing how exposure to female

[1] In Norway the gender gap in wages is 50 percent higher among college graduates than among full time working men and women in general, and before quotas were introduced in corporate boards only 5 percent of board members were women (Bertrand et al., 2014).

colleagues affects attitudes towards women that wants to take a leader position. We examine exposure in a context that is close to ideal with respect to the conditions for bias reduction specified in inter-group contact theory, namely one of equal status, common goals, cooperation, and enforcing authority (Allport, 1954; Pettigrew, 1998). We identify the peer effects by separating men and women in the estimation, and hence follow recent advances in the peer-effects literature (Angrist, 2014).

Random exposure to female village leaders in India (Beaman et al., 2009), and to black roommates in college (e.g. Boisjoly et al. 2006) or in the US Air Force (Carrell et al. 2015), has been shown to reduce bias. To the best of our knowledge, however, the question of whether peer exposure to female colleagues reduces the amount of discrimination has not been tested before. Finally, previous literature has identified a clear pattern, whereby gender discrimination covaries positively with the gender composition of the sector of employment (Albert et al. 2011; Booth and Leigh 2010; Carlsson 2011; Correll et al. 2007; Firth 1982; Guryan and Charles 2013; Levinson 1975; Neumark et al. 1996; Petit 2007; Riach and Rich 1987, 2006; Rich 2014; Weichselbaumer 2000; Zhou et al. 2013). The Norwegian Armed Forces have fewer women in top positions than any other Norwegian sector, including the church (Teigen, 2014). Hence our results are of interest in order to understand the advancement of women in a hyper male setting.

2 Exposure and discrimination

The discrimination literature often acknowledges that exposure to the discriminated group is potentially important, but empirical tests are troublesome due to non-random variation in exposure. Experimental analyses are sometimes combined with observational data on attitudes or the ethnic mix of the area (e.g. Doleac and Stein 2013; Ewens et al. 2014), however, self-selection into areas implies that these studies do not necessarily capture the effect of exposure. We avoid such selection problems since we randomly assign males to share room with female soldiers during the first eight weeks of their military service. The room is an important unit during this period. Apart from living together, roommates solve a number of tasks together and operate as a team within the platoon.

How do we expect that this type of exposure will affect discrimination in our case? A handful of studies have found that exposure to peers with other

characteristics reduce biased perceptions. Boisjoly et al. (2006) find that white students who were randomly assigned to live with an African-American in college were more positive towards African-Americans and towards affirmative action, than were white students who had white roommates. Carrell et al. (2015) find that white freshman cadets at the U.S. Air Force Academy become more positive towards blacks if randomly assigned to squadrons with black students, and Van Laar et al. (2005) find improved inter-group attitudes among college students using randomized exposure.

Theoretically, exposure to the discriminated group may decrease or increase bias, and its effect is likely to depend on the type of exposure and the setting in which contact takes place. If exposure takes place in a competitive environment bias is more likely to increase (e.g. Semyonov et al. 2006). The so called inter-group contact theory (Allport, 1954; Pettigrew, 1998) argues that prejudice and negative stereotyping of minorities might decline with contact with out-group members, but only under some conditions: Contact will reduce discrimination only if those in contact have equal status in the particular context, if they share common goals, if they are in a cooperative context, and if the contact takes place under some form of authority (see Pettigrew (1998)). The setting in which our soldiers interact makes us therefore expect contact to decrease bias: Soldiers of private rank have equal social status within the army, they share the common goals of the unit, they need to cooperate to solve their tasks, and contact takes place in a context with an explicit, enforcing authority. In fact, the army explicitly promote views of unity and equality among soldiers of the same rank.

There are different reasons why exposure could reduce discrimination. The most straightforward mechanism is that exposure leads to experiences that makes men update their beliefs about the suitability of female leaders (as suggested by Carrell et al. (2015) to be the reason why interracial exposure reduces bias). Other reasons may be linked to identity, homosociality, and critical mass. Norms about gender differences are salient in leadership perceptions in male dominated settings (Ridgeway and Correll, 2004), and as people tend to favor leaders that are similar to themselves, a self-fulfilling process of homosocial reproduction may occur (Kanter, 1977a,b). A qualitative field study of gender mixed rooms, conducted after the boot-camp period from two camps in Northern Norway (including the camps of our soldiers) concludes that mixed rooms reduces gender essentialist notions and increases feelings of sameness among the soldiers (Hellum, 2015). Hence, it is possible that intense exposure makes male

soldiers perceive themselves as more similar to female soldiers and therefore less skeptical to having them as leaders. Another mechanism that may potentially reduce discrimination is reduced tokenism as under-representation of women in the group may lead to them being viewed as symbols or tokens (Kanter, 1977a,b). Such a perspective is, however, related to non-linear effects of the amount of exposure, and previous research suggests a critical mass, whereby the perspective of the minority members and the nature of the relations in the group change qualitatively as the minority grows from a few token individuals into a considerable minority (Kanter, 1977a,b; Dahlerup, 1988). Where and if such a cutoff exists necessarily depends on the context, but it is generally thought to lie at around 30 percent representation (Dahlerup, 1988).

3 The Experiments

3.1 The Field Experiment

Military service is mandatory for men in Norway, but conscription is based on need, and only about one in six men are needed in duty. Since 2010, screening and testing for military service has been mandatory for both sexes, but women serve on a voluntary basis. Hence, both the men and the women are selected based on ability and motivation, and the women more so. Our sample includes all incoming soldiers in the August 2014 contingent to the The Second Battalion of the North Brigade of the Norwegian Armed Forces.

The soldiers met on their first day of service at a military camp close to Oslo. They were tested for medical and psychological fitness and flown to Northern Norway if they passed the tests. We had a session with a questionnaire to the soldiers during this day, it included questions on motivation, intentions to attend higher education, as well as a set of background characteristics. The soldiers were not told the purpose of the study. The instructor stressed that the survey results are merely used for research purposes and that it is anonymous to all representatives of the armed forces. At the point of testing, the soldiers had not met before, and they did not know with whom they would share rooms until they arrived to the camp in Northern Norway. Hence, the first survey constitutes the baseline data for the field experiment.

In Northern Norway soldiers were immediately assigned to rooms they would live in during the whole recruit period. The room assignment followed a procedure that randomly assigned individuals within platoons to rooms following

a decision rule that assigns at least two women to the same room if possible. Together with their room-mates, they perform various tasks, such as cleaning the room for inspection each morning. They also serve in the same platoon, and they constitute a squad within the platoon. There are strict rules for what soldiers can and cannot do during the boot camp; they have to wear uniforms at all times and are not allowed to sleep outside of the base.

After eight weeks we surveyed the soldiers a second time and link their answers to the first round using an anonymous reference number for each soldier. At this time we conducted a vignette experiment to test if there are discriminatory attitudes toward female soldiers, and to test if exposure reduces the amount of discrimination.

3.2 The Vignette Experiment

To detect discrimination we constructed a hypothetical (but realistic) case description of a candidate applying for a position as squad leader. We choose a squad leader position,since it is relevant for the soldiers in question.

The soldiers were asked to rate a fictional candidate on a scale from 1 to 6 based on a short text, presented in Table 1. The experiment had four between-subject treatments. The treatments differ with respect to the gender of the candidate, and in how much information they receive about the candidate. In the first treatment ("Ida basic"), the soldiers are provided with basic information about the female Ida Johansen: They get information about her high school grades, career plans, family background, and motivation. The second treatment ("Ida more") provides more information about the candidate: In addition to the basic information, they get information about her physical capacity and her leadership experience (in bold text). The other two treatments are identical to "Ida basic" and "Ida more", with the exception that the female name Ida Johansen is replaced by the male name Martin Hansen. The forenames are gender specific, and to avoid any name effects, we chose, as in Carlsson and Eriksson (2014), the most common names of the soldiers' age group. The surnames are the most common in Norway (Statistics Norway, 2014).

Table 1: From the instructions.

SQUAD LEADER
The unit is choosing new squad leaders. The squad leader is the link between officers and soldiers. For some, this position can be very physically and mentally demanding. The position requires high skills. As squad leader, one is responsible not just for oneself, but also for the team.

A potential candidate

Name: **Ida Johansen/ Martin Hansen**

- Grades from high school: 4.1 (average).
- Career plans: Does not wish to continue in the armed forces, plans to pursue higher education in the field of economics and administration.
- Family background: Has a sister, dad is an engineer, and mother is a teacher. Comes from a middle-sized city in the eastern part of Norway.
- Motivation: Thinks that serving in the armed forces is both meaningful and important.
- **Physical capacity: Among the top 20 percent in his/ her cohort (armed forces). Exercise regularly.**
- **Leadership experience: Was the leader of a youth organization.**

Ida Johansen/ Martin Hansen would very much like to become a squad leader, indicate how well suited you think he/ she is for the job: (1=very badly, 6=very well) - put a circle around your choice.
1 2 3 4 5 6

We ran the experiment on September 26, 2014, and in total 413 people participated in eight sessions. Session size varies depending on the size of the room and on when the soldiers were available for answering the questionnaire and the vignette, see Table 2. The experiment was conducted on a military base, and soldiers used pen and paper to answer the questions.

Table 2: Session Statistics

Session	Room	Subjects	Males	Males living with females
1	Small	31	29	6
2	Big	29	27	5
3	Big	65	63	11
4	Big	70	67	21
5	Small	23	22	4
6	Big	107	92	15
7	Small	33	29	7
8	Small	40	38	9
Total	8 sessions	398	367	78

Note: Observations with missing or erroneous information about rooms or candidate number were dropped (15 observations).

In all the following tables we restrict the sample to men only, for reasons that are explained in Section 4.[2] In Table 3, we see the raw difference across cases in how they are evaluated as squad leaders. The male candidate is evaluated as better than the female candidate. More information leads to more positive evaluations, but it does not reduce discrimination. There are no statistically significant differences in the background characteristics across the four treatments. As we also present results with the treatments pooled by gender we present the differences across the pooled cases in Table 4, and we see that they are only statistically significantly different with respect to one background variable (mothers education) and only at the 10 percent level. We will present regression results both with and without controlling for the background variables.

[2]The results from the basic vignette experiment, as well as summary statistics for the whole sample, are presented and discussed in the Appendix.

Table 3: Descriptive statistics across assigned cases

	(1) Ida basic		(2) Martin basic		(3) Ida more		(4) Martin more	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Dependent variable								
Score on the candidate (1=very bad, 6=very good)	3.771	(1.004)	4.145	(0.926)	4.376	(0.893)	4.720	(0.817)
Background characteristics								
Mother has high education	0.763	(0.428)	0.620	(0.488)	0.707	(0.458)	0.685	(0.467)
Father has high education	0.882	(0.325)	0.797	(0.404)	0.837	(0.371)	0.815	(0.390)
Mother works	0.855	(0.354)	0.886	(0.320)	0.868	(0.340)	0.902	(0.299)
Father works	0.947	(0.225)	0.962	(0.192)	0.989	(0.105)	0.978	(0.147)
Parents are divorced	0.276	(0.450)	0.253	(0.438)	0.366	(0.484)	0.253	(0.437)
Plan higher education	0.750	(0.436)	0.633	(0.485)	0.774	(0.420)	0.750	(0.435)
IQ	5.795	(1.488)	5.602	(1.306)	5.810	(1.555)	5.687	(1.353)
N (on dependent variable)	83		83		101		100	

Table 4: Descriptive statistics across the pooled cases

	(1) Ida		(2) Martin	
	Mean	SD	Mean	SD
Dependent variable				
Score on the candidate (1=very bad, 6=very good)	4.103	(0.989)	4.459	(0.912)
Background characteristics				
Mother has high education	0.732*	(0.444)	0.655	(0.477)
Father has high education	0.857	(0.351)	0.807	(0.396)
Mother works	0.862	(0.346)	0.895	(0.308)
Father works	0.970	(0.171)	0.971	(0.169)
Parents are divorced	0.325	(0.470)	0.253	(0.436)
Plan higher education	0.763	(0.426)	0.696	(0.461)
IQ	5.803	(1.521)	5.648	(1.329)
N (on dependent variable)	184		183	

4 Empirical strategy

4.1 Detecting discrimination and separating between types of discrimination

To test if female candidates are given a lower score than male candidates we estimate the following equation:

$$Score_{irt2} = \alpha_J + \gamma_S + \beta_1 Female\,candidate_{ir} + \beta_n X_{irt1} + \epsilon_{ir} \quad (1)$$

Where the variable Score for individual i in room j at time $t2$ is the score given to the candidate in the vignette. α_J refers to platoon fixed effects, γ_S are session fixed effects, $Female\,candidate$ is a dummy variable that equals 1 if Ida is the candidate, and 0 if Martin is the candidate evaluated. Hence, β_1 is our measure of discrimination against women. X is a vector of control variables collected at baseline (including parents' education, marital status, and employment status, the individuals' plans for taking higher education, and IQ score). We present results with and without these controls.

The fact that our outcome variable is the score from one to six implies that we circumvent the well-known problem in correspondence studies that potential differences in variance across groups in combination with a cutoff value for hiring renders discrimination unidentified (Heckman and Siegelman, 1993; Heckman, 1998; Neumark, 2012). Guryan and Charles (2013) offer the example of a hiring situation where candidates are matched at a level that is low relative to the hiring threshold. In such a case, the more heterogeneous group will have a higher share exceeding the threshold for hiring. In the opposite case, where candidates are matched at a relatively high level, the less heterogeneous group will have a higher share exceeding the threshold (see Carlsson et al. (2014) for a very illustrative graphical exposition of the problem).[3] In our case, the evaluation of the candidate is linear on a scale from one to six and hence we can recover the average evaluation without having to depend on a cutoff value.

Adding more information has become a standard way to try to identify sta-

[3]Neumark (2012) develops a method to work around this problem by testing how the discrimination changes when adding more positive information to the candidates. Crucially, this added information must be assumed to be equally correlated with the perceived productivity of the candidates across the groups. Such an assumption is often difficult to defend empirically and it further precludes a separation of the discrimination into taste based and statistical discrimination.

tistical discrimination (e.g. Ahmed et al. 2010; Andersson et al. (2012); Ewens et al. 2014; List 2004; Zussman, 2013; and Doleac and Stein 2013) . If there is less discrimination when more positive attributes are added we can conclude that the discrimination is statistical with respect to those attributes. We follow this approach and add positive information and use a difference in difference approach in order to identify statistical discrimination. This is done by estimating equation 2:

$$\begin{aligned} Score_{irt2} &= \alpha_J + \gamma_S + \beta_1 Female\,candidate_{ir} + \beta_2 More\,information_{ir} \\ &\quad + \beta_3 Female\,candidate * More\,information_{ir} + \beta_n X_{irt1} + \epsilon_{ir} \end{aligned} \tag{2}$$

Where $More\,information_{ir}$ is a dummy variable equal to 1 if the respondent receives the treatment with added information. The interaction term, $Female\,candidate_{ir} * More\,information_{ir}$, captures the difference in the effect of added information for the female and the male candidate.

4.2 Exposure

Peer effects interest social scientists across a range of disciplines (see Sacerdote, 2011, for a review of the literature). The notion that people are affected by other people is commonly held, yet it turns out to be difficult to prove because people generally self select into networks and the whole group usually face a similar environment (Manski, 1993). We solve this by randomly allocating soldiers into rooms. Angrist (2014) argues that the most compelling evidence of peer effects comes from studies whereby there is a clear separation of the individuals thought to be affected and the peers thought to affect them.[4] For this reason, we limit the sample to only men. The women will only be used to define the room characteristics. The following regression model will be estimated:

[4] This type of design is applied in Kling et al. (2007), who analyze the effects of neighborhoods on individuals randomly assigned to receive housing vouchers in the Moving to Opportunity program. The neighborhood effects are estimated by using characteristics of the neighbors, and no effects on the old neighbors are estimated. Similarly, Angrist and Lang (2004) investigate the effects of low-income peers in the classroom, where low-income individuals were bussed in as part of the Metco program. The low income students' own outcomes were not included in the regression, they were only used to calculate peer characteristics.

$$Score_{irt2} = \alpha_J + \gamma_S + \beta_1 Room\,Treatment_r + \beta_n X_{irt1} + \epsilon_{ir} \quad (3)$$

Where $Score_{irt2}$ refers to the Score given by individual i in room r at time t2. α_J refers to the platoon fixed effects, γ_S are session fixed effects, while β_n is the vector of coefficients of the covariates. In this specification, β_1 captures the causal effect of being assigned to a room with at least one woman. In order to investigate treatment intensity, we run regressions with the share of women in the room, and we also investigate non-linearities. Since we compare soldiers within the same platoon, but with different treatment status at the room level, it is possible that there are spillovers such that being exposed to women in the platoon also affects discriminatory attitudes. Hence, the effect we measure is the difference between intense exposure at the room and team level in addition to any effect of exposure at the platoon level. We expect that the spillover reduces the estimated effect. Hence, our results are lower bound estimates of intense exposure to female colleagues.

5 Results

5.1 Evidence for discrimination

As we saw in Table 3, the female candidate with little information receives the lowest score, while the male candidate with more information receives the highest score. We now test these differences formally. We start by asking if there is discrimination of female candidates, by regressing the score on the gender of the candidate, as described in equation 1. In column 1 of Table 5, we see that the woman is evaluated as less fit to be a squad leader. Hence, there is discrimination of the female candidate by men in our sample. The coefficient for female candidate in column 1 captures the combined effect across the cases with more and less information. In column 2, we add baseline controls, the results are similar.

Adding positive information about the candidates we can test if discrimination is statistical based on the added information. Column 3 shows the difference in difference results where we separate the cases with and without information, as described in equation 2. We find that while information improves the evaluation of both our male and female candidates, it does not reduce the degree of discrimination. The interaction term is negative, implying that if anything information helps the male candidate more, but the coefficient is not statisti-

cally significant. Adding baseline controls (column 4) yields similar results. The results are consistent with the discrimination being taste based, or with discrimination being statistical with respect to other characteristics not included in the vignette, but inconsistent with discrimination being statistical with respect to the information added.

Table 5: Gender discrimination: Dependent variable is score of the candidate.

	(1)	(2)	(3)	(4)
	Pooled		Less and more information	
Female candidate	-0.326***	-0.365***	-0.275*	-0.318**
	(0.108)	(0.103)	(0.140)	(0.143)
Information added			0.551***	0.456***
			(0.134)	(0.135)
Female candidate*Information			-0.109	-0.096
			(0.166)	(0.176)
Mean of dependent variable	4.281	4.307	4.281	4.307
Observations	367	335	367	335
R-squared	0.128	0.191	0.190	0.232
Platoon and Session FE	Yes	Yes	Yes	Yes
Baseline controls	No	Yes	No	Yes

Notes: The sample only includes male respondents.
Standard errors clustered at the room level in parantheses.
*** p<0.01, ** p<0.05, * p<0.1

5.2 Exposure reduces discrimination

When we now move on to test whether random variation in exposure to female soldiers reduces discrimination, the analysis is restricted to men, and the female peers merely inform the treatment status. In total we have 89 rooms, with four to eight persons in each room. Eight percent of the soldiers are women and between zero and four women live in the rooms. The share of women in the rooms ranges from 0-0.67 with a mean of 0.07 and a standard deviation of 0.15. In total, 21 percent of the men are treated, i.e. they share room with at least one woman. The share of exposure for the treated varies from 17 to 67 percent and the full distribution of treatment shares is shown in Table 6. We see that the modal exposure rate is to have one-third of the soldiers in the room being female, conditional on having at least one women in the room.

Table 6: Share of women in the room for treated soldiers.

Share of women in room	Number of exposed men	Percent
17 % women in the room	5	6.41
20 % women in the room	4	5.13
25 % women in the room	11	14.10
29 % women in the room	9	10.26
33 % women in the room	35	44.87
50 % women in the room	13	16.67
67 % women in the room	2	2.56
Total	78	100

In Table 7, we present coefficients and t-statistics from regressions of the treatment indicator dummy on the pre-determined variables. Platoon fixed effects are included in all regressions since room assignment is randomized within platoons, and session fixed effects are included since the vignette experiment took place within 8 different sessions. The table also reports an F-test of joint significance. The differences between the treatment and the control group are small and not statistically significant. Most importantly, the small F-value in the joint test of whether all variables together predict treatment status allows us to conclude that the randomization was successful.

Table 7: Regressions of treatment status on pre-determined variables.

	Coefficient	t-statistic
Mother has high education	0.020	0.489
Father has high education	0.003	0.081
Mother is employed	0.023	0.517
Father is employed	-0.039	-0.476
Parents are divorced	0.017	0.319
Plan to take higher education	0.005	0.138
IQ	0.007	0.544
F-test of joint significance	0.02 (p=.90)	

Note: Each row presents the results from one regression. Platoon and session fixed effects are included in all regressions. t-values adjusted for room clustering. *** p<0.01, ** p<0.05, * p<0.1

In column 1 of Table 8, we present results based on regressing the score of the candidate on the gender of the fictitious candidate, a treatment dummy equal to 1 if sharing room with a woman, and treatment interacted with gender of the candidate. There is discrimination among men sharing room with only

men (as shown by the negative and statistically significant coefficient for female candidate not interacted with treatment). Men sharing room with women discriminate significantly less against women (as shown by the positive and statistically significant interaction term). These results show that the random intense, and relevant, exposure to women that comes from sharing room and forming a squad with them, not only reduces discrimination, but actually eliminates it. The results are similar when we add baseline controls in column 2, albeit the interaction term is only statistically significant at the 10 percent level. In column 3, we interact treatment also with information and the results indicate that exposure reduces discrimination most in combination with provided information albeit we do not have enough power to estimate the effects separately. Adding baseline controls yields similar results as seen in column 4.

Table 8: Exposure and discrimination: Dependent variable is score of the candidate.

	(1)	(2)	(3)	(4)
	Treatment		Information and Treatment	
Female candidate	-0.430***	-0.438***	-0.277	-0.305*
	(0.124)	(0.119)	(0.170)	(0.179)
Information added			0.657***	0.550***
			(0.153)	(0.162)
Female*Information			-0.254	-0.232
			(0.186)	(0.203)
Treated	-0.230	-0.201	0.085	0.062
	(0.145)	(0.141)	(0.221)	(0.226)
Treated*Female candidate	0.513**	0.358*	0.111	0.005
	(0.204)	(0.213)	(0.277)	(0.281)
Treated*Information			-0.493*	-0.437*
			(0.250)	(0.257)
Treated*Female candidate*Information			0.635	0.637
			(0.396)	(0.433)
Baseline controls	No	Yes	No	Yes
Mean of dependent variable	4.281	4.266	4.281	4.266
Observations	367	335	367	335
R-squared	0.139	0.196	0.204	0.242
Platoon and Session FE	Yes	Yes	Yes	Yes
Baseline controls	No	Yes	No	Yes

Notes: Standard errors clustered at the room level in parantheses.
*** p<0.01, ** p<0.05, * p<0.1

In Table 9 we exploit the variation in treatment intensity and regress the share of women in the room on the evaluation of the candidate as well as the interaction with evaluating the female candidate. In column 1 we see a large and highly statistically significant effect of the share of women in the room interacted with evaluating a female candidate. As the unit of measurement is one of going from zero women in the room to a situation where everyone in the room are women it implies an extrapolation as there is by definition no man in a room with only women. In column 2 we instead present standardized beta coefficients and the results show that a one standard deviation increase in the share of women (corresponding to an increase in the share of women of 15 percent) corresponds to a decline in the discrimination and an increase in the evaluation of the female candidate by 0.165 standard deviations.

In Table 10 we allow for non-linear treatment intensity by splitting the treatment variable into different categories by share of exposure. We interact each share of exposure with the dummy for evaluating the female candidate. We have to collapse the two highest shares of exposure, however, as only two men are exposed to 67 percent women in the room and none of them received the vignette with the female candidate. We see in column 1 that there are is a clear non-linear pattern whereby having the lowest share of exposure, with only 17 percent women in the room, actually has a negative effect on the discrimination of the female candidate. Having at least 20 percent women in the room, however, always leads to a decline in the discrimination of the female candidate, albeit not statistically significantly so for having 25 percent women in the room. These non-linear effects show that there is perhaps a need of a critical mass of women for the effects to materialize. This may be due to reduced tokenism as women become a sizable minority so that the men start evaluating female soldiers based on their individual characteristics rather than mainly as women, but it may also be an effect via homosocialization whereby men exposed to a high enough share of women feel that female soldiers are like themselves and thereby discriminate less against them. In the appendix we present results where we split the treatment variable into number of women in the room and it seems to be important to have at least two women in the room.

Table 9: Share of exposure and discrimination: Dependent variable is score of the candidate.

	(1) Baseline	(2) Standardized Beta coefficients
Female candidate	-0.434***	-0.225***
	(0.122)	(0.001)
Share of women	-0.567	-0.086
	(0.441)	(0.202)
Share of women*Female candidate	1.665***	0.165***
	(0.625)	(0.009)
Observations	367	367
R-squared	0.141	0.141
Platoon and Session FE	Yes	Yes

Standard errors are clustered on rooms in both regressions and are presented in parentheses in column 1. The parantheses in column 2 presents p-values (col 2). *** p<0.01, ** p<0.05, * p<0.1

Table 10: Non-linear effects of exposure on discrimination: Dependent variable is score of the candidate.

	(1)	(2)
	Score	Score
Female candidate	-0.437***	-0.433***
	(0.127)	(0.127)
Female candidate*17 % women in the room	-0.514***	-0.519***
	(0.185)	(0.185)
Female candidate*20 % women in the room	0.437***	0.433***
	(0.127)	(0.127)
Female candidate*25 % women in the room	0.097	0.094
	(0.138)	(0.138)
Female candidate*29 % women in the room	1.087***	1.077***
	(0.230)	(0.232)
Female candidate*33 % women in the room	0.759***	0.756***
	(0.231)	(0.232)
Female candidate*> 50 % women in the room	0.986*	0.985*
	(0.557)	(0.555)
17 % women in the room	0.133	0.135
	(0.274)	(0.274)
20 % women in the room	-1.006***	-0.886***
	(0.159)	(0.190)
25 % women in the room	-0.118	-0.128
	(0.130)	(0.126)
29 % women in the room	-1.006***	-1.131***
	(0.159)	(0.198)
33 % women in the room	-0.618***	-0.624***
	(0.133)	(0.133)
> 50 % women in the room	0.034	0.027
	(0.284)	(0.284)
Number of persons in room		0.122
		(0.112)
Constant	4.688***	4.003***
	(0.380)	(0.724)
Observations	367	367
R-squared	0.159	0.160
Platoon and Session FE	Yes	Yes

Notes: Standard errors clustered at the room level in parantheses.
*** p<0.01, ** p<0.05, * p<0.1

6 Conclusion

Fewer women than men reach higher levels of leadership, also in Norway (Bertrand et al., 2014), and especially in the military (Teigen, 2014). Such differences can be explained by supply side factors, such as differences in preferences and differences in competitiveness across the sexes (Croson and Gneezy, 2009). The differences may, however, also stem from demand side discrimination, i.e., that men are valued more highly than women with identical qualities and aspirations. Discrimination may be statistical, in the sense that it is based on unbiased statistical inference, or it may be preference based, so that it is driven by negative attitudes or biased perceptions of women's abilities. This paper aims to shed light on three important questions related to gender discrimination. First, to what extent are women seeking leadership positions in a male dominant environment discriminated against? Second, if women are discriminated, does additional information, revealing that the female and male applicant is equally suitable for the position, help? Third, will working together with women in a male dominant environment induce men to discriminate less against women aspiring for leadership roles?

Identifying the presence and the type of discrimination is difficult with observational data, as many of the factors that may influence the valuation of a candidate are not observed by the researcher. We use a vignette experiment to detect discrimination and find that male soldiers give a fictitious male candidate for a squad leader position higher rating than an otherwise identical female candidate. A random sample of the soldiers got additional information on qualifications of the candidates. Providing additional positive information increased the rating of both candidates but did, however, not reduce the degree of discrimination.

A novel element of this study is that a random sample of male soldiers was allocated to share rooms with female soldiers. By sharing rooms, they also share the responsibility for many different tasks and they often form a team within the platoon. We find that discrimination evaporates if we expose male soldiers to female peers in an environment that is relevant for the leader position. We further find that the share of exposure is important, and there appears to be a critical mass of women needed to disrupt the discriminatory behavior in this hyper male setting.

By combining a vignette experiment with a randomized field experiment, our results have strong internal validity. Previous literature finds discrimination of

women in male dominated spheres, and we believe our results are generalizable to such settings. The particular selection of men and especially of women in our setting is a fact in all male dominated settings. Limits to the external validity may arise, however, from other peculiarities of the military setting. The advantage of our context, in addition to the ability of establishing causality, is that we can derive the clear theoretical prediction that bias should be reduced. If our findings extend beyond the army setting they have important policy implications. We have shown that the glass ceiling that prevents female candidates to obtain leader positions in a masculine context can be broken by exposure.

References

Ahmed, A. M., L. Andersson and M. Hammarstedt, "Can discrimination in the housing market be reduced by increasing the information about the applicants?," *Land Economics* 86 (2010), 79–90.

Albert, R., L. Escot and J. A. Fernández-Cornejo, "A field experiment to study sex and age discrimination in the Madrid labour market," *The International Journal of Human Resource Management* 22 (2011), 351–375.

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

Andersson, L., N. Jakobsson and A. Kotsadam, "A field experiment of discrimination in the Norwegian housing market: Gender, class, and ethnicity," *Land Economics* 88 (2012), 233–240.

Angrist, J. D., "The perils of peer effects," *Labour Economics* 30 (2014), 98–108.

Angrist, J. D. and K. Lang, "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program," *American Economic Review* (2004), 1613–1634.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande and P. Topalova, "Powerful Women: Does Exposure Reduce Bias?*," *The Quarterly journal of economics* 124 (2009), 1497–1540.

Bertrand, M., S. E. Black, S. Jensen and A. Lleras-Muney, "Breaking the glass ceiling? The effect of board quotas on female labor market outcomes in Norway," Technical Report, National Bureau of Economic Research, 2014.

Boisjoly, J., G. J. Duncan, M. Kremer, D. M. Levy and J. Eccles, "Empathy or Antipathy? The Impact of Diversity," *American Economic Review* 96 (2006), 1890–1905.

Booth, A. and A. Leigh, "Do employers discriminate by gender? A field experiment in female-dominated occupations," *Economics Letters* 107 (2010), 236–238.

Carlsson, M., "Does hiring discrimination cause gender segregation in the Swedish labor market?," *Feminist Economics* 17 (2011), 71–102.

Carlsson, M. and S. Eriksson, "Discrimination in the rental market for apartments," *Journal of Housing Economics* 23 (2014), 41–54.

Carlsson, M., L. Fumarco and D.-O. Rooth, "Does the design of correspondence studies influence the measurement of discrimination?," *IZA Journal of Migration* 3 (2014), 1–17.

Carrell, S. E., M. Hoekstra and J. E. West, "The Impact of Intergroup Contact on Racial Attitudes and Revealed Preferences," Technical Report, National Bureau of Economic Research, 2015.

Correll, S. J., S. Benard and I. Paik, "Getting a Job: Is There a Motherhood Penalty? 1," *American journal of sociology* 112 (2007), 1297–1339.

Croson, R. and U. Gneezy, "Gender differences in preferences," *Journal of Economic literature* (2009), 448–474.

Dahlerup, D., "From a small to a large minority: women in Scandinavian politics," *Scandinavian Political Studies* 11 (1988), 275–298.

Doleac, J. L. and L. C. Stein, "The visible hand: race and online market outcomes," *The Economic Journal* 123 (2013), F469–F492.

Ewens, M., B. Tomlin and L. C. Wang, "Statistical discrimination or prejudice? A large sample field experiment," *Review of Economics and Statistics* (2014).

Firth, M., "Sex discrimination in job opportunities for women," *Sex Roles* 8 (1982), 891–901.

Guryan, J. and K. K. Charles, "Taste-based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots," *The Economic Journal* 123 (2013), F417–F432.

Heckman, J. J., "Detecting discrimination," *Journal of Economic Perspectives* (1998), 101–116.

Heckman, J. J. and P. Siegelman, "The Urban Institute audit studies: Their methods and findings," (1993).

Kanter, R. M., *Men and Women of the Corporation*, volume 5049 (Basic books, 1977a).

———, "Some effects of proportions on group life: Skewed sex ratios and responses to token women," *American journal of Sociology* (1977b), 965–990.

Kling, J. R., J. B. Liebman and L. F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica* 75 (2007), 83–119.

Levinson, R. M., "Sex discrimination and employment practices: An experiment with unconventional job inquiries," *Social Problems* 22 (1975), 533–543.

List, J. A., "The nature and extent of discrimination in the marketplace: Evidence from the field," *The Quarterly Journal of Economics* (2004), 49–89.

Manski, C., "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies* 60 (1993), 531–542.

Midtbøen, A., "Segregering og diskriminering," in M. Teigen and R. Liza, eds., *Kjønnsdeling og etniske skiller på arbeidsmarkedet* (Gyldendal, 2014).

Neumark, D., "Detecting discrimination in audit and correspondence studies," *Journal of Human Resources* 47 (2012), 1128–1157.

Neumark, D., R. J. Bank and K. D. Van Nort, "Sex Discrimination in Restaurant Hiring: An Audit Study*," *The Quarterly journal of economics* 111 (1996), 915–941.

Pager, D., "The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future," *The Annals of the American Academy of Political and Social Science* 609 (2007), 104–133.

Pande, R. and D. Ford, "Gender quotas and female leadership: A review," (2011).

Petit, P., "The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector," *Labour Economics* 14 (2007), 371–391.

Pettigrew, T. F., "Intergroup Contact Theory," *Annual Review of Psychology* 49 (1998), 65–85.

Riach, P. A. and J. Rich, "TESTING FOR SEXUAL DISCRIMINATION IN THE LABOUR MARKET*," *Australian Economic Papers* 26 (1987), 165–178.

———, "Field Experiments of Discrimination in the Market Place*," *The economic journal* 112 (2002), F480–F518.

———, "An experimental investigation of sexual discrimination in hiring in the English labor market," *Advances in Economic Analysis & Policy* 5 (2006).

Rich, J., "What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000," (2014).

Ridgeway, C. L. and S. J. Correll, "Unpacking the gender system a theoretical perspective on gender beliefs and social relations," *Gender & society* 18 (2004), 510–531.

Sacerdote, B., "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?," in E. A. Hanushek, S. J. Machin and L. Woessmann, eds., *Handbook of the Economics of Education* (Elsevier, 2011).

Schwab, K., B. Brende, S. Zahidi, Y. Bekhouche, A. Guinault, A. Soo, R. Hausmann and L. Tyson, *The Global Gender Gap Report*, volume 6 (World Economic Forum, 2014).

Semyonov, M., R. Raijman and A. Gorodzeisky, "The Rise of Anti-Foreigner Sentiment in European Societies, 1988-2000," *American Sociological Review* 71 (2006), 426–449.

StatisticsNorway, "Names 2014," (2014).

Teigen, M., "Kjønnsdeling på langs," in M. Teigen and R. Liza, eds., *Kjønnsdeling og etniske skiller på arbeidsmarkedet* (Gyldendal, 2014).

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

Weichselbaumer, D., "Is it sex or personality? The impact of sex-stereotypes on discrimination in applicant selection," *Univ. of Linz Economics Working Paper* (2000).

Zhou, X., J. Zhang and X. Song, "Gender Discrimination in Hiring: Evidence from 19,130 Resumes in China," (2013).

Zussman, A., "Ethnic discrimination: Lessons from the Israeli online market for used cars," *The Economic Journal* 123 (2013), F433–F468.

Appendix

Question wordings and recoding of survey items

Do your parents have higher education (university/college)?
Categories: 1= Yes, both have higher education, 2=My father has higher education, my mother has not, 3= My mother has higher education, my father has not, 4=No, neither of them have higher education
Recode: We recode into two variables: Father has high education (1/2=1, 3/4 = 0) and Mother has high education (1/3=1, 2/4=0)

Are your parents in work?
Categories: 1= Yes, both, 2=My father is in work, my mother is not, 3=My mother is in work, my father is not, 4=No, neither of them is in work
Recode: We recode into two variables: Father is employed (1/2=1, 3/4 = 0) and Mother is employed (1/3=1, 2/4=0)

Are your parents divorced/separated?
Categories: 1=Yes, 2=No, 3=Don't know
Recode: 3 to missing.

Do you plan to take higher education?
Categories: 1=No, 2=Yes
Recode: We rely on the original coding

The IQ measure is a composite score from three speeded ability tests of arithmetics, word similarities, and figures. These results are provided to us from the military registers. The composite test score is an unweighted mean of the three subtests. The scores are reported in stanine (Standard Nine) units, a method of standardizing raw scores into a nine point standard scale with a normal distribution (mean=5, SD= 2).

Results when including women in the vignette experiment

In the main analysis we restrict the sample to only include men as this is a necessary restriction in analyzing peer effects. It is not necessary to exclude

them in the initial analyzes, however, and we here present the first set of results when all individuals are included and we also test if there is differential discrimination for men and women in the sample. In Table 11 we see the raw difference across cases in how they are evaluated as squad leaders when all individuals are included and the results are very similar to the ones presented in the main text.

In Table 12 we estimate equation 1 for the whole sample and we also test for gender differences in the discrimination by including an indicator variable for female respondents and interacting this variable with female candidate. In column 1 we see that the woman is evaluated as less fit for being a squad leader in the total sample. In column 2 we see that female respondents evaluate the female candidates statistically significantly better and by adding terms we cannot reject that women do not discriminate against neither the male nor the female candidate. In column 3 we show the difference in difference results for the whole sample and note that they are similar and in column 4 we interact all terms with being a female respondent. Doing this, we do not find any evidence of that men and women interpret the information differently.

Table 11: Descriptive statistics across assigned cases

	(1)		(2)		(3)		(4)	
	Ida less info		Martin less info		Ida more info		Martin more info	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Dependent variable								
Score on the candidate (1=very bad, 6=very good)	3.816	(1.018)	4.144	(0.955)	4.405	(0.888)	4.727	(0.823)
Background characteristics								
Mother has high education	0.775	(0.420)	0.628	(0.486)	0.703	(0.459)	0.653	(0.478)
Father has high education	0.887	(0.318)	0.802	(0.401)	0.822	(0.385)	0.802	(0.400)
Mother works	0.850	(0.359)	0.884	(0.322)	0.860	(0.349)	0.891	(0.313)
Father works	0.950	(0.219)	0.965	(0.185)	0.990	(0.100)	0.970	(0.171)
Parents are divorced	0.263	(0.443)	0.256	(0.439)	0.373	(0.486)	0.270	(0.446)
Plan higher education	0.762	(0.428)	0.640	(0.483)	0.794	(0.406)	0.762	(0.428)
IQ	5.782	(1.466)	5.562	(1.314)	5.845	(1.557)	5.606	(1.381)
N (on dependent variable)	87		90		111		110	

Table 12: Gender discrimination including both sexes: Dependent variable is score of the candidate.

	(1)	(2)	(3)	(4)
	No information		With information	
	Baseline	Gender difference	Baseline	Gender difference
Female candidate	-0.285***	-0.325***	-0.230*	-0.275*
	(0.104)	(0.107)	(0.135)	(0.141)
Information added			0.554***	0.550***
			(0.127)	(0.135)
Female candidate*Information			-0.124	-0.112
			(0.157)	(0.166)
Female respondent		-0.107		-0.168
		(0.293)		(0.466)
Female respondent*Female candidate		0.641*		0.962
		(0.346)		(0.639)
Female respondent*Information				0.104
				(0.474)
Female candidate*Info*Female resp.				-0.534
				(0.721)
Mean of dependent variable	4.306	4.306	4.306	4.306
Observations	398	398	398	398
R-squared	0.125	0.134	0.185	0.195
Platoon and Session FE	Yes	Yes	Yes	Yes

Notes: Standard errors clustered at the room level in parantheses.

More on the non-linear effects

Instead of interacting the share of exposure with the female candidate we can also interact the number of women with the female candidate.Table 13 shows the results and we again a non-linear effect whereby it seems to be important to have at least two women in the room. The room with four women drops out from the interaction as these two individuals did not receive the vignette with the female candidate. Again,the results are similar if we control for the number of people in the room.

Table 13: Non-linear effects of exposure on discrimination: Dependent variable is score of the candidate.

	(1)	(2)
	Score	Score
Female candidate	-0.433***	-0.428***
	(0.125)	(0.125)
Female candidate*One woman	-0.199	-0.161
	(0.318)	(0.319)
Female candidate*Two women	0.614***	0.590***
	(0.210)	(0.214)
Female candidate*Three women	0.718	0.716
	(0.535)	(0.532)
One woman	-0.213	-0.163
	(0.317)	(0.291)
Two women	-0.417***	-0.425***
	(0.135)	(0.133)
Three women	0.318	0.310
	(0.216)	(0.214)
Four women	-1.309***	-1.311***
	(0.456)	(0.450)
Number of persons in room		0.128
		(0.095)
Constant	4.681***	3.967***
	(0.366)	(0.645)
Observations	367	367
R-squared	0.163	0.165
Platoon and session FE	Yes	Yes

Notes: Standard errors clustered at the room level in parantheses.
*** p<0.01, ** p<0.05, * p<0.1