Does personal contact with ethnic minorities affect anti-immigrant sentiments?
Evidence from a field experiment / Henning Finseraas, Andreas Kotsadam

Does personal contact with ethnic minorities affect anti-immigrant sentiments? Evidence from a field experiment

Henning Finseraas[a] and Andreas Kotsadam[b]

Abstract

We study the causal effect of personal contact with ethnic minorities on majority members' views on immigration, immigrants' work ethic, and support for lower social assistance benefits to immigrants than to natives. We get exogenous variation in personal contact by randomizing soldiers into different rooms during the basic training period for conscripts in the Norwegian Army's North Brigade. Based on contact theory of majority-minority relations, we spell out why the army can be regarded as an ideal contextual setting for exposure to reduce negative views on minorities. We find a substantive effect of contact on views on immigrants' work ethic, but small and insignificant effects on support for welfare dualism and on views on whether immigration makes Norway a better place to live.

KEYWORDS: Contact theory, field experiment, immigration attitudes, welfare state

[a] Institute for Social Research
P.box 3233 Elisenberg, 0208 Oslo, Norway e-mail: henning.finseraas@samfunnsforskning.no

[b] The Ragnar Frisch Centre for Economic Research
University of Oslo, Gaustadalléen 21, 0349 Oslo, Norway e-mail:andreas.kotsadam@frisch.uio.no.

1. Introduction

The topic of this paper is to what degree majority-minority contact influences anti-immigration sentiments, with a particular emphasis on the relationship between immigration, diversity and the welfare state. In the US, majority-minority conflicts have long been linked to White Americans' welfare state preferences (Gilens 1995). Since Alesina, Glaeser and Sacerdote (2001) and Alesina and Glaeser (2004), research on the relationship between immigration, diversity, and native Europeans' welfare state preferences has thrived. The empirical literature provides no consensus on the effects of ethnic diversity on welfare state preferences (compare e.g. Dahlberg, Edmark, & Lundqvist 2012; Eger 2010; and Brady & Finnigan 2013).

The empirical literature on immigration and welfare state preferences has three important shortcomings that we address in this paper. First, in the existing literature there is a lack of attention to context: To what degree majority-minority tensions are likely to grow or diminish will be highly dependent on the particular context. It is generally accepted that we tend to develop social group identifications, and because language, culture and traditions often differ across ethnic lines, ethnicity will often function as group boundaries for which in-group and out-groups can be constructed. Indeed, the political science literature has been overwhelmingly inspired by the threatened responses to diversity (Enos 2014; 2016; Dancygier 2010). Competition between your in-group and out-groups over scarce resources, social rights and social status can cause out-group prejudice (see e.g. Bobo 1999; Semyonov, Raijman, & Gorodzeisky 2006) which might undermine welfare state support. However, there is obviously no determinism in the saliency of ethnicity as the most important group boundary (Wimmer 2008). Intergroup contact theory (Allport 1954, Pettigrew 1998) specifies the degree of social segregation as key in this respect. According to this perspective, prejudice and negative stereotyping of minorities might decline with contact with out-group members, especially under some conditions: Contact will reduce tensions 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). Friendship potential in the contact has been proposed as a fifth condition, as it increases the probability of affective ties and the willingness to learn about out-group members (Van Laar et al. 2005). Under these four to five conditions, we should expect integration and de-emphasizing of ethnic boundaries, while absent these conditions "every superficial contact we make with an out-group member could (...) strengthen the adverse associations we have" (Allport 1954, 264). Pettigrew and Tropp's (2006:760-761) meta-analysis of empirical studies of contact theory find that contact can reduce prejudice even absent these conditions, however, the effect is larger when the conditions are met, in particular in the most rigorous empirical studies. Thus, diversity can lead to conflicts in contexts of segregation, but to tolerance in contexts of integration (Uslaner 2011, see also van der Waal et al. 2013). Much of existing empirical literature on the consequences of ethnic diversity does not take this contextual distinction into account (e.g. Brady & Finnigan 2013; Senik, Stichnoth, & Van der Straeten 2009; Alesina, Glaeser, & Sacerdote 2001; Ervasti & Hjerm 2012). The discrepancy between the theoretical and empirical models implies that the empirical estimates are not very informative about the importance of minority-majority contact. Null-findings can easily occur if one disregards the contextual situation, as laboratory experiments in cooperative settings often find relations across groups to improve while the opposite holds in competitive settings (Boisjoly et al. 2006). We take the assumptions of contact theory seriously and set up a research design which is informative about the role of social segregation for welfare state support.

Second, the causal relationship between immigration or views on immigrants on the one hand and support for the welfare state on the other hand is rarely addressed empirically. Variables used to assess the impact of immigration and diversity are correlated with too many other variables to make the selection on observables assumption plausible, implying that we need an

explicit research design for causal inference to study the various theoretical accounts for how diversity might influence welfare state support. The design of our study allows a causal interpretation of our results.

The third shortcoming is conceptual. Most of the literature has examined the impact of immigration/diversity on broad or abstract measures of welfare state support, such as support for income redistribution or level of social spending. Presumably, this is because much of the literature has been heavily inspired by the research on majority-minority tensions in the US. We suspect that the impact of increasing ethnic diversity on European welfare states can be different from in the US, mainly because large-scale welfare states were already in place when immigration took off (see Pontusson 2006 for similar claims). To retrench a large-scale welfare state for which most citizens rely on for periods of their life-time might not be comparable to the American experiences of developing a large-scale welfare state in an already ethnically heterogeneous context. The different sequencing of immigration and welfare state development in the US compared to in Europe might contribute to explain why there is a strong link between views on minorities and support for the welfare state in the US (see e.g. Gilens 1995) while the empirical evidence on the impact of immigration on Europeans' welfare state preferences is less clear (Pontusson 2006). In the European context, retrenchment of the welfare state might not be the first best option for xenophobic voters or voters concerned about the fiscal impact of immigration. Instead we suspect that a dual welfare state where one discriminates welfare rights based on for instance citizenship, might be the first-best option for voters who perceive immigration as a drain on public budgets or as a cultural threat (Bay, Finseraas, & Pedersen 2013; Brady & Finnigan 2013; Larsen 2011; Van der Waal et al. 2010). This line of policy has been actively advocated and pursued n some European countries (Careja & Emmenegger 2013), Denmark being one prominent example. By studying support for welfare dualism we study a highly policy relevant outcome which is likely to be affected by views on immigrants.

We conduct an explicit test of contact theory and its relevance for support for welfare dualism. We conduct this test as a field experiment in the military, which provides an institutional context where the specified conditions for contact to improve tolerance are fulfilled. 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. Moreover, the army explicitly promotes views of unity and equality among soldiers of the same rank. Thus, contact theory should operate in this context. Furthermore, the army is a promising venue to study social interaction since the soldiers cannot determine who they want to serve with. To ensure that majority-minority contact and cooperation is real and not superficial, we define contact as room sharing. As friendship is more likely to occur with repetitive contact, and roommate situations have high acquaintance potential (Van Laar et al. 2005), the setting also fulfills the fifth condition for the contact hypothesis.

To make sure that room sharing is exogenous and to reduce biases due to self-selection into social interactions based on own preferences (such as prejudice), we randomize soldiers to different rooms and hence to direct personal contact with minorities. Next we compare outcomes for majority soldiers who were randomized to share room with a minority soldier to majority soldiers who were randomized to a room which consisted of majority soldiers only. Following contact theory, we expect majority member soldiers who are randomly allocated a roommate of ethnic minority background to develop more positive attitudes toward ethnic minorities, and we expect support for welfare dualism to decrease among those with a roommate with minority background. In particular, we propose that contact with minorities might reduce support for dualism by changing the soldiers' views on immigrants' work ethic. A large literature points to perceptions about different groups' work ethic as important for their views on welfare spending/benefits directed to that group (see e.g. Hasenfeld & Rafferty 1989; Gilens 1995; Dyck

5

& Hussey 2008; Rosenthal, Levy, & Moyer 2011). Perceptions about work ethics are likely to be biased, and, if so, close contact and intense daily cooperation with minorities might reduce the bias. Given the close association between perceptions of work ethics and welfare attitudes, it is plausible that support for dualism will be affected as well. This would not be the case, however, if support for dualism is mainly driven by deep-rooted and stable normative considerations or views on majority-minority competition over scarce resources, which are mechanisms not affected by our treatment.

The rest of the paper is organized as follows. In the next section we describe the field experiment, before we describe the construction of the key variables. Next we describe how we deal with the well-known empirical challenges involved in estimating the effect of exposure to others (peer effects). Then we describe the treatment effect equations that we estimate, before we present the empirical results in section 6. To avoid concerns that the data analysis is a "fishing expedition" (Humphreys, de la Sierra, & Van der Windt 2013), we comprehensively described the field experiment, the hypotheses, the construction of the variables, the treatment effect equations, power calculations, and more, in an analysis plan which we submitted to the AEA Registry prior to the data collection. Thus, our hands are tied and we cannot choose the empirical specification which yields the results that we for ideological or publication strategic reasons might prefer. We explicitly mention it when the analysis deviates from the pre-analysis plan.

2 The field experiment

The field experiment was set-up to be conducted on all incoming soldiers of the August 2014-contingent of the North Brigade of the Norwegian Armed Forces (NAF). The soldiers had their first day in the army at the military camp Sessvollmoen, a camp close to Oslo Gardermoen airport. When they meet for their first day in the army the soldiers do not know each other, and

they do not know who will be their roommates. At Sessvollmoen the soldiers go through a program of medical and psychological testing. We got permission to set up a station in this program where we asked the soldiers to complete a survey questionnaire. The data from this survey constitute our baseline data. The field experiment was approved by the Data Protection Official (DPO) of the Norwegian Social Science Data Services (NSD).

After completing the program at Sessvollmoen, the soldiers boarded planes to Northern Norway to start their recruit period. When the soldiers arrived in Northern Norway, they were bussed to a number of different military camps where they were assigned to rooms. The assigned room is where they live for the eight weeks of the recruit period. Roommates perform tasks together, such as cleaning the room for inspection each morning. They also serve in the same platoon, and usually they constitute a team within the platoon. Thus, sharing room during the recruit period constitute intense treatment in the form of forced personal contact.

The first eight weeks of military service is the basic training period, which is known for strict enforcement of military rules and regulation. During these eight weeks the soldiers are to wear their uniform 24/7 and are not allowed to sleep outside the base. The first extended leave is normally granted after completion of the basic training period. Because of the remote location of the bases, this means that the soldiers basically spend all their time with their roommates and fellow conscripts in the platoon. Most of the training in the first eight weeks takes place in platoon formation. After the eight weeks of recruit period the soldiers are sorted into new platoons based on skills and tasks.

We provided the personnel officers in charge of room assignment with an excel sheet which they were instructed to use to randomize soldiers within platoons into rooms. The excel sheet randomized soldiers into rooms when the personnel officer entered the list of soldiers in the

platoon and the size of the rooms. Copies of the excel sheets where emailed to FFI for verification. The procedure allows for a construction of a treatment group consisting of soldiers with an ethnic Norwegian background who were randomized into a room with at least one soldier with an ethnic minority background (see definitions of majority and minority backgrounds below). The control group is soldiers who did not share room with an ethnic minority soldier. We surveyed the soldiers for the second time at the end of the recruit period so that we have pre- and post-treatment data on the outcomes.

The intention, as we spell out in the pre-analysis plan, was for all soldiers in the August 2014-contingent of the North Brigade to be part of the field experiment. However, it turns out that only three battalions, about half of the contingent, followed our instructions and used the excel sheet to randomize soldiers into rooms. It is unclear why many battalions did not follow the procedure, but it appears to be mainly due to lack of communication of the importance of room randomization from battalion commanders down in the hierarchy to personnel officers.[1]

The sample

Norway has military conscription, but the military's demand for soldiers is lower than the size of the age cohorts, which implies that the majority of the soldiers are doing military service voluntarily. According to our survey, 34 percent are unsure of whether they would have served in the military if it was completely voluntary. Since the army has a degree of selection of who they allow to serve in the army, the soldiers are positively selected on background characteristics like grades in high school. Similar to e.g. lab experiments, the positive selection into the army does not invalidate our experiment, but it might have consequences for the

---

[1] The personnel officers who did not follow the randomization procedure apparently decided to follow their usual practice when assigning rooms to the soldiers. The usual practice varies between personnel officers, thus, we cannot use this data in the study. We restrict the analysis to the battalions who followed our procedure, which are Andre Bataljon Nord-Norge (the Second Battalion of Northern Norway), Artilleribataljonen (the Artillery Battalion) and Panserbataljonen (the Armoured Battalion). The power calculations in the pre-analysis plan were based on the assumption that all battalions followed the protocol.

external validity of our results. We return to this issue when interpreting the results.

The total number of soldiers from these battalions participating in the first round is 826, while 577 participated in both rounds of the survey. Most of the attrition comes from soldiers having been dismissed from the Army at the time of the second round. A high dismissal rate is normal during the recruit period. Importantly, we test and confirm that attrition in the panel is unrelated to treatment status as well as to baseline values of the outcome variables (see Appendix Table A2 and the discussion there). The rooms vary in size between 3 and 12 persons, but 73 percent of the sample lives in 6 person rooms. Out of the 577 soldiers, 5 percent (27 soldiers) have a minority ethnic background and 20 percent (116 soldiers) shared room with at least one ethnic minority soldier. These soldiers constitute the treatment group. Ten of the majority soldiers share room with two persons of a minority ethnic background. Since the rooms also vary in size we have variation in the share of minority exposure in the room, ranging from zero to 40 percent.

In the Appendix Table A1 we report descriptive statistics on a set of background characteristics for the full sample and for the treatment and the control group. There are no notable differences between the treatment and the control group (see Table 1 below for formal tests of group differences), with the exception that there are more female soldiers in the control group. This is because there are few female minority soldiers in the Army, and because the Army ideally wants at least two female soldiers in each room (male and female soldiers share room), conditional on there being one woman in the room. In summary, the sample consists of young men where a large majority report that their parents have high education and are in paid employment, and most of the soldiers plan to take more education after the military service.

3. Key variable operationalizations

In this section we describe the operationalization of the outcome variables and ethnic background. In the Appendix we describe the additional background variables used in the analysis.

Ethnic background

The main independent variable is a dummy variable which equals one if there is at least one person with at least one parent born in a non-Western country sharing room with the respondent. Thus, treatment is sharing room with a second generation immigrant with a minority background. This variable is based on the answers on questions regarding parents' country of birth: "In what country is your mother/father born?" 1=Norway, 2=Other Nordic country, 3=Other European country, 4=A country in North America, 5= A country in South America, 6= A country in Africa, 7= A country in Asia, 8= A country in Oceania. We code the person as having a non-Western parent if s/he answers categories 5 to 8.[2] We choose to emphasize non-Western ethnic background rather than foreign background as the effect is likely to be larger for this group. Having a parent from e.g. another Nordic country will not be visible and hence not noticed by the other peers. As an alternative to using a dummy variable of whether there were any minority soldiers in the room, we will also study the effect of the share of minority soldiers in the room. We acknowledge that other definitions of minority background are possible, such as religion or whether both parents are born abroad, however, the sample size precludes us from analyzing variation across definitions.

Outcomes

Our main outcome of interest is support for welfare dualism (*same rights*). The variable is a categorical variable based on the question: "Do you agree or disagree with the statement: Refugees and immigrants should not have the same rights to social assistance as Norwegians."

---

[2] It is not obvious that Oceania should be coded as non-Western, but the decision to do so does not influence the results.

The answer categories were 1= Strongly agree, 2= Agree, 3= Neither agree nor disagree, 4= Disagree, 5= Strongly disagree. This question precisely captures support for separate welfare benefits for social assistance, a benefit for which European Union legislation does not rule out separate benefits based on citizenship or length of stay in the country. Thus, the question is directly policy relevant.

We explore two prejudice-related mechanisms which can explain why contact might decrease support for dualism. First we test whether the respondents think the work ethic of immigrants and natives is more similar if exposed to minorities. View on immigrants' work ethic (*work ethics*) is measured with the following question: "In general, immigrants have poorer work ethics than Norwegians". The answer categories were the same as for same rights. This question is directly linked to the experiences of the soldiers as they work together in the Army, but again involves a generalization from the second generation immigrants to the greater immigrant population. Second, we explore whether there is an effect on attitudes towards immigrants more generally by using the question "Is Norway made a worse or better place to live by people coming to live here from other countries?" (*better country*). The soldiers were asked to answer on a 7-point scale were 1= Worse place to live, 7= Better place to live. In the pre-treatment survey, work ethics and better country are as expected strongly correlated with same rights.[3]

If we compare the distribution of answers on these three outcomes in our sample of soldiers to a sample of men aged 18-30 years from the general population, we find that the soldiers are more positive towards giving immigrants the same rights.[4] About 54 percent in our sample disagree or disagree strongly that immigrants should not have the same rights, compared to about 41 percent in the general population. They are also less likely to agree or strongly agree with the claim that

---

[3] In the pre-treatment survey, the spearman correlation between *same rights* and *work ethics* is .47, between *same rights* and *better country* is .48, and .34 between *work ethics* and *better country*. In the post-treatment survey, the corresponding spearman correlations are .56, .48, and .41.
[4] The data for the general population are described in Bay, Finseraas, & Pedersen (2013).

immigrants have poorer work ethics: Eight percent agree/agree strongly in our sample, versus 22

percent in the general population. For the question on the overall impact of immigration,

however, there is no difference, as about 42 percent in both samples answer on the positive side

of the scale.[5]

While treatment is exposure to a second generation immigrant, the outcomes refers to the rights

of refugees and immigrants. Thus, a treatment effect on this outcome requires that the contact

effect generalizes to a broader out-group than of the treatment. Previous studies have found that

positive effects of contact tend to generalize to distant out-groups (Pettigrew 1998), but it might

be harder to spread from second generation immigrants to the overall immigrant population and

then to policy preferences. We return to this issue in the interpretation of the results.

4. Identification of peer effects

We are interested in the effect of sharing room with at least one ethnic minority soldier on

attitudes. The notion that people are affected by other people is commonly held, yet it is

difficult to establish empirically. The by far most commonly estimated model of peer effects

(Sacerdote 2011) is some version of the following equation:

$$Y_i = a + \beta_1 \overline{Y}_{-i} + \gamma_1 X_i + \gamma_2 \overline{X}_{-i} + \varrho_i \qquad (1)$$

where Y_i is the outcome of interest for individual i which is thought to be a function of the

average outcomes of the peers ($\overline{Y}_{-i}$), the individuals own characteristics (X_i), and the

characteristics of the peers ($\overline{X}_{-i}$). Being interested in welfare dualism, one can imagine a test of

attitudes towards welfare dualism as a function of the peers' attitudes (i.e. room mates'

attitudes) toward dualism and the individuals' own and the peers' background characteristics (including e.g. ethnicity). Without random (or at least plausibly exogenous) allocation of individuals to peers, identification of equation 1 will most likely be subject to severe selection bias due to homophily: Individuals with negative attitudes toward immigrants are more likely to support welfare dualism and less likely to be friends with people of other ethnic groups.

For illustration, we run a set of "naïve" regressions of our outcomes on the share of non-Norwegian friends in high school and the share of immigrants in the soldiers' home municipality.[6] Table A3 in the Appendix shows that having minority friends in high school is positively correlated with all three outcomes. We get similar results for the share of immigrants in the municipality of origin (Panel B). These regressions suggest strong support for the contact hypothesis if interpreted causally. However, the estimation is likely to be severely biased by selection into friend networks and municipalities.

The selection problem is not the only problem facing researchers interested in identifying equation 1. Following Manski (1993) it is common to distinguish between three types of effects in equation 1:

1)   Endogenous effects whereby the individual is affected by the behavior of the other individuals. People try to estimate this effect by looking at β_1.

2)   Exogenous effects whereby individuals are affected by the characteristics of the peers. The hope of the researcher is to identify this by looking at γ_2.

3)   Correlated effects whereby there is a correlation between individuals and their peers because they face similar environments or because of selection.

The selection part of the problem of correlated effects can be solved by randomly allocating

---

[6] The sample is restricted to soldiers with a majority ethnic background.

peers to individuals. In estimating endogenous effects the problem is that if peers affect the outcomes of each other it becomes difficult to separate the effect of the peers on individual i's outcome from the effect of individual i on the peers' outcomes. Manski (1993) labels this the reflection problem. Moreover, even with random assignment of peer groups, separate identification of β_1 and γ_2 is often difficult since peer characteristics affect peer outcomes. Most peer effect papers do not separate between the two, but estimate the combined effect (Sacerdote 2011).

Identification of β_1 is further complicated by the possibility of common variance in outcomes, since individual i and the peers share a common environment (Angrist 2014). For this reason, Angrist (2014) strongly cautions against using outcome-on-outcome estimations and advocates a clear separation between i) the individuals assumed to be affected and ii) the peers assumed to provide the mechanisms for the peer effects. Separation implies that the individuals with the background characteristic which provide the suggested mechanism (those with an ethnic minority background) are excluded from the sample of those assumed to be affected (those with a ethnic majority background).

5. The treatment effect equation

Based on the peer effects discussion, we limit the sample to soldiers without a minority ethnic background, and those with a minority background are used only to define the room characteristics. The following regression models are estimated:

$$Y_{irt2} = \alpha_J + \beta_1 Treated_r + \beta_2 Y_{irt1} + \beta_n X + \varrho_{ir} \tag{2}$$

Where Y_{irt2} is one of the outcomes for individual i in room r at time period t_2. α_J refers to platoon fixed effects and Y_{irt1} is the outcome measured at baseline (i.e. the first survey at day

1). Adding the baseline outcomes is not necessary for identification, but they are included to increase power. Platoon fixed effects are included since randomization occurred at the platoon level, while standard errors will be clustered on rooms as treatment is at the room level. The platoon fixed effects also ensures that the people we are comparing are facing as similar circumstances as possible. Randomization solves the selection issue, but we might still worry that common environmental factors drive the results (see e.g. An 2011). With platoon fixed effects this is less likely. As we compare soldiers within the same platoon, but with different treatment status at the room level, the results have to be interpreted accordingly. In particular, it is possible that there are spillovers such that also being exposed to immigrants in the platoon affects attitudes.

Hence, the effect we measure is the difference between intense contact at the room level net of any effect of contact at the platoon level. To investigate the severity of the spillover effects we estimate the effect of having a second generation immigrant in the platoon but not in the room, and, reassuringly, we find no effects of platoon exposure (see Appendix Table A4). We therefore conclude that the spillovers probably have a very small impact on our results. βn is the vector of coefficients for the covariates and the vector X contains either control variables for which the treatment and the control group differ, all baseline controls, or no controls.[7]

Treatment effect heterogeneity

We further expect there will be a stronger positive effect of minority roommate if the minority roommate has a higher relative ability score. This test is inspired by Carrell, Hoekstra, & West (2015), who finds that only high-aptitude blacks are able to influence the attitudes of whites. We

---

[7] In the pre-analysis plan we also suggest an IV-approach where we use assignment to a room with an ethnic minority soldier as an instrument for actually sharing room with an ethnic minority soldier. We suggested this approach in case the initial allocation was not completely followed. Unfortunately, the Army has only provided information on room assignment, but we have been assured that room switching during the recruit period is very rare. The use of room assignment is in any case most reliable from a causal inference perspective, as the intention to treat estimator relies on less restrictive assumptions than the IV-strategy.

expect roommate ability to matter in so far as negative views on minorities reflect statistical discrimination which will be more strongly updated if one has contact with a high-ability minority person Carrell, Hoekstra, & West (2015).

The soldiers completed three speeded ability tests of arithmetics, word similarities, and figures (see Sundet, Barlaug, & Torjussen 2004), prior to entering military service. We rely on the composite test score, which is an unweighted mean of the three subtests.[8] The ability of ethnic minority roommate is measured as a dummy equal to 1 if the ethnic minority roommate has an IQ score above the median of the minority soldiers in the respective platoon (platoons with only one minority soldiers are excluded).

The treatment heterogeneity across minority IQ will be estimated in the following models:

$$Y_{irt2} = \alpha_J + \beta_1 HighAbilityMin_r + \beta_2 LowAbilityMin_r + \beta_3 Y_{irt1} + \varrho_{ir} \qquad (3)$$

Where *HighAbilityMin* is a dummy representing a high ability-score minority roommate, *LowAbilityMin* is a dummy representing a low ability-score minority roommate. The reference category is having no minority roommate (these three categories are mutually exclusive). β_1 and β_2 test whether high ability and low ability groups differ from the control group. We are also interested in the difference between β_1 and β_2 and will rely on F-tests to examine whether they are statistically significant from each other.

6. Empirical results

Balance

Before presenting the treatment effects, we examine whether the treatment and the control group is balanced across a range of background characteristics (see Appendix for

---

[8] 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).

operationalizations of the variables). Since room allocation is randomized, we should not expect large and significant differences across pre-determined variables. Table 1 reports results from regressions of the treatment indicator dummy on the pre-determined variables.[9] Platoon fixed effects are included in all regressions since room assignment is randomized within platoons. The table also reports an F-test of joint significance.

As to be expected, the differences between the treatment and the control group are small, with one exception. The proportion with an employed mother is nine percentage points smaller in the treatment group (adjusted for platoon fixed effects), a difference which is statistically significant. In light of the generally small differences and the small F-value in the joint test, we nonetheless conclude that the randomization was successful and we will present results when controlling for whether the mother is employed separately.

Table 1: Regressions of treatment status on pre-determined variables.

	Coefficient	t	Standardized Coefficient	N
Same rights t1	-0.132	-1.198	-0.050	552
Work ethics t1	-0.164	-1.498	-0.073	552
Better country t1	0.048	0.328	0.015	552
Mother has high education	-0.019	-0.377	-0.017	550
Father has high education	0.003	0.066	0.003	550
Mother is employed	-0.094**	-2.051	-0.117	549
Father is employed	-0.023	-1.401	-0.063	549
Parents are divorced	0.001	0.009	0.000	549
Plan to take higher education	0.008	0.162	0.007	551
IQ	-0.013	-0.085	-0.004	601
F-test of joint significance	1.07 (p=.28)			

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

Main results

---

[9] We write in the pre-analysis plan that we will analyze imbalance on differences in sibling composition. Unfortunately we have a large proportion missing on the sibling variables, which we suspect is because many without brothers/sisters left the question blank rather than filling in zero. We therefore exclude these questions from the analysis.

The main results are presented in Table 2. In Panel A we present the results without any controls other than the baseline outcome and the platoon fixed effects. In Panel B we add a control for whether the mother is employed since there is a baseline difference between the treatment and the control group on this variable. Finally, in Panel C we add all individual level controls, irrespective of whether there were significant differences between the groups at baseline.

The first column shows the results for *same rights*. These results are very clear: Not only is the treatment coefficient insignificant, but it is also very small. The coefficient decreases further when we add controls. Without controls the estimated difference between the groups is .04 which is small in light of the standard deviation of *same rights* (mean=3.5, SD=1.1). Thus, we conclude that sharing room with a soldier with a minority ethnic background did not change views on whether immigrants should have the same rights to social assistance as natives. These results question a causal interpretation of the impact of contact with minorities on welfare policy preferences which we found in the naive regressions, and which has been identified in purely observational data (e.g. Alesina et al. 2001: 48, Ervasti & Hjerm 2012).

Table 2: Main results

Panel A: No controls	(1) Same Rights t2	(2) Work Ethics t2	(3) Better Country t2
Treated	0.037 (0.085)	0.196** (0.085)	0.083 (0.124)
Same rights t1	0.610*** (0.039)		
Work ethics t1		0.582*** (0.046)	
Better country t1			0.635*** (0.043)
Observations	534	535	534
R-squared	0.383	0.331	0.378
Platoon FE	Yes	Yes	Yes
Panel B: Control for mother employment			
Treated	0.012 (0.084)	0.187** (0.085)	0.080 (0.124)
Same rights t1	0.619*** (0.039)		
Mother is employed	-0.068 (0.111)	-0.007 (0.116)	-0.152 (0.153)
Work ethics t1		0.586*** (0.047)	
Better country t1			0.635*** (0.043)
Observations	531	532	531
R-squared	0.390	0.332	0.379
Platoon FE	Yes	Yes	Yes
Panel C: Full set of individual levels controls			
Treated	0.000 (0.084)	0.187** (0.085)	0.058 (0.126)
Same rights t1	0.605*** (0.040)		
Work ethics t1		0.589*** (0.049)	
Better country t1			0.649*** (0.043)
Observations	522	523	522
R-squared	0.396	0.341	0.411
Platoon FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes

Robust standard errors adjusted for clustering on room in parentheses *** p<0.01, ** p<0.05, * p<0.1

Moving to *work ethic*, we find positive treatment coefficients which are significant at the 5

percent level. The coefficient is stable across panels. In particular the coefficient is not driven by the baseline difference in mothers' employment. The estimated difference between the groups is about .2. Since the standard deviation of the dependent variable is 1, the difference of .2 implies that the substantive size of the effect is non-negligible. Thus, while we find no effect of contact on the policy preference variable, contact improves views on the work ethic of immigrants.

Our interpretation is that by sharing room and cooperating on task solving, treated soldiers have received information on majority-minority differences in work ethics, and updated their priors on these differences. Clearly, since the outcome concerns the work ethic of the overall immigrant population, the effect generalizes from second generation immigrants to the overall immigrant population. However, the contact effect does not spread further to the policy preference. The null result on the policy preference variable in light of the strong effect on work ethics, suggests that view on work ethic is not a major driver of differences in preferences on welfare dualism.

Finally, we find no treatment effect on the general, less-specific question of whether immigrants make the country a better place to live. Moreover, the treatment coefficient is less than .1 which is small in view of *better country*'s standard deviation of 1.4. Again, this result should be compared to the naive regression where we find a strong positive "effect" of having minority friends on the same question.

Treatment heterogeneity

Next, we examine whether the treatment effect depends on measured ability of the minority soldier. We do so by creating one dummy representing whether the soldier shared room with a high ability minority soldier (i.e. a minority soldier with an IQ score above the median of the

minority soldiers in the respective platoon), and one dummy representing a low ability minority

soldier. The reference group is, as before, the control group. We test for treatment

heterogeneity using an F-test of whether the two treatment coefficients are significantly

different from each other.

Table 3: Treatment heterogeneity

	(1) Same Rights t2	(2) Work Ethics t2	(3) Better Country t2
Treated high ability	0.389*	0.126	-0.027
	(0.202)	(0.222)	(0.163)
Treated low ability	0.034	0.178*	0.101
	(0.102)	(0.106)	(0.164)
Same rights t1	0.635***		
	(0.042)		
Work ethics t1		0.603***	
		(0.048)	
Better country t1			0.679***
			(0.051)
F-test of diff high-low	0.47 (p=.49)	0.28 (p=.60)	0.01 (p=.93)
Observations	391	392	391
R-squared	0.412	0.348	0.406
Platoon FE	Yes	Yes	Yes

Robust standard errors adjusted for clustering on room in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The results in Table 3 show that there are indications of treatment heterogeneity on the same

rights-question, as the coefficient for the high-ability treated is larger than the one for low-

ability treated. The difference between coefficients is, however insignificant. For work ethic,

both treatment coefficients have a positive sign, and somewhat surprisingly, the size of the low

ability coefficient is larger than the high ability coefficient. However, the difference is not large

and is statistically insignificant. For better country, the high ability coefficient is negative.

Thus, there is no clear pattern in the results, and all differences are statistically insignificant.

We therefore keep the null hypotheses of no treatment heterogeneity depending on ability.[10]

---

[10] The samples in these regressions are smaller as we have to exclude the platoons that only have one minority soldier. We reach the same conclusion if we define high and low ability minority soldiers based on the total sample

Robustness checks

In the online appendix we present and discuss a large number of robustness checks, all specified in the pre-analysis plan prior to the data collection. We show that conclusions are the same if we i) estimate ordered probit models rather than OLS models, ii) dichotomize the dependent variables, iii) rely on a continuous measure of share of minority soldiers in the room rather than the dummy treatment indicator, and iv) if we control for the share of educated fathers to account for the fact that having a minority soldier in the room implies that the average socio-economic status of the roommates is lower. We further discuss adjustments of p-values for testing multiple outcomes (Rosenblum & van der Laan 2011) and show that the finding for work ethics is significant at the 10 percent level if we adjust the p-values according to the classical Bonferroni method or the false discovery rate method (Benjamini & Hochberg 1995).

Next, in Table 4 we examine treatment effects on two placebo outcomes. These outcomes are both linked to views on gender equality. The first, "Equality not important", is the answer to the item "It is important that men and women share household work equally" (1=Strongly agree, 5=Strongly disagree). We examine the treatment effect on this variable in its original ordinal form and in a recode of those who agree and strongly agree versus the other responses. The second, "Gender not important", is the answer to the item "Which sex do you think is the best in leading a platoon?" (1=Equally good, while those answering Men or Women (almost none) are coded 0. This recoding was determined in the pre-analysis plan). While one may of course imagine circumstances whereby attitudes toward gender equality are affected by sharing room with someone from an ethnic minority, one should expect that the effects on these variables should be smaller. We present results with and without rooms where there were

---

of minority soldiers and not only within platoons.

female soldiers present (some places male and female soldiers share room), since exposure to female soldiers might change views on gender equality, and is correlated with the probability of sharing room with an immigrant as fewer immigrant women serve.

Table 4: Placebo regressions

	(1) Equality not important (ordinal) All rooms	(2) Equality not important (ordinal) Male rooms	(3) Equality not important (dummy) All rooms	(4) Equality not important (dummy) Male rooms	(5) Gender not important (dummy) All rooms	(6) Gender not important (dummy) Male rooms
Treated	-0.075	-0.072	0.008	0.007	-0.044	-0.044
	(0.090)	(0.105)	(0.026)	(0.031)	(0.050)	(0.062)
Baseline	0.578***	0.628***	0.376***	0.438***	0.476***	0.443***
	(0.038)	(0.045)	(0.073)	(0.097)	(0.040)	(0.052)
Obs	535	349	535	349	537	350
R-sq	0.326	0.348	0.163	0.206	0.241	0.240
Platoon FE	Yes	Yes	Yes	Yes	Yes	Yes

Columns 1, 3, and 5 present results using the total sample, while columns 2, 4, and 6 present results when the sample is restricted to male soldiers living in rooms with only men, i.e. excluding mixed rooms. Robust standard errors adjusted for clustering on room in parentheses *** p<0.01, ** p<0.05, * p<0.1

As expected we find small and statistically insignificant treatment effects on "Equality not important". The treatment effect coefficients are larger on "Gender not important", but the t-values are small.

Exploratory analysis

One type of treatment effect heterogeneity which we explore above is to test the effect of exposure to different types of second generation immigrants. Another type is to investigate whether the treatment effect varies across subgroups, in particular whether the effect depends on prior contact with minorities. In the online appendix we present a set of exploratory analyses of this latter question. These analyses were not part of the pre-analysis plan, thus, all findings should be interpreted purely as suggestive for future research. Nonetheless, in these analyses we find that the treatment effect is larger for individuals coming from municipalities with a higher share of immigrants, but not for having had a higher share of immigrant friends

in high school. Neither do we find treatment heterogeneity depending on baseline values.

7. Concluding discussion

In this paper we have examined the effects of direct personal contact with ethnic minorities on majority members' support for welfare dualism, views on immigrants' work ethic, and views on the consequences of immigration. By running a field experiment with randomized personal contact with minorities in a context which allows clear theoretical expectations of reduced prejudice due to personal contact, we overcome important theoretical and methodological shortcomings in the previous empirical literature on this topic.

We find large and statistically significant effects of personal contact on views on immigrants' work ethic. Soldiers with a majority background who have lived and served with a soldier with a minority background are significantly less likely to agree with the statement that immigrants have weaker work ethic than Norwegians. We interpret this result as reflecting the existence of a negative bias in the soldiers' views on minorities' work ethic, which becomes updated and reduced from observing minorities' work ethic through direct personal contact and cooperation. Since treatment is exposure to second generation minorities, while the work ethic question is about the overall immigrant population, the treatment effect appears to generalize beyond the second generation minorities and to the overall immigrant population.

We find small and statistically insignificant treatment effects on support for welfare dualism. Thus, contrary to our expectation, the improved view on immigrants' work ethic is not reflected in reduced support for welfare dualism. The same is true for views on whether immigration makes the country a better place to live. Thus, personal contact changes the outcome which is the one closest to the treatment, but it does not spill over to affecting welfare

24

policy preferences.[11] This finding is surprising in light of the well-established association between views on work ethics and welfare policies (see e.g. Rosenthal et al.'s 2011 meta-analysis). This finding might illustrate the limits of inference from non-experimental data: If you simply estimate the correlation between views on work ethics and welfare policy preferences you are likely to get a biased estimate of the effect of work ethic since it is extremely unlikely that you are able to account for all confounding variables. In our case, we have variation in views on work ethic which is a direct result of the randomization, yet the support for same rights is unmoved. We therefore believe that the role of views on work ethics is overstated in studies using non-experimental data, because the estimates will partly reflect confounding variables that are not observed. For instance, Hasenfeld and Rafferty (1989) understand the role of views about work ethics as an integrated part of the ideology of economic individualism. Obviously, it is hard to observe all parts of this ideology in a specific study, which implies that the parts of the ideology that one do observe will also represent the influence of correlated ideological traits.

Recent developments of contact theory suggest that the effects of contact are moderated by contact quality (Pettigrew 2008), and negative contact is found to be an even stronger predictor than positive contact (Barlow et al. 2012). One reason we do not find effects for all variables might be that some of the contacts have been negative. Our point of departure in this project is that prejudice might exist and create negative biases that potentially will be reduced by direct personal contact. Clearly, if the contact is negative it might reinforce the existing biases. Shook and Fazio (2011) investigate roommate integration for interracial college roommates and find that the effect depends on the relationship quality. Unfortunately, we do not have any questions in our survey on relationship quality, but we find no difference between treated and control

---

[11] Finseraas et al. (2016) investigate the effects of random assignment of women into mixed gender rooms and find that it affects perceptions of women as leaders in a vignette experiment. Similar to in the present study, the outcome is one where information is likely to be updated as a result of the daily interaction.

individuals on self-reported wellbeing in the room (.13, robust SE=.13, p=.33). Furthermore, Shook and Fazio (2008a, 2008b) find that interracial relationships seem to entail less quality than same-race relationships. In particular, the interracial rooms were more likely to dissolve and the roommate satisfaction and involvement was lower. Still, the effects of exposure on attitudes and intergroup anxiety were nonetheless positive and seemingly unrelated to the quality of the relationship. Hence, it is not obvious that relationship quality is a moderator. This feature should be investigated further in future experimental research.

Our results indirectly suggest that other concerns than those regarding work ethics are more important for support for welfare dualism than prejudiced views on ethnic differences in work ethics. These, not mutually exclusive, concerns could be e.g. deep-rooted normative views on reciprocity and deservingness (Van Oorschot 2006), cultural threat (Van der Waal et al. 2010) or ethnic economic competition over public resources (Kitschelt & McGann 1995). Van Oorschot (2006) finds that the ranking of social groups' welfare deservingness is similar across European countries, with immigrants at the bottom, which suggests that deservingness rankings are deep-rooted. Immigrants have contributed to the tax base for a shorter period and might therefore not be perceived as deserving similar welfare benefits as natives (Duffy and Frere-Smith 2014). Changes in perceptions about work ethics will not influence this type of reasoning. The same will be true if preferences for welfare dualism are driven by educational differences in cultural capital (Van der Waal et al. 2010). Regarding economic competition, the explanatory power of personal labour market competition for anti-immigration attitudes is often considered as weak, but sociotropic concerns about the national level impact of immigration, including economic concerns, appear to have more explanatory power (Hainmueller & Hopkins 2014). Concerns about the financial consequences of lower work ethics among immigrants fit within this perspective—implying that we should expect a change in dualism according to this perspective—however, other sociotropic concerns might be more important, for instance concerns that skills mismatch might make labour market integration of

refugees difficult. More generally, the results support the view that at least in the short run, and even with intense treatment, policy preferences can be sluggish and hard to change (see Kuziemko et al., 2015, for recent evidence).

Another possible explanation for the null result is the "atypicalness" of the minority soldiers. Brown and Hewstone (2005) propose that positive changes are more likely to generalize if the out group members can be regarded as typical for their group. The minority soldiers who provide the treatment in our setting are positively selected (better integrated) in comparison to the overall immigrant population for whom the outcomes are about. Thus, treated soldiers might not conceive these soldiers as representative for the overall immigrant population. However, treatment *does* generalize from second generation to the overall immigrant population on the work ethic outcome, so it is not obvious that this is the explanation for the null result for same rights. The null results could also reflect ambiguousness in the same right question, because it refers simultaneously to refugees and immigrants. People might think that labour immigrants have contributed with taxes more than refugees and thus be perceived as more worthy of welfare benefits.

We can make strong claims of high internal validity of our study. Regarding external validity, we study a sample of (mainly) young men which of course implies that results might not generalize to, say, old women. Furthermore, our sample is slightly more positive towards minorities than the Norwegian population of young men. It is possible that treatment effects will be different in populations with different initial distributions of attitudes. Finally, we study people in an unusual context. Although the context of our study is in part a necessity in order to derive clear theoretical expectations, it restricts external validity to contexts with some similarity to ours. Cooperation at workplaces, in classrooms, and in team sports has similarities to our context. That said, the structure of contact in these contexts are weaker and less

streamlined, which might imply that treatment effects from direct contact might be weaker than what we find here. We strongly urge future research to conduct field experiments in other contexts so that more general knowledge can be reached.

References

Alesina, A. & Glaeser, E. (2004). *Fighting Poverty in the US and Europe: A World of Difference*. Oxford: Oxford University Press.

Alesina, A., Glaeser, E. & Sacerdote, B. (2001). Why doesn't the US have a European-style welfare system? *Brookings Papers on Economic Activity* 2001(2).

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

An, W. (2011). Models and Methods to Identify Peer Effects. In J. Scott & P.J.Carrington, eds., *The Sage Handbook of Social Network Analysis.* London: Sage.

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

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R., Harwood, J., & Sibley, C. G. (2012). The contact caveat negative contact predicts increased prejudice more than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin*, *38*(12), 1629-1643.

Bay, A-H., Finseraas, H. & Pedersen, A. W. (2013). Welfare dualism in two Scandinavian welfare states: Public opinion and party politics. *West European Politics* 36(1): 199–220.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289–300.

Bobo, L. D. (1999). Prejudice as Group Position: Microfoundations of a Sociological Approach to Racism and Race Relations. *Journal of Social Issues* 55(3): 445–472.

Boisjoly, J., et al. (2006). Empathy or Antipathy? The Impact of Diversity. *American Economic Review* 96(5): 1890–1905.

Brady, D., & Finnigan, R. (2013). Does Immigration Undermine Public Support for Social Policy? *American Sociological Review* 79(1): 17–42.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. *Advances in experimental social psychology*, *37*, 255-343.

Careja, R., & Emmenegger, P. (2013). Keeping Them Out: Migration and Social Policies in the 'Reluctant Countries of Immigration'. In A. M. Suszycki, and I. P. Karolewski, eds. *Citizenship and Identity in the Welfare State*. Baden-Baden: Nomos Verlag.

Carrell, S. E., Hoekstra, M., & West, J.E. (2015). The Impact of Intergroup Contact on Racial Attitudes and Revealed Preferences. NBER Working Paper No. 20940.

Dahlberg, M., Edmark, K. & Lundqvist, H. (2012). Ethnic Diversity and Preferences for Redistribution. *Journal of Political Economy* 120(1): 41–76.

Dancygier, R. M. (2010). *Immigration and Conflict in Europe*. Cambridge, MA: Cambridge University Press.

Dixon, J., Durrheim, K. & Tredoux, C. (2005). Beyond the Optimal Contact Strategy: A Reality Check for the Contact Hypothesis. *American Psychologist* 60(7): 697-711.

Duffy, B. & Frere-Smith, T. (2014), *Perceptions and Reality: Public Attitudes to Immigration*, London: Ipsos MORI.

Dyck, J. J. & Hussey, L.S. (2008). The End of Welfare as We Know It? Durable Attitudes in a Changing Information Environment. *Public Opinion Quarterly* 72(4): 589–618.

Eger, M. A. (2010). Even in Sweden: The Effect of Immigration on Support for Welfare State Spending. *European Sociological Review* 26(2): 203–217.

Enos, R. D. (2014). Causal effect of intergroup contact on exclusionary attitudes. *Proceedings of the National Academy of Sciences* 111(10): 3699–3704.

Enos, R. D. (2016). What the Demolition of Public Housing Teaches us About the Impact of Racial Threat on Political Behavior. *American Journal of Political Science* 60(1):123–142.

Ervasti, H., & Hjerm, M. (2012). Immigration, trust and support for the welfare state. In H. Ervasti et al., eds, *The Future of the Welfare State: Social Policy Attitudes and Social Capital in Europe*. Camberley, UK: Edward Elgar Publishing.

Finseraas, H., et al. (2016). Exposure to female colleagues breaks the glass ceiling—Evidence from a combined vignette and field experiment. *European Economic Review*: 90: 363-374.

Gilens, M. (1995). Racial Attitudes and Opposition to Welfare. *Journal of Politics* 57(4):994-1014.

Hainmueller, J., & Hopkins, D. J. (2014). Public Attitudes Toward Immigration. *Annual Review of Political Science*, 17, 225-249.

Hasenfeld, Y., & Rafferty, J. (1989). The Determinants of Public Attitudes Toward the Welfare State. *Social Forces* 67(4): 1027–1048.

Humphreys, M., Sanchez de la Sierra, R. & Van der Windt, P. (2013). Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration. *Political Analysis* 21(1): 1–20.

Kitschelt, H. & McGann, A.J. (1995). *The Radical Right in Western Europe: A Comparative Analysis*. Ann Arbor: University of Michigan Press.

Kuziemko, I et al. (2015). How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments. *American Economic Review* 105(4): 1478–1508.

Larsen, C. A. (2011). Ethnic heterogeneity and public support for welfare: Is the American experience replicated in Britain, Sweden and Denmark? *Scandinavian Political Studies* 34(4): 332–353.

Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* 60(3): 531–542.

McLaren, L. M. (2003). Anti-immigrant Prejudice in Europe: Contact, Threat Perception, and Preferences for the Exclusion of Migrants. *Social Forces* 81(3): 909–936.

Pettigrew, T. F. (1998). Intergroup Contact Theory. *Annual Review of Psychology* 49(1): 65–85.

Pettigrew, T. F. & Tropp, L.R. (2006). A Meta-Analytic Test of Intergroup Contact Theory. Journal of Personality and Social Psychology 90(5): 751-783.

Pettigrew, T. F. & Tropp, L.R.. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology*, *38*(6), 922-934.

Pettigrew, T. F. (2008). Future directions for intergroup contact theory and research. *International Journal of Intercultural Relations*, 32, 187-199.

Pontusson, J. (2006). The American Welfare State in Comparative Perspective: Reflections on Alberto Alesina and Edward L. Glaeser, "Fighting Poverty in the US and Europe. *Perspectives on Politics* 4(2): 315–326.

Rosenblum, M. & van der Laan, M.J. (2011). Optimizing Randomized Trial Designs to Distinguish Which Subpopulations Benefit from Treatment. *Biometrika* 98(4): 845–860.

Rosenthal, L., Levy, S. & Moyer, A. (2011). Protestant work ethic's relation to intergroup and policy attitudes: A meta-analytic review. *European Journal of Social Psychology* 41(7): 874–885.

Sacerdote, B. (2011). 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*. Amsterdam: North Holland.

Shook, N. J., & Fazio, R. H. (2008a). Interracial roommate relationships an experimental field test of the contact hypothesis. *Psychological Science*, *19*(7), 717-723.

Shook, N. J., & Fazio, R. H. (2008b). Roommate relationships: A comparison of interracial and same-race living situations. *Group Processes & Intergroup Relations*, *11*(4), 425-437.

Shook, N. J., & Fazio, R. H. (2011). Social network integration: A comparison of same-race and interracial roommate relationships. *Group Processes & Intergroup Relations*, *14*(3), 399-406.

Semyonov, M., Raijman, R. & Gorodzeisky, A.(2006). The Rise of Anti- Foreigner Sentiment in European Societies, 1988-2000. *American Sociological Review* 71(3): 426–449.

Senik, C., Stichnoth, H.& Van der Straeten, K. (2009). Immigration and Natives' Attitudes towards the Welfare State: Evidence from the European Social Survey. *Social Indicators Research* 91(3): 345–370.

Sundet, J. M, Barlaug, D. G.& Torjussen, T. M. (2004). The end of the Flynn effect?: A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence* 32(4): 349–362.

Uslaner, E, M. 2011. Trust, Diversity, and Segregation in the United States and the United Kingdom. *Comparative Sociology* 10(2): 221–247.

Van der Waal, J et al. (2010). 'Some are more equal than others': Economic egalitarianism and welfare chauvinism in the Netherlands. *Journal of European Social Policy* 20(4): 350– 363.

Van der Waal, J., de Koster, W. & Achterberg, P. (2013). Ethnic Segregation and Radical right-Wing Voting in Dutch Cities. *Urban Affairs Review* 49 (5), 748-777

Van Laar, C., et al. (2005). The Effect of University Roommate Contact on Ethnic Attitudes and Behavior. *Journal of Experimental Social Psychology* 41(4): 329–345.

Van Oorschot, W. (2006). Making the difference in social Europe: Deservingness perceptions among citizens of European welfare states. *Journal of European Social Policy* 16(1): 23–42.

Wimmer, A. (2008). The Making and Unmaking of Ethnic Boundaries: A Multilevel Process Theory. *American Journal of Sociology* 113(4): 970–1022.

Online Appendix

Question wordings and recoding of survey items for tests of balance

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

During your last school year, what share of your friends had a non-Norwegian ethnic background?

Categories: 1=Less than 20 percent, 2=20-40 percent, 3=40-60 percent, 4=More than 60 percent

Recode: We rely on the original coding.

Descriptive statistics

Table A1: Descriptive statistics

	All	Treatment group	Control group
	Mean	Mean	Mean
	(SD)	(SD)	(SD)
Female	0.12	0.06	0.14
	(0.33)	(0.24)	(0.35)
Mother has high education	0.67	0.67	0.67
	(0.47)	(0.47)	(0.47)
Father has high education	0.80	0.82	0.80
	(0.40)	(0.39)	(0.40)
Mother is in paid work	0.88	0.82	0.90
	(0.33)	(0.39)	(0.31)
Father is in in paid work	0.97	0.96	0.98
	(0.15)	(0.21)	(0.13)
Parents are divorced	0.30	0.32	0.30
	(0.46)	(0.47)	(0.46)
Plan to take highed education	0.74	0.74	0.74
	(0.44)	(0.44)	(0.44)
Age	19.35	19.26	19.37
	(0.81)	(0.61)	(0.86)
Observations	546	114	432

Attrition

We have two sources of attrition. One source is due to people leaving the population because they are discharged from the military. We use these observations to calculate room characteristics, but otherwise they are discarded. The second source is due to missing data in the survey.

As described in the pre-analysis plan, the first test is to see whether attrition is related to treatment status. To check this we estimate the following regression:

$$Attrition_i = \alpha_J + \beta_1 Treatment + \beta_n X + \varrho \qquad (A1)$$

We see that attrition is unrelated to treatment status in columns 1 (without control variables) and 2 (with individual level control variables) of Table A1. In addition we run the following regression:

$$Attrition_i = \alpha_J + \beta_1 Y_{t1} + \beta_n X + \varrho, \qquad (A2)$$

Here we test whether our outcomes at baseline are related to the probability of not being in the sample in the second period. The results in columns 3 to 5 show no statistically significant relationship between attrition and our outcomes of interest.

Table A2:Tests of non-random attrition

VARIABLES	(1) Treatment	(2) Treatment	(3) Outcome	(4) Outcome	(5) Outcome
Treated	0.016	0.021			
	(0.051)	(0.051)			
Better country t1			-0.014		
			(0.011)		
Work ethics t1				-0.007	
				(0.015)	
Same rights t1					-0.011
					(0.013)
Observations	783	765	826	826	826
Platoon FE	Yes	Yes	Yes	Yes	Yes
Individual controls	No	Yes	No	No	No

Robust standard errors adjusted for clustering on room in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Naïve estimation of peer effects

Table A3: Naive estimation of peer effects

	(1)	(2)	(3)
Panel A: Minority friends in highschool	Same Rights t2	Work Ethics t2	Better Country t2
Minority friends	0.138*	0.156**	0.230**
	(0.074)	(0.063)	(0.109)
Observations	532	533	532
R-squared	0.049	0.058	0.035
Platoon FE	Yes	Yes	Yes
Panel B: Share of immigrants in the municipality			
Share of immigrants	1.754***	0.810*	1.365***
	(0.481)	(0.429)	(0.501)
Observations	528	529	528
R-squared	0.061	0.056	0.034
Platoon FE	Yes	Yes	Yes

Robust standard errors adjusted for clustering on room in Panel A and municipality in Panel B.
*** p<0.01, ** p<0.05, * p<0.1

Platoon level regression

Table A4: Regressions examining the effect of exposure at the platoon level

	(1) Same rights t2	(2) Work ethics t2	(3) Better country t2
Treated platoon	0.045 (0.140)	0.029 (0.143)	0.119 (0.204)
Same rights t1	0.619*** (0.045)		
Work ethics t1		0.629*** (0.044)	
Better country t1			0.615*** (0.048)
Platoon FE	No	No	No

The sample consists only of individuals not exposed to immigrants at the room level. Robust standard errors adjusted for clustering on room. *** p<0.01, ** p<0.05, * p<0.1

Robustness checks

In Table 2 we estimate linear regression models, despite that the dependent variables are categorical variables. In Table A5, Panel A, we show that conclusions are the same if we instead estimate ordered probit models. In Panel B, we present results when dichotomizing the dependent variables (see table note). Doing so produces a less precise estimate for work ethic (t=1.81). In addition, the treatment coefficient on better country is of the same size as for work ethic and statistically significant at the 5 percent level (t=2.03). We should point out, however, that we did not suggest dichotomizing better country in the pre-analysis plan, but presents it here for completeness. Results are similar if we add individual level controls (see Table A6).

The quite large treatment effect on the dichotomized version of better country, but the small treatment effect on the continuous original measure might suggest that treatment led to increased polarization on this issue. We did not suggest a potential polarization effect in the pre-analysis plan, so the suggestion of a polarization effect should be considered as purely speculative. Nonetheless, to examine this further we create a worse country dummy, which is equal to 1 if they answer 1-3 on the scale. The treatment coefficient is positive also on this outcome (not shown), which suggest a degree of polarization, but the coefficient is very small (.01) and not statistically significant.

In Panel C, we replace the treatment indicator dummy with a continuous measure of share of minority soldiers in the room. The share variable takes into account that room size varies, and, perhaps more importantly, the possibility of sharing room with more than one minority soldier. Our conclusions are the same as when we rely on the treatment indicator dummy, but the relationship between share and work ethic is less precisely estimated compared to when we use the dummy approach (t=1.86).

In interpreting the effects as running via the ethnicity of the peers we might be worried that we pick up something correlated with ethnicity of peers, in particular the education level of parents. Since the education level of parents of second generation immigrants is lower than for natives, the peer effect might in part be due to effects of sharing room with soldiers with low educated parents. In Table A7, Panel A we report results when controlling for the share of high educated fathers[1] which shows that the treatment coefficients do not change much when including this control. In Panel B we present results without the treatment indicator, effectively making the share of high educated fathers the treatment.[2] The correlations between the share of high educated fathers and the outcomes are small, and for work ethics the size of the coefficient is only one third of the correlation between share of minority soldiers and work ethics. Thus, we conclude that the main results are not much biased by differences in the share of high educated parents.

---

[1] Conclusions are the same if we rely on the share of high educated mothers, or if we include them both.

[2] We exclude the individual himself from the calculation of the shares so that we follow the "leave-out-oneself"-approach (see section 3). This regression suffers from the bias of not separating the effects for the treated from those providing the treatment. Excluding those having a father with high education from the estimation solves this problem. The conclusions are the same, but the sample size is small.

Table A5: Robustness checks

Panel A: Ordered probit regressions	(1) Same Rights t2	(2) Work Ethics t2	(3) Better Country t2
Treated	0.063	0.274**	0.093
	(0.105)	(0.119)	(0.117)
Same rights t1	0.740***		
	(0.064)		
Work ethics t1		0.786***	
		(0.084)	
Better country t1			0.610***
			(0.054)
Observations	534	535	534
Platoon FE	Yes	Yes	Yes
Panel B: Linear probability models			
Treated	-0.012	0.077*	0.093**
	(0.041)	(0.042)	(0.046)
Same rights t1	0.477***		
	(0.042)		
Work ethics t1		0.464***	
		(0.040)	
Better country t1			0.500***
			(0.040)
Observations	534	535	534
Platoon FE	Yes	Yes	Yes
Panel C: Share of minority soldiers			
Treated	0.213	0.713*	0.205
	(0.391)	(0.384)	(0.474)
Same rights t1	0.611***		
	(0.039)		
Work ethics t1		0.582***	
		(0.046)	
Better country t1			0.635***
			(0.043)
Observations	534	535	534
Platoon FE	Yes	Yes	Yes

Robust standard errors adjusted for clustering on room in parentheses. In Panel B *same rights* and *work ethics* are recoded to binary indicators of support for by collapsing the categories "disagree" and "disagree strongly", while *better country* is dichotomized by recoding categories 5-7 to 1 and the others to 0. *** p<0.01, ** p<0.05, * p<0.1

Table A6: Regressions (LPM) on dichotomized DVs with controls

Panel A: Control for mother employment	(1) Same Rights t2	(2) Work Ethics t2	(3) Better Country t2
Treated	-0.024	0.062	0.098**
	(0.041)	(0.042)	(0.047)
Same rights t1	0.481***		
	(0.042)		
Mother is employed	-0.027	-0.049	-0.021
	(0.053)	(0.067)	(0.056)
Work ethics t1		0.469***	
		(0.040)	
Better country t1			0.499***
			(0.040)
Observations	530	531	530
R-squared	0.266	0.263	0.284
Platoon FE	Yes	Yes	Yes
Panel B: Full set of individual levels controls			
Treated	-0.029	0.064	0.096**
	(0.040)	(0.042)	(0.048)
Same rights t1	0.461***		
	(0.045)		
Work ethics t1		0.475***	
		(0.040)	
Better country t1			0.486***
			(0.042)
Observations	521	522	521
R-squared	0.282	0.282	0.303
Platoon FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes

Robust standard errors adjusted for clustering on room in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A7: Regressions with share for high educated fathers

	(1) Same Rights t2	(2) Work Ethics t2	(3) Better Country t2
Panel A: Control for share of high educated fathers			
Treated	0.032	0.200**	0.081
	(0.085)	(0.084)	(0.124)
Same rights t1	0.615***		
	(0.038)		
Share of high educated fathers	-0.025	0.277	-0.208
	(0.176)	(0.172)	(0.266)
Work ethics t1		0.588***	
		(0.047)	
Better country t1			0.633***
			(0.043)
Observations	532	533	532
R-squared	0.389	0.334	0.379
Platoon FE	Yes	Yes	Yes
Panel B: Share of high educated fathers w/o treated			
Treated	-0.028	0.257	-0.215
	(0.176)	(0.174)	(0.263)
Same rights t1	0.615***		
	(0.038)		
Work ethics t1		0.581***	
		(0.047)	
Better country t1			0.633***
			(0.043)
Observations	532	533	532
R-squared	0.389	0.328	0.378
Platoon FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes

Robust standard errors adjusted for clustering on room in parentheses *** p<0.01, ** p<0.05, * p<0.1

Multiple comparisons

One of the most critical features of the pre analysis plan is to specify as exactly as possible the outcomes to be tested, to avoid that the researcher selects the most publishable results ex post. Doing so would be of little use, however, without an understanding of the limits to power of testing multiple hypotheses. If we specified to test, say 50 hypotheses we would end up with several being statistically significant by chance alone. It is easy to generate many hypotheses about peer effects, in particular regarding treatment heterogeneity. An open ended investigation should however be considered as a hypotheses generating process rather than as a test of hypotheses. To take into account the limits to power from testing multiple hypotheses we need to be restrictive with regard to the number of hypotheses tested and to impose pre-specified decision rules on thresholds for statistical significance (Rosenblum and van der Laan 2011).

The only significant treatment effects we have are on the work ethics-outcome. To account for having four different outcomes (three outcomes and one heterogeneity test) at the outset we adjust our critical levels for rejection of the null hypothesis for work ethics to be: $.05/4 = .0125$ for the .05-level, and $.10/4 = .025$ for the .10-level. This correction is the same whether we use the classical Bonferroni method or the false discovery rate method (Benjamini and Hochberg 1995), as we only have statistically significant results for one outcome. In the main results (Table 2), the p-value is .022 in Panel A and .03 in Panel B and C. Thus, using the adjusted critical levels, the treatment effect is borderline significant at the ten percent level. Given the substantive size of the effects on this outcome, we attribute the impreciseness to a somewhat weak level of statistical power.

Exploratory analysis

Table A8: Exploratory analysis of heterogeneous effects on immigrants *work ethics* based on previous exposure

	(1) Municipality share	(2) Friends	(3) Municipality share	(4) Friends	(5) Baseline attitudes	(6) All
Treated	-0.141	0.237	0.045	0.187	0.780	0.710
	(0.150)	(0.298)	(0.134)	(0.298)	(0.516)	(0.498)
High im. share	-0.010		0.049			0.043
	(0.094)		(0.073)			(0.073)
T*High im. share	0.410**		0.283			0.298*
	(0.196)		(0.177)			(0.178)
Immigrant friend		0.181***		0.058		0.046
		(0.062)		(0.059)		(0.061)
Treated*Im. friend		-0.128		0.001		-0.068
		(0.198)		(0.188)		(0.194)
Baseline attitude			0.583***	0.579***	0.624***	0.623***
			(0.046)	(0.047)	(0.045)	(0.046)
Treated*Baseline					-0.160	-0.161
					(0.140)	(0.140)
Observations	589	534	535	534	535	534
R-squared	0.057	0.060	0.337	0.332	0.335	0.342
Platoon FE	Yes	Yes	Yes	Yes	Yes	Yes
Ind. controls	No	No	No	No	No	No

Robust standard errors adjusted for clustering on room in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A9: Exploratory analysis of heterogeneous effects on immigrants *better country* based on previous exposure

	(1) Municipality share	(2) Friends	(3) Municipality share	(4) Friends	(5) Baseline attitudes	(6) All
Treated	-0.011	-0.073	0.079	-0.048	-0.329	-0.398
	(0.185)	(0.375)	(0.147)	(0.340)	(0.482)	(0.543)
High im. share	0.143		0.093			0.086
	(0.131)		(0.108)			(0.110)
T*High im. share	0.279		0.001			-0.086
	(0.294)		(0.225)			(0.256)
Immigrant friend		0.198*		0.119		0.111
		(0.113)		(0.113)		(0.113)
Treated*Im.friend		0.123		0.088		0.083
		(0.280)		(0.258)		(0.277)
Baseline attitude			0.631***	0.632***	0.615***	0.610***
			(0.044)	(0.043)	(0.050)	(0.050)
Treated*Baseline					0.096	0.093
					(0.102)	(0.102)
Observations	588	533	534	533	534	533
R-squared	0.025	0.036	0.379	0.381	0.379	0.383
Platoon FE	Yes	Yes	Yes	Yes	Yes	Yes
Ind. controls	No	No	No	No	No	No

Robust standard errors adjusted for clustering on room in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A10: Exploratory analysis of heterogeneous effects on immigrants *same rights* based on previous exposure

	(1) Municipality share	(2) Friends	(3) Municipality share	(4) Friends	(5) Baseline attitudes	(6) All
Treated	-0.337** (0.138)	0.101 (0.290)	-0.030 (0.124)	0.167 (0.249)	0.098 (0.340)	0.296 (0.410)
High im. share	0.130 (0.106)		0.031 (0.086)			0.014 (0.084)
T*High im. share	0.533** (0.228)		0.122 (0.190)			0.214 (0.187)
Immigrant friend		0.172** (0.083)		0.021 (0.078)		0.019 (0.078)
Treated*Im.friend		-0.139 (0.196)		-0.101 (0.169)		-0.146 (0.174)
Baseline attitude			0.604*** (0.040)	0.615*** (0.038)	0.614*** (0.044)	0.619*** (0.044)
Treated*Baseline					-0.018 (0.096)	-0.053 (0.093)
Observations	588	533	534	533	534	533
R-squared	0.056	0.050	0.384	0.390	0.383	0.391
Platoon FE	Yes	Yes	Yes	Yes	Yes	Yes
Ind. controls	No	No	No	No	No	No

Robust standard errors adjusted for clustering on room in parentheses. *** p<0.01, ** p<0.05, * p<0.1