Does Personal Contact with Ethnic Minorities Affect Support for Welfare Dualism? Evidence From a Field Experiment

Henning Finseraas and Andreas Kotsadam*

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.

*Henning Finseraas: Institute for Social Research, P.box 3233 Elisenberg, 0208 Oslo, Phone: +47 95169855, Norway e-mail: henning.finseraas@samfunnsforskning.no, Andreas Kotsadam: Department of Economics, University of Oslo, P.box 1095 Blindern, 0317 Oslo, Norway, Phone: +47 40338176, e-mail: andreas.kotsadam@econ.uio.no. We are grateful to Jon Fiva, Axel West Pedersen, Terje Wessel and seminar participants at University of Oslo, Oslo University College, Norwegian Social Research, Institute for Social Research, and the 2015 annual meeting of Norwegian Political Scientists for helpful comments and suggestions. We thank Ada Fuglset, Eirik Strømland, and Wiktoria Szczesna for excellent research assistance. We would also like to thank Frank Steder and Torbjørn Hanson at the Norwegian Defence Research Establishment, and the soldiers and staff at the North Brigade. Funding from the Norwegian Research Council (grant number 236801) is acknowledged.

1 Introduction

The impacts of immigration on social, political and economic outcomes in receiving countries are hot topics in the social sciences as well as in the public debate. Although immigration might improve productivity (Peri 2012), promote innovation (Hunt and Gauthier-Loiselle 2010), and reduce global poverty (Clemens 2011), concerns are raised about the potential for majority-minority tensions and conflicts due to distributional effects of immigration (Dustmann, Frattini, and Preston 2013) or as a consequence of ethnic segregation (Uslaner 2011). Majority-minority conflicts can influence policy preferences and outcomes, and the welfare state has been singled out as particularly vulnerable in this respect. Such conflicts can influence the size of the welfare state spending through various channels. Political scientists have typically emphasized that ethnic tensions might make it more difficult to organize interest groups (Stephens 1979), in particular labour unions, which some consider as essential in order to develop a generous welfare state (Korpi and Palme 2003). Moreover, such conflicts might shift political competition from being organized along income and class divisions to being organized along ethnic lines (Przeworski and Sprague 1986). If so, coalitions in favour of generous redistribution and universal social insurance programs might be undermined (Austen-Smith and Wallerstein 2006).[1]

The topic of this paper is to what degree majority-minority contact directly influences voters' public policy preferences, more specifically over the design of the welfare state. In the US, majority-minority conflicts have long been linked to White Americans' welfare state preferences (Gilens 1995), and, starting with Alesina et al. (2001) and Alesina and Glaeser (2004), research on the relationship between immigration, diversity, and native Europeans' welfare state preferences have thrived. The literature on the effects of ethnic diversity on the electorate point, however, in all directions (Dahlberg, Edmark, and Lundqvist 2012; Eger 2010; Brady and Finnigan 2013; Finseraas 2008).

The empirical literature on immigration and welfare state preferences have three important shortcomings that we address in this paper. First, the causal relationship between

[1] In addition, if there are cultural differences in spending preferences, then diversity might shift the composition of public spending as immigrants are enfranchised (Luttmer and Singhal 2011; Vernby 2013).

immigration or views on immigrants on the one hand and support for the welfare state on the other hand is rarely addressed empirically.[2] 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 establish whether diversity is a threat to e.g. welfare state support. While most empirical studies suggest that intergroup contact reduces intergroup prejudice (see Pettigrew and Tropp 2006), the worry that most of these results are driven by selection, reverse causality, or both, looms large in this literature, since people self-select into networks based on pre-existing attitudes, or since people in different networks face different circumstances. Only a handful of studies have randomly assigned peers of different ethnicities to study the causal effects of minority-majority contacts. Boisjoly et al. (2006) find that being assigned to an African-American roommate in college makes White students more positive toward affirmative action. Carrell, Hoekstra, and West (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 more so if the Black peers were of high aptitude. Finally, Van Laar et al. (2005) find improved intergroup attitudes by randomized exposure which generalize to other out-groups, such as having a Black roommate affecting attitudes toward Latinos, and vice versa. We are not aware of similar studies outside the US context, and none of these studies investigate the impact of randomized exposure on welfare state preferences.

The second 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. The main reason, we believe, is that 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 the European welfare state can be quite different from the US experiences, mainly

[2] Dahlberg, Edmark, and Lundqvist (2012) is one important exception. They exploit arguably exogenous variation in municipality immigrant shares by refugee placement in Sweden and identify a negative effect on preferences for redistribution. Neighborhood effects are, however, unlikely to be generalizable to effects of interpersonal contact since physical proximity does not necessarily imply personal contact. McLaren (2003) finds that living in areas with high immigration without contact is correlated with a higher, not lower, threat perception.

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 is probably not 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, and Pedersen 2013; Brady and Finnigan 2013; Larsen 2011).[3] In some European countries this line of policy has been actively advocated and pursued in the last decade (Careja and Emmenegger 2013), Denmark being one prominent example.

The third shortcoming is theoretical, and concerns a lack of attention to in what contexts majority-minority tensions are likely to grow or diminish. It is 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. 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, and Gorodzeisky 2006) and thus support for welfare dualist policies. 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, but only under some quite

[3] What we label welfare dualism is sometimes labeled welfare chauvinism. We prefer dualism since it is a more descriptive and less value-laden term.

restrictive conditions: Contact will reduce tensions 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). 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). The take-home point is that diversity can lead to conflicts in contexts of segregation, but to tolerance in contexts of integration (Uslaner 2011). The existing empirical literature on the consequences of ethnic diversity tends to overlook the importance of segregation (e.g. Brady and Finnigan 2013; Senik, Stichnoth, and Van der Straeten 2009; Alesina, Glaeser, and Sacerdote 2001; Ervasti and Hjerm 2012).[4] 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 and allows a causal interpretation of our results.

Specifically, we conduct an explicit test of contact theory in the military, which provides an institutional context where the specified conditions for contact to cause tolerance is fulfilled.[5] 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 promote 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

[4]See Dixon, Durrheim, and Tredoux (2005) for a similar critique of the empirical literature on contact theory in social psychology.

[5]In fact, the initial inspiration and early empirical support for contact theory comes from a study of integration of Black soldiers into the US Army (see Pettigrew 1998).

study social interaction also 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 extended and repetitive contact (Wessel 2009) 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. 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.

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" (Miguel et al. 2014; Humphreys, de la Sierra, and Van der Windt 2013), we comprehensively describe 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.[6] 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 state clearly when the analysis deviate from the pre-analysis plan.

[6]An anonymous version of the plan is attached to the submission as an online appendix for the referees. The link to the online document will be provided here upon acceptance.

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).[7] These 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.

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 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.

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. This allows for a construction of a treatment group consisting of soldiers with an ethnic

[7] Norway has military conscription, i.e. military service is mandatory, however, 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.

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, and examine whether outcomes differ between treatment and control group, controlling for outcomes at baseline and platoon fixed effects (see below).

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 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.[8]

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. 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 A1 and the discussion there). The rooms vary in size between 3 and 12 persons, but 73 percent of the sample live in 6 person rooms. Out of the 577 soldiers, 5 percent (27 soldiers) have a non-Western 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 non-Western ethnic background. Since the rooms also vary in size we also have variation in the share of immigrant exposure in the room, ranging from zero to 40 percent.

[8] 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 randomization from battalion commanders down in the hierarchy to personnel officers. The personnel officers apparently decided to do what they have always done when assigning rooms, which appears to vary between personnel officers. Thus, we cannot use this data in the study and are forced to restrict the analysis to the battalions who followed procedure. These battalions are Andre Bataljon Nord-Norge (the Second Battalion of Northern Norway), Artilleribataljonen (the Artillery Battalion) and Panserbataljonen (the Armoured Battalion).

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.[9] 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 rely on the share of minority soldiers in the room.

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." The answer categories were 1= Strongly agree, 2= Agree, 3= Neither agree nor disagree, 4= Disagree, 5= Strongly disagree. This is a widely used question to tap welfare dualism and is included in e.g. the European Social Survey. Note that while treatment is exposure to a second generation immigrant, the policy outcome refers to the

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

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* to policy preferences.

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.

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.[10] 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 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.[11]

[10]The data for the general population are described in Bay, Finseraas, and Pedersen (2013)

[11]The latter number is from the sixth round of the European Social Survey (ESS). The scale is different in the ESS where it ranges from 0 to 10.

4 Identification of Peer Effects

We are interested in the effect of sharing room with at least one ethnic minority soldier on attitudes, i.e. a peer effect. 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:

$$Y_i = a + \beta_1 \overline{Y}_{-i} + \gamma_1 X_i + \gamma_2 \overline{X}_{-i} + \epsilon_i \tag{1}$$

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 "naive" regressions of our outcomes on the share of non-Norwegian friends in high school and the share of immigrants in the soldiers' home municipality.[12] Table 1 shows that having minority friends in high school is positively correlated with all three outcomes, albeit only statistically significant at the 7 percent level (t=1.86) for the variable *same rights*. Very similar results are seen in Panel B for the regression using the share of immigrants in the municipality of origin. These regressions would suggest strong support for the contact hypothesis if interpreted causally. This would involve a great leap of faith, however, as the estimation is likely to be severely biased by selection into friend networks and municipalities.

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

Table 1: Naive estimation of peer effects

	Same rights t2	Work ethics t2	Better country t2
Panel A: Minority friends			
Minority friends	0.138*	0.156**	0.230**
	(0.074)	(0.063)	(0.109)
Observations	533	534	533
Platoon FE	Yes	Yes	Yes
Panel B: Share of immigrants in the municipality			
Share of immigrants	1.592***	0.770*	1.011**
	(0.462)	(0.408)	(0.493)
Observations	584	585	584
R-squared	0.049	0.053	0.021
Platoon FE	Yes	Yes	Yes

Note: Robust standard errors adjusted for clustering on room in Panel A and municipality in Panel B. All regressions include a constant. *** p<0.01, ** p<0.05, * p<0.1

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 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. Such 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). Kling, Liebman, and Katz (2007) apply this design to 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, but the neighbors themselves do not otherwise play any role in the analysis. Similarly, Angrist and Lang (2004) investigate the effects of low-income peers in the classroom by estimating the effects on individuals where low-income individuals were bused in to the school as part of the Metco program. Again, the low income students themselves were not included in the regression, but only used to calculate peer characteristics.

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:

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

Where Y_{irt2} is one of the outcomes for individual i in room r at time period t2. 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 exposure at the room and team level net of any effect of exposure 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 on our three outcomes of interest, and reassuringly we find no effects of platoon exposure (see Appendix Table A2). We therefore conclude that the spillovers probably have a very small impact on our results. β_n 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.

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.

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

The soldiers completed three speeded ability tests of arithmetics, word similarities, and figures (see Sundet, Barlaug, and Torjussen 2004), prior to entering military service. We rely on the composite test score, which is an unweighted mean of the three subtests.[13] 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:

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

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 operationalizations). Since room allocation is randomized, we should not expect large and significant differences across pre-determined variables. Table 2 reports results from

[13] 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).

Table 2: Regressions of treatment status on pre-determined variables.

	Coeff	t	Standardized coeff	N
Same rights t1	-.13	1.20	-.05	589
Work ethics t1	-.16	1.50	-.07	552
Better country t1	.05	0.33	.01	552
Mother has high education	-.02	0.38	-.02	550
Father has high education	.00	0.07	.00	550
Mother is employed	-.09**	2.05	-.12	549
Father is employed	-.02	0.40	-.06	549
Parents are divorced	.00	0.01	.00	549
Plan to take higher education	.01	0.16	.01	551
IQ	-.01	0.09	-.00	601
F-test of joint significance	1.07 (p=.38)			

Note: 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

regressions of the treatment indicator dummy on the pre-determined variables.[14] 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.

Main Results

The main results are presented in Table 3. 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

[14]In the pre-analysis plan we write 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.

add all individual level controls, irrespective of whether there were a significant difference 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 have been identified in purely observational data (e.g. Alesina et al. 2001: 48, Ervasti and Hjerm 2012).

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. One 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. The null result on the policy preference variable 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 found a strong positive "effect" of having minority friends on the same question.

Table 3: Main results

	Same rights t2	Work ethics t2	Better country t2
Panel A: No controls			
Treated	0.037	0.196**	0.083
	(0.085)	(0.085)	(0.124)
Same rights t1	0.610***		
	(0.039)		
Work ethics t1		0.582***	
		(0.046)	
Better country t1			0.635***
			(0.043)
Platoon FE	Yes	Yes	Yes
Observations	534	535	534
Panel B: Control for baseline difference in mother's employment			
Treated	0.012	0.187**	0.080
	(0.084)	(0.085)	(0.124)
Same rights t1	0.619***		
	(0.039)		
Work ethics t1		0.586***	
		(0.047)	
Better country t1			0.635***
			(0.043)
Mother is employed	-0.068	-0.007	-0.152
	(0.111)	(0.116)	(0.153)
Platoon FE	Yes	Yes	Yes
Observations	531	532	531
Panel C: Full set of individual level controls			
Treated	0.000	0.187**	0.058
	(0.084)	(0.085)	(0.126)
Same rights t1	0.605***		
	(0.040)		
Work ethics t1		0.589***	
		(0.049)	
Better country t1			0.649***
			(0.043)
Platoon FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Observations	522	523	522

Note: Robust standard errors adjusted for clustering on room. All regressions include a constant. *** p<0.01, ** p<0.05, * p<0.1

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.

The results in Table 4 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 coefficients are, however, both insignificant, and the difference between them is insignificant as well. 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. This result differs from the one by Carrell, Hoekstra, and West (2015) in the US Air Force Academy, who find an effect on attitudes only for whites exposed to high-aptitude blacks.[15]

Robustness checks

In Table 3 we estimate linear regression models, despite that the dependent variables are categorical variables. In Table 5, 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).

[15] We arrive at the same conclusion if we instead define high and low ability minority soldier based on the total sample of minority soldiers and not only within platoons (results are available upon request).

Table 4: Treatment heterogeneity

	(1) Same rights t2	(2) Work ethics t2	(3) Better country t2
Treated high ability	0.242	0.129	-0.063
	(0.152)	(0.146)	(0.234)
Treated low ability	0.016	0.181*	0.065
	(0.096)	(0.100)	(0.156)
Same rights t1	0.568***		
	(0.042)		
Work ethics t1		0.573***	
		(0.052)	
Better country t1			0.647***
			(0.048)
F-test of diff high-low	1.8 (p=.19)	0.1 (p=.75)	0.25 (p=.62)
Observations	436	437	436
Platoon FE	Yes	Yes	Yes

Note: Robust standard errors adjusted for clustering on room in parentheses. All regressions include a constant. *** p<0.01, ** p<0.05, * p<0.1

We should point out, however, that we did not suggest to dichotomize *better country* in the pre-analysis plan, but present it here for completeness. Results are similar if we add individual level controls (see Table A3 in the Appendix).

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 A4, Panel A we report results when controlling for the share of high educated fathers[16] 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.[17] 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.

Next, 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). 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 effect on these variables should be smaller. We present results with and without rooms where there were 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

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

[17]While we are excluding the individual himself from the calculation of the shares so that we follow the "leave-out-oneself"-approach (see section 3), we acknowledge that 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 latter problem. The conclusions remain exactly the same, but the sample then only consists of 106 individuals (results are available upon request).

Table 5: Robustness checks

	Same rights t2	Work ethics t2	Better country t2
Panel A: Ordered probit regressions			
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)
Platoon FE	Yes	Yes	Yes
Observations	534	535	534
Panel B: Linear probability models of binary dependent variables			
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)
Platoon FE	Yes	Yes	Yes
Observations	534	535	534
Panel C: Share of minority soldiers in the room			
Treated Share	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)
Platoon FE	Yes	Yes	Yes
Observations	534	535	534

Note: Robust standard errors adjusted for clustering on room. All regressions include a constant. 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 dicotomized by recoding categories 5-7 to 1 and the others to 0. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 6: Placebo regressions

	Equality not imp All	Equality not imp Men	Equality not imp dummy All	Equality not imp dummy Men	Gender not imp All	Gender not imp Men
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)
Equality not important t1	0.578***	0.628***	0.376***	0.438***		
	(0.038)	(0.045)	(0.073)	(0.097)		
Gender not important t1					0.476***	0.443***
					(0.040)	(0.052)
Platoon FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	535	349	535	349	537	350

Note: Robust standard errors adjusted for clustering on room. All regressions include a constant. *** p<0.01, ** p<0.05, * p<0.1

room with an immigrant as fewer immigrant women serve.

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.

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 select the most publishable results ex post at the expense of accuracy (Miguel et al. 2014). 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. Taking into account the limits to power by testing multiple hypotheses forces us instead to be restrictive in the number of hypotheses tested and to impose pre-specified decision rules (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.[18] In the main results (Table 3), 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

One type of treatment effect heterogeneity is to test the effect of exposure to different types of second generation immigrants as we do above. Another type is to investigate whether the treatment effect varies across subgroups. In this section we present some exploratory analyses that were not part of pre-analysis plan. All findings in this section should therefore be interpreted as suggestive or as hypotheses to be tested in the future.

We start by investigating the effects for different types of people based on their prior exposure to immigrants. First we divide individuals into those from municipalities with above median and below median shares of immigrants in the population and interact this variable with the treatment indicator. Second we interact the treatment with the share of immigrant friends during last year in high school.

It is ambiguous what results to expect from these interactions as prior exposure might

[18]In the pre-analysis plan we write that we will follow the recommendations of Fink, McConnell, and Vollmer (2014) and use a method developed by Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001) to minimizes the false non-discovery rate. While the classical Bonferroni method–which corrects the critical level for rejection of an hypothesis to be the desired level a divided by the number of hypotheses, 0.05/4 in our case–is too restrictive when the hypotheses are correlated, which they are in our case, the false discovery rate method retain statistical power. In Benjamini and Hochberg's (1995) approach, the m p-values of the i hypotheses are ordered from low to high, and adjusted critical values are found by using the formula $p(i) <= a \times i/m$. The main advantage is that the method only adjusts the critical values based on other true hypotheses.

be correlated with more prior information about minorities as well as a host of omitted variables, in particular baseline attitudes toward minorities. We find it instructive to examine the two different measures of prior exposure since they are likely to pick up different things. Being exposed to a high share of immigrants in the municipality of origin is probably correlated with previous exposure, but not in such an intense fashion as being friends with immigrants. In fact, exposure in the municipality may increase rather than decrease misconceptions about immigrants as suggested in some of the literature on neighborhood effects (see Wessel 2009 a review). This negative effect is especially likely if there is micro-segregation within the municipality so that the immigrants are visible, but only as "others". Being friends with an immigrant is probably more linked to issues of selection and is also a function of the share of immigrants in the area.[19]

The results are seen in Table 7 where we include the two previous exposure variables as well as their interaction with the treatment variables. In the first set of columns we do not include the baseline attitudes as we want to capture the total reduced form effect, including the one via baseline attitudes. In column 1 we see that the treatment effect is larger for individuals coming from municipalities with a higher share of immigrants. This is consistent with a view that the individuals from such municipalities are updating their previous misconceptions regarding immigrants work ethics that they had gained before in their (perhaps segregated) exposure to immigrants. Consistency is not proof, however, and there are many other differences between the two groups of municipalities. In column two we see that there is no positive interaction effect for having had a higher share of immigrant friends and being exposed to second generation immigrants in the army. This is consistent with a view that these individuals already possessed the relevant information gained from more intense exposure to immigrants but it may of course also be related to the fact that they already had a more positive view of immigrants.

In columns three and four we add the baseline attitudes toward immigrant work ethics as a control and the results all point in the same direction, albeit the immigration share

[19] It is likely that the self-selection problem is more limited in the analysis of municipality of origin share of immigrants as the individuals have not chosen to live there themselves, but rather their parents. Nonetheless, the attitudes of the parents is an obvious omitted factor.

interaction is not statistically significant. We also add an interaction between baseline attitudes and treatment in column 5 and we find no statistically significant difference in the treatment effect based on pre-existing attitudes but the coefficient for the interaction is negative which is consistent with a larger positive effect for those who did not have as positive attitudes to start with. In column six we add all interactions terms to the regression for completeness and we see that the only interaction being statistically significant at the 10 percent level is the one for high immigration share in the municipality. In the Appendix we present the corresponding interactions for the other two main variables (see Tables A5 and A6). There are no statistically significant interaction terms for the same rights variable but for the better country variable there is a statistically significant positive interaction between treatment and coming from a municipality with a high share of immigrants, albeit only without controlling for baseline attitudes.

We also interact treatment with all our control variables used for tests of balance one at a time (available upon request). Only one out of the seven regressions produce a statistically significant interaction term and this is only at the 10 percent level. The variable is planning to take higher education and the results indicate a smaller effect on individuals planning to take higher education.

7 Conclusion

The social, economic and political consequences of immigration is a major topic across the developed countries. 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. An important shortcoming in the existing literature is the ability to reach causal statements, which we address by running a field experiment where personal contact with minorities is randomized, and takes place in a context which allows clear theoretical expectations of reduced prejudice due to personal contact. We pre-registered the study, i.e. all decisions on hypotheses, variable operationalizations and empirical model specifications were decided prior to data collection.

Table 7: Exploratory analysis of heterogenous effects on immigrants' *work ethics* based on previous exposure

	(1) Municipality share	(2) Friends	(3) Municipality share	(4) Friends	(5) Baseilne 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 immigrant share	-0.010		0.049			0.043
	(0.094)		(0.073)			(0.073)
Treated*High imm	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*Immigrant 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
Individual controls	No	No	No	No	No	No

Note: Robust standard errors adjusted for clustering on room in parentheses. All regressions include a constant.
** p<0.01, ** p<0.05, * p<0.1

We find quite 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 more 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.

Contrary to our expectation, the improved view on immigrants' work ethic is not reflected in reduced support for welfare dualism, as we find small and statistically insignificant treatment effects on 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 theoretically the one closest to treatment, but does not spill over to affecting welfare policy preferences. The results therefore support the view that policy preferences are sluggish and hard to change (see Kuzuemko et al., 2015, for recent evidence of how hard it is to change welfare policy preferences). Our results indirectly suggest that other concerns–which could be 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 and McGann 1995)–are more important for support for welfare dualism than prejudiced views on ethnic differences in work ethics.

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, older women. Furthermore, our sample appears to be somewhat more positive towards minorities than the population of young men. In addition, 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 have similarities to our context, yet 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. In light of the small and insignificant treatment effects we find on two of our outcomes, studies are unlikely to find substantive effects of contact in such environments.

References

Alesina, Alberto, and Edward L. Glaeser. 2004. *Fighting Poverty in the US and Europe: A World of Difference*. Oxford: Oxford University Press.

Alesina, Alesina, Edward Glaeser, and Bruce Sacerdote. 2001. "Why doesn't the US have a European-style welfare system?" *Brookings Papers on Economic Activity* 2001(2).

Allport, Gordon W. 1954. *The Nature of Prejudice*. Reading: Addison-Wesley.

An, Weihua. 2011. "Models and Methods to Identify Peer Effects." *The Sage Handbook of Social Network Analysis. London: Sage* pp. 515–532.

Angrist, Joshua D. 2014. "The perils of peer effects." *Labour Economics* 30: 98–108.

Angrist, Joshua D., and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review* 94(5): 1613–1634.

Austen-Smith, David, and Michael Wallerstein. 2006. "Redistribution and Affirmative Action." *Journal of Public Economics* 90(10): 1789–1823.

Bay, Ann-Helén, Henning Finseraas, and Axel West Pedersen. 2013. "Welfare dualism in two Scandinavian welfare states: Public opinion and party politics." *West European Politics* 36(1): 199–220.

Benjamini, Yoav, and Daniel Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing under Dependency." *Annals of Statistics* 29(4): 1165–1188.

Benjamini, Yoav, and Yosef Hochberg. 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, Lawrence 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, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." *American Economic Review* 96(5): 1890–1905.

Brady, David, and Ryan Finnigan. 2013. "Does Immigration Undermine Public Support for Social Policy?" *American Sociological Review* 79(1): 17–42.

Careja, Romana, and Patrick Emmenegger. 2013. "Keeping Them Out: Migration and Social Policies in the 'Reluctant Countries of Immigration'." In *Citizenship and Identity in the Welfare State*, ed. Andrzej M. Suszycki, and Ireneusz P. Karolewski. Baden-Baden: Nomos Verlag.

Carrell, Scott E., Mark Hoekstra, and James E. West. 2015. "The Impact of Intergroup Contact on Racial Attitudes and Revealed Preferences." NBER Working Paper No. 20940.

Clemens, Michael A. 2011. "Economics and Emigration: Trillion-Dollar Bills on the Sidewalk?" *Journal of Economic Perspectives* pp. 83–106.

Dahlberg, Matz, Karin Edmark, and Heléne Lundqvist. 2012. "Ethnic Diversity and Preferences for Redistribution." *Journal of Political Economy* 120(1): 41–76.

Dixon, John, Kevin Durrheim, and Colin Tredoux. 2005. "Beyond the Optimal Contact Strategy: A Reality Check for the Contact Hypothesis." *American Psychologist* 60(7): 697.

Dustmann, Christian, Tommaso Frattini, and Ian P. Preston. 2013. "The Efect of Immigration along the Distribution of Wages." *Review of Economic Studies* 80(1): 145–173.

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

Ervasti, Heikki, and Mikael Hjerm. 2012. "Immigration, trust and support for the welfare state." In *The Future of the Welfare State: Social Policy Attitudes and Social Capital in Europe*, ed. Patrick Emmenegger, Bruno Palier, and Martin Seeleib-Kaiser. Camberley, UK: Edward Elgar Publishing.

Fink, Günther, Margaret McConnell, and Sebastian Vollmer. 2014. "Testing for Heterogeneous Treatment Effects in Experimental Data: False Discovery Risks and Correction Procedures." *Journal of Development Effectiveness* 6(1): 44–57.

Finseraas, Henning. 2008. "Immigration and Preferences for Redistribution: An Empirical Analysis of European Survey Data." *Comparative European Politics* 6(4): 407–431.

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

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

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. 2010. "How Much Does Immigration Boost Innovation?" *American Economic Journal: Macroeconomics* 2(2): 31–56.

Kitschelt, Herbert, and Anthony J. McGann. 1995. *The Radical Right in Western Europe: A Comparative Analysis.* Ann Arbor: University of Michigan Press.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83–119.

Korpi, Walter, and Joakim Palme. 2003. "New Politics and Class Politics in the Context of Austerity and Globalization: Welfare State Regress in 18 Countries, 1975–95." *American Political Science Review* 97(3): 425–446.

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015. "How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments." *American Economic Review* 105(4): 1478–1508.

Larsen, Christian Albrekt. 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.

Luttmer, Erzo F.P., and Monica Singhal. 2011. "Culture, Context, and the Taste for Redistribution." *American Economic Journal: Economic Policy* 3(1): 157–179.

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

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

Miguel et al., Edward. 2014. "Promoting Transparency in Social Science Research." *Science* 343(6166): 30–31.

Peri, Giovanni. 2012. "The Effect of Immigration on Productivity: Evidence from US States." *Review of Economics and Statistics* 94(1): 348–358.

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

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

Przeworski, Adam, and John D. Sprague. 1986. *Paper Stones.* Chicago: University of Chicago Press.

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

Sacerdote, Bruce. 2011. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen J. Machin, and Ludger Woessmann. Elsevier.

Semyonov, Moshe, Rebeca Raijman, and Anastasia Gorodzeisky. 2006. "The Rise of Anti-Foreigner Sentiment in European Societies, 1988-2000." *American Sociological Review* 71(3): 426–449.

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

Stephens, John D. 1979. *The Transition From Capitalism to Socialism.* London: Macmillian.

Sundet, Jon Martin, Dag G. Barlaug, and Tore M. Torjussen. 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, Eric M. 2011. "Trust, Diversity, and Segregation in the United States and the United Kingdom." *Comparative Sociology* 10(2): 221–247.

Van der Waal, Jeroen, Peter Achterberg, Dick Houtman, Willem De Koster, and Katerina Manevska. 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 Laar, Colette, Shana Levin, Stacey Sinclair, and Jim Sidanius. 2005. "The Effect of University Roommate Contact on Ethnic Attitudes and Behavior." *Journal of Experimental Social Psychology* 41(4): 329–345.

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

Vernby, Kåre. 2013. "Inclusion and Public Policy: Evidence from Sweden's Introduction of Noncitizen Suffrage." *American Journal of Political Science* 57(1): 15–29.

Wessel, Terje. 2009. "Does diversity in urban space enhance intergroup contact and tolerance?" *Geografiska Annaler: Series B, Human Geography* 91(1): 5–17.

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

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

Attrition

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

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:

$$Attrition_i = \alpha_J + \beta_1 Treatment + \beta_n X + \epsilon \tag{A1}$$

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 also run the following regression:

$$Attrition_i = alpha_J + \beta_1 Y_{t1} + \beta_n X + \epsilon, \tag{A2}$$

where 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 A1: Tests for non-random attrition

	(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)
Platoon FE	Yes	Yes	Yes	Yes	Yes
Individual controls	No	Yes	No	No	No
Observations	783	765	826	826	826

Note: Robust standard errors adjusted for clustering on room. All regressions include a constant. *** p<0.01, ** p<0.05, * p<0.1

Robustness checks

Table A2: Regressions examining the effect of exposure at the platoon level

	Same rights t2	Work ethics t2	Better country t2
Treated platoon	0.045	0.029	0.119
	(0.140)	(0.143)	(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
Observations	421	422	421

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

Table A3: Regressions (LPM) on dichotomized DVs with controls

	Same rights t2	Work ethics t2	Better country t2
Panel A: Control for baseline difference in mother's employment			
Treated	-0.022	0.068*	0.095**
	(0.041)	(0.042)	(0.046)
Same rights t1	0.482***		
	(0.042)		
Work ethics t1		0.466***	
		(0.040)	
Better country t1			0.500***
			(0.040)
Mother is employed	-0.026	-0.046	-0.022
	(0.053)	(0.067)	(0.056)
Platoon FE	Yes	Yes	Yes
Observations	531	532	531
Panel B: Full set of individual levels controls			
Treated	-0.027	0.071*	0.094**
	(0.040)	(0.043)	(0.047)
Same rights t1	0.462***		
	(0.045)		
Work ethics t1		0.472***	
		(0.040)	
Better country t1			0.486***
			(0.042)
Platoon FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Observations	522	523	522

Note: Robust standard errors adjusted for clustering on room. All regressions include a constant. *** p<0.01, ** p<0.05, * p<0.1

Table A4: Regressions with share for high educated fathers

	Same rights t2	Work ethics t2	Better country t2
Panel A: Control for share of high educated fathers			
Treated	0.036	0.204**	0.078
	(0.085)	(0.084)	(0.123)
Same rights t1	0.610***		
	(0.039)		
Work ethics t1		0.587***	
		(0.047)	
Better country t1			0.633***
			(0.043)
Share of high educated fathers	-0.055	0.270	-0.197
	(0.179)	(0.172)	(0.266)
Platoon FE	Yes	Yes	Yes
Observations	534	535	534
Panel B: Share of high educated fathers w/o treated			
Share of high educated fathers	-0.059	0.250	-0.204
	(0.179)	(0.174)	(0.263)
Same rights t1	0.609***		
	(0.039)		
Work ethics t1		0.579***	
		(0.047)	
Better country t1			0.633***
			(0.043)
Platoon FE	Yes	Yes	Yes
Observations	534	535	534

Note: Robust standard errors adjusted for clustering on room. All regressions include a constant. *** p<0.01, ** p<0.05, * p<0.1

Table A5: Exploratory analysis of heterogenous effects on *better country* based on previous exposure

	(1) Municipality share	(2) Friends	(3) Municipality share	(4) Friends	(5) Baseilne 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_imm	0.143		0.093			0.086
	(0.131)		(0.108)			(0.110)
T*high_imm	0.279		0.001			-0.086
	(0.294)		(0.225)			(0.256)
Friends_imm		0.198*		0.119		0.111
		(0.113)		(0.113)		(0.113)
T*friends_imm		0.123		0.088		0.083
		(0.280)		(0.258)		(0.277)
Baseline			0.631***	0.632***	0.615***	0.610***
			(0.044)	(0.043)	(0.050)	(0.050)
T*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
Individual controls	No	No	No	No	No	No

Note: Robust standard errors adjusted for clustering on room in parentheses. All regressions include a constant.
** p<0.01, ** p<0.05, * p<0.1

Table A6: Exploratory analysis of heterogenous effects on *same rights* based on previous exposure

	(1) Municipality share	(2) Friends	(3) Municipality share	(4) Friends	(5) Baseilne attitudes	(6) All
Treated	-0.337**	0.101	-0.030	0.167	0.098	0.296
	(0.138)	(0.290)	(0.124)	(0.249)	(0.340)	(0.410)
High_imm	0.130		0.031			0.014
	(0.106)		(0.086)			(0.084)
T*high_imm	0.533**		0.122			0.214
	(0.228)		(0.190)			(0.187)
Friends_imm		0.172**		0.021		0.019
		(0.083)		(0.078)		(0.078)
T*friends_imm		-0.139		-0.101		-0.146
		(0.196)		(0.169)		(0.174)
Baseline			0.604***	0.615***	0.614***	0.619***
			(0.040)	(0.038)	(0.044)	(0.044)
T*baseline					-0.018	-0.053
					(0.096)	(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
Individual controls	No	No	No	No	No	No

Note: Robust standard errors adjusted for clustering on room in parentheses. All regressions include a constant.
** p<0.01, ** p<0.05, * p<0.1