European Journal of Political Research ●●: ●●–●●, 2017
doi: 10.1111/1475-6765.12199

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

HENNING FINSERAAS[1] & ANDREAS KOTSADAM[2]
[1] *Institute for Social Research, Oslo, Norway;* [2] *Ragnar Frisch Centre for Economic Research, University of Oslo, Norway*

Abstract. This article explores the causal effect of personal contact with ethnic minorities on majority members' views on immigration, immigrants' work ethics, and support for lower social assistance benefits to immigrants than to natives. Exogenous variation in personal contact is obtained by randomising 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, the study spells out why the army can be regarded as an ideal contextual setting for exposure to reduce negative views on minorities. The study finds a substantive effect of contact on views on immigrants' work ethics, but small and insignificant effects on support for welfare dualism, as well as on views on whether immigration makes Norway a better place in which to live.

Keywords: contact theory; field experiment; immigration attitudes; welfare state

Introduction

The topic of this article is the degree to which majority–minority contact influences anti-immigration sentiments, with particular emphasis on the relationship between immigration, diversity and the welfare state. In the United States, majority–minority conflicts have long been linked to white Americans' welfare state preferences (Gilens 1995). Since the work by Alesina et al. (2001) and Alesina and Glaeser (2004), research on the relationship between immigration, diversity and native Europeans' welfare state preferences has thrived. However, empirical literature provides no consensus on the effects of ethnic diversity on welfare state preferences (compare, e.g., Eger 2010; Dahlberg et al. 2012; Brady & Finnigan 2013).

The empirical literature on immigration and welfare state preferences has three important shortcomings that we address in this article. First, in the existing literature there is a lack of attention to context: The degree to which 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 (Dancygier 2010; Enos 2014, 2016). Competition between an in-group and out-groups over scarce resources, social rights and social status can cause out-group prejudice (see, e.g., Bobo 1999; Semyonov et al. 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 that the degree of social segregation is 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 conditions, we should expect integration and the de-emphasis of ethnic boundaries, while without 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 finds that contact can reduce prejudice even when these conditions are absent; however, the effect is larger when the conditions are met, particularly 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 the existing empirical literature on the consequences of ethnic diversity does not take this contextual distinction into account (e.g., Alesina et al. 2001; Senik et al. 2009; Ervasti & Hjerm 2012; Brady & Finnigan 2013). 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 therefore have set up a research design that is informative about the role of social segregation for welfare state support.

The second shortcoming is that the causal relationship between immigration or views on immigrants, on the one hand, and support for the welfare state, on the other, 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 for 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 United States. We suspect that the impact of increasing ethnic diversity on European welfare states can be different from the impact in the United States, 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 on which most citizens rely for periods of their lifetime 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 United States compared to Europe might contribute to explain why there is a strong link between views on minorities and support for the welfare state in America (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 cultural threat (Van der Waal et al. 2010; Larsen 2011; Bay et al. 2013; Brady & Finnigan 2013). This line of policy has been actively advocated and pursued in some European countries (Careja & Emmenegger 2013), with 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 conducted an explicit test of contact theory and its relevance for support for welfare dualism. We ran 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 the context of 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 soldiers cannot determine with whom they want to serve. To ensure that majority–minority contact and cooperation is real and not superficial, we have defined '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 randomised soldiers to different rooms and hence to direct personal contact with minorities. Next we compared outcomes for majority soldiers who were randomised to share a room with a minority soldier to majority soldiers who were randomised to share a room that 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 ethics. A large literature points to perceptions about different groups' work ethics as important for their views on welfare spending/benefits that are directed to that group (see, e.g., Hasenfeld & Rafferty 1989; Gilens 1995; Dyck & Hussey 2008; Rosenthal et al. 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 this article is organised as follows. First, we describe the field experiment before we describe the construction of the key variables. Then we explain how we deal with the well-known empirical challenges involved in estimating the effect of exposure to others (peer effects). Following that, we discuss the treatment effect equations that we estimate before we present the empirical results. To avoid concerns that the data analysis is a 'fishing expedition' (Humphreys et al. 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 that we submitted to the American Economic Association (AEA) Registry prior to the data collection. Thus, our hands are tied and we cannot choose the empirical specification that would yield the results that we, for ideological or publication strategic reasons, might prefer. In the article, we explicitly mention when the analysis deviates from the pre-analysis plan.

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 have their first day in the army at Sessvollmoen military camp, close to Oslo Gardermoen airport. When they meet on that first day the soldiers do not know each other, and they do not know who will be their roommates. The soldiers go through a programme of medical and psychological testing at the camp. We obtained permission to set up a station in this programme 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 programme at Sessvollmoen, the soldiers board planes to Northern Norway to start their recruit period. When they arrive in Northern Norway, they are bussed to a number of different military camps where they are 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 constitute a team within the platoon. Thus, room sharing during the recruit period constitutes 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 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 base, 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 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 spreadsheet that they were instructed to use to randomise soldiers within platoons into rooms. The randomization occurred when the personnel officer entered the list of soldiers in the platoon and the size of the rooms. Copies of the spreadsheets were emailed to Norwegian Defence Research Establishment (FFI) for verification. The procedure allows for the construction of a treatment group consisting of soldiers with an ethnic Norwegian background who were randomised into a room with at least one soldier with an ethnic minority background (see definitions of 'majority' and 'minority' backgrounds below). The control group consists of soldiers who did not share a room with an ethnic minority soldier. We surveyed the soldiers for the second time at the end of the recruit period so 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 spreadsheet to randomise 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 randomisation from battalion commanders down the hierarchy to personnel officers.[1]

The sample

Norway has 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 per cent of soldiers are unsure whether they would have served in the military if it was completely voluntary. Since the Army has a degree of control over who they allow to serve, the soldiers are positively selected on background characteristics like grades in high school. Similar to, for example, laboratory experiments, positive selection into the army does not invalidate our experiment, but it might have consequences for the external validity of our results. We will 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 by the time of the second round. A high dismissal rate is normal during the recruit period. Importantly, we tested and confirmed that attrition in the panel is unrelated to treatment status as well as to baseline values of the outcome variables (see Table A2 in the Online Appendix and the discussion there). The rooms vary in occupancy between 3 and 12 persons, but 73 per cent of our sample lived in six-person rooms. Out of the 577 soldiers, 5 per cent (27 soldiers) had a minority ethnic background and 20 per cent (116 soldiers) shared a room with at least one ethnic minority soldier. These soldiers constituted the treatment group. Ten of the majority soldiers shared a 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 0 to s 40 per cent.

In Table A1 in the Online Appendix, 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 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 a 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.

Table 1. Regressions of treatment status on predetermined variables

	Coefficient	t	Standardised 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 ($p = 0.28$)		

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

Key variable operationalisations

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

Ethnic background

The main independent variable is a dummy that equals 1 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 she or he answers categories 5 to 8.[2] We chose to emphasise non-Western ethnic background rather than foreign background as the effect is likely to be larger for this group. Having a parent from, for example, 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 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 analysing 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. 1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree.

This question 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 that can explain why contact might decrease support for dualism. First, we test whether the respondents think the work ethics of immigrants and natives is more similar if exposed to minorities. View on immigrants' work ethics (*Work ethics*) is measured with the following question:

> Do you agree or disagree with the statement: In general, immigrants have poorer work ethics than Norwegians. 1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree.

This question is directly linked to the experiences of the soldiers as they work together in the Army, but again involves generalisation from the second-generation immigrants to the greater immigrant population.

Second, we explore whether there is an effect on attitudes towards immigrants more generally 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 seven-point scale where 1 = Worse place to live, and 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 per cent in our sample disagree or disagree strongly that immigrants should not have the same rights, compared to about 41 per cent in the general population. They are also less likely to agree or strongly agree with the claim that immigrants have poorer work ethics: 8 per cent agree/agree strongly in our sample, versus 22 per cent in the general population. For the question on the overall impact of immigration, however, there is no difference, as about 42 per cent 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 generalises to a broader out-group than of the treatment. Previous studies have found that positive effects of contact tend to generalise 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 will return to this issue in the interpretation of the results.

Identification of peer effects

We are interested in the effect of sharing a 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 most commonly estimated model of peer effects (Sacerdote 2011) is some version of the following equation:

$$Y_i = a + \beta_1 \bar{Y}_{-i} + \gamma_1 X_i + \gamma_2 \bar{X}_{-i} + Q_i \quad (1)$$

where *Yi* is the outcome of interest for individual *i* which is thought to be a function of the average outcomes of the peers ($\bar{Y}\ -i$), the individuals own characteristics (*Xi*), and the characteristics of the peers ($\bar{X}-i$). 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 Online 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 one 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 behaviour 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 studies 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 (1) the individuals assumed to be affected, and (2) the peers assumed to provide the mechanisms for the peer effects. Separation implies that the individuals with the background characteristic providing the suggested mechanism (i.e., those with an ethnic minority background) are excluded from the sample of those assumed to be affected (i.e., those with an ethnic majority background).

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:

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

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 randomisation 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. Randomisation 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 Table A4 in the Online Appendix). 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 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 the minority roommate if the minority roommate has a higher relative ability score. This test is inspired by Carrell et al. (2015), who find that only high-aptitude blacks are able to influence the attitudes of whites. We expect roommate ability to matter insofar as negative views on minorities reflect statistical discrimination that will be more strongly updated if one has contact with a high-ability minority person (Carrell et al. 2015).

The soldiers completed three speeded ability tests of arithmetics, word similarities and figures (see Sundet et al. 2004) prior to entering military service. We rely on the composite test score, which is an unweighted mean of the three subtests.[8] 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 model:

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

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). β1 and β2 test whether high ability and low ability groups differ from the control group. We are also interested in the difference between β1 and β2 and will rely on F tests to examine whether they are statistically significant from each other.

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

As is 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 that is statistically significant. In light of the generally small differences and the small F value in the joint test, we nonetheless conclude that the randomisation 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 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.

Table 2. Main results

	(1) Same rights t2	(2) Work ethics t2	(3) Better country t2
Panel A: No controls			
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^2	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^2	0.390	0.332	0.379
Platoon FE	Yes	Yes	Yes
Panel C: Full set of individual-level 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^2	0.396	0.341	0.411
Platoon FE	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes

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

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 0.04, which is small in light of the standard deviation of *Same rights* (mean = 3.5, standard deviation = 1.1). Thus, we conclude that sharing a 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).

Moving to *Work ethics*, we find positive treatment coefficients which are significant at the 5 per cent 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 0.2. Since the standard deviation of the dependent variable is 1, the difference of 0.2 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 ethics of immigrants.

Our interpretation is that by sharing a room and cooperating on task solving, treated soldiers have received information on majority–minority differences in work ethics and updated their prior views on these differences. Clearly, since the outcome concerns the work ethics of the overall immigrant population, the effect generalises 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 views on work ethics are 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 in which to live. Moreover, the treatment coefficient is less than 0.1, 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 the measured ability of the minority soldier. We do so by creating one dummy representing whether the soldier shared a 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 3 show that there are indications of treatment heterogeneity on the *Same rights* question as the coefficient for high ability treated is larger than the one for low ability treated. The difference between coefficients is, however, insignificant. For *Work ethics*, 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]

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: (1) estimate ordered probit models rather than ordinary least squares models; (2) dichotomise the dependent variables; (3) rely on a continuous measure of share

Table 3. Treatment heterogeneity

	(1) Same rights t2	(2) Work ethics t2	(3) Better country t2
Treated high ability	0.389* (0.202)	0.126 (0.222)	–0.027 (0.163)
Treated low ability	0.034 (0.102)	0.178* (0.106)	0.101 (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 = 0.49)	0.28 (p = 0.60)	0.01 (p = 0.93)
Observations	391	392	391
R^2	0.412	0.348	0.406
Platoon FE	Yes	Yes	Yes

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

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 (0.090)	–0.072 (0.105)	0.008 (0.026)	0.007 (0.031)	–0.044 (0.050)	–0.044 (0.062)
Baseline	0.578*** (0.038)	0.628*** (0.045)	0.376*** (0.073)	0.438*** (0.097)	0.476*** (0.040)	0.443*** (0.052)
Observations	535	349	535	349	537	350
R^2	0.326	0.348	0.163	0.206	0.241	0.240
Platoon FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: 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.

of minority soldiers in the room rather than the dummy treatment indicator; and (4) 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 socioeconomic 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 per cent 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 a 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 female soldiers present (in some places male and female soldiers share a room) since exposure to female soldiers might change views on gender equality, and is correlated with the probability of sharing a 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.

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, particularly 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.

Concluding discussion

In this article we have examined the effects of direct personal contact with ethnic minorities on majority members' support for welfare dualism, views on immigrants' work ethics and views on the consequences of immigration. By running a field experiment with randomised personal contact with minorities in a context that 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 ethics. 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 a weaker work ethic than Norwegians. We interpret this result as reflecting the existence of a negative bias in the soldiers' views on minorities' work ethics, which becomes updated and reduced from observing minorities' work ethics through direct personal contact and cooperation. Since treatment is exposure to second-generation minorities, while the work ethics question is about the overall immigrant population, the treatment effect appears to generalise 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 ethics is not reflected in reduced support for welfare dualism. The same is true for views on whether immigration makes the country a better place in which 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 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 ethics 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 ethics that is a direct result of the randomisation, 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 does observe will also represent the influence of correlated ideological traits.

Recent developments in 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) investigated roommate integration for interracial college roommates and found 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 individuals on self-reported wellbeing in the room (0.13, robust standard error = 0.13, $p = 0.33$). Furthermore, Shook and Fazio (2008a, 2008b) found that interracial relationships seem to entail less quality than same-race relationships. In particular, the interracial rooms were more likely to dissolve and 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, for example, 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, suggesting 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 & 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 to be 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 a 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 generalise 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 who are the subjects of the outcomes. Thus, treated soldiers might not conceive these soldiers as representative of the overall immigrant population. However, treatment *does* generalise from second generation to the overall immigrant population on the work ethics 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 rights 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 generalise 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 in workplaces, classrooms and 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.

Acknowledgments

We would like to thank the FFI, in particular Frank Steder and Torbjørn Hanson. This study could not have been conducted without the help of FFI and their project 'Research on Cohorts'. A previous version of this article has been circulated with the title 'Does

Personal Contact with Ethnic Minorities Affect Support for Welfare Dualism? Evidence from a Field Experiment'. We thank Ada Fuglset, Eirik Strømland and Wiktoria Szczesna for excellent research assistance. Thanks also to the soldiers and staff at the North Brigade. We are grateful to Jon Fiva, Don Green, Øyvind Skorge, Axel West Pedersen, Margaret Peters, Terje Wessel and seminar participants at University of Oslo, Oslo University College, Norwegian Social Research, Institute for Social Research, University of Oxford, the 2015 annual meeting of Norwegian Political Scientists and the 2015 annual meeting of the American Political Science Association for helpful comments and suggestions. The project is part of the research activities at the Centre for the Study of Equality, Social Organisation and Performance (ESOP) at the Department of Economics, University of Oslo. ESOP is supported by the Research Council of Norway. Funding from the Norwegian Research Council (grant numbers 236801 and 236992) is also acknowledged.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table A1: Descriptive statistics
Table A2: Tests of non-random attrition
Table A3: Naive estimation of peer effects
Table A4: Regressions examining the effect of exposure at the platoon level
Table A5: Robustness checks
Table A6: Regressions (LPM) on dichotomized DVs with controls
Table A7: Regressions with share for high educated fathers
Table A8: Exploratory analysis of heterogeneous effects on immigrants *work ethics* based on previous exposure
Table A9: Exploratory analysis of heterogeneous effects on immigrants *better country* based on previous exposure
Table A10: Exploratory analysis of heterogeneous effects on immigrants *same rights* based on previous exposure

Notes

1. The personnel officers who did not follow the randomisation 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.
2. It is not obvious that Oceania should be coded as non-Western, but the decision to do so does not influence the results.
3. In the pre-treatment survey, the Spearman correlation between *same rights* and *work ethics* is 0.47; between *same rights* and *better country* it is 0.48; it is 0.34 between *work ethics* and *better country*. In the post-treatment survey, the corresponding Spearman correlations are 0.56, 0.48, and 0.41.
4. The data for the general population are described in Bay et al. (2013).

5. 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.
6. The sample is restricted to soldiers with a majority ethnic background.
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 a 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 the estimator relies on less restrictive assumptions than the IV strategy.
8. The scores are reported in stanine ('standard nine') units, a method of standardising raw scores into a nine-point standard scale with a normal distribution (mean = 5; standard deviation = 2).
9. We write in the pre-analysis plan that we will analyse 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.
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 of minority soldiers and not only within platoons.
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 the present study, the outcome is one where information is likely to be updated as a result of the daily interaction.

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). Available online at: http://post.economics.harvard.edu/hier/2001papers/2001list.html

Allport, G.W. (1954). *The nature of prejudice*. Reading, MA: 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. et al. (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 & 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 20940. Cambridge, MA: National Bureau of Economic Research.

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: Cambridge University Press.

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*. Cheltenham: Edward Elgar.

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, MI: 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.

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

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

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.

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 & L. Woessmann (eds), *Handbook of the economics of education*. Amsterdam: North Holland.

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.

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 and 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 and Intergroup Relations* 14(3): 399–406.

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., Koster, W. de & 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.

Address for correspondence: Henning Finseraas, Institute for Social Research, Postbox 3233 Elisenberg, 0208 Oslo, Norway. E-mail: henning.finseraas@samfunnsforskning.no