Supplementary Information for

Parochialism, Social Norms, and Discrimination Against Immigrants

Donghyun Danny Choi, Mathias Poertner, and Nicholas Sambanis

Donghyun Danny Choi, Mathias Poertner and Nicholas Sambanis
E-mail: dhchoi@sas.upenn.edu, mathias.poertner@berkeley.edu, sambanis@sas.upenn.edu

This PDF file includes:

Supplementary text
Figs. S1 to S7
Tables S1 to S20
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1820146116

Supporting Information Text

1. Materials and Methods

The full replication code that produces this report will be made available at the Penn Identity and Conflict Lab's webpage.

Experimental design. The experiment focuses specifically on exploring whether host populations reward immigrants for their enforcement of social norms that are well-established in the host society, and whether such behavior is sufficient to offset the discrimination towards immigrants that are driven by intergroup differences in ascriptive characteristics. We focus on the willingness of the host population to offer assistance to immigrants in the context of common day-to-day interactions regarding the enforcement of the littering norm. The setup and procedures are diagrammatically presented in Figure S1, shown below.

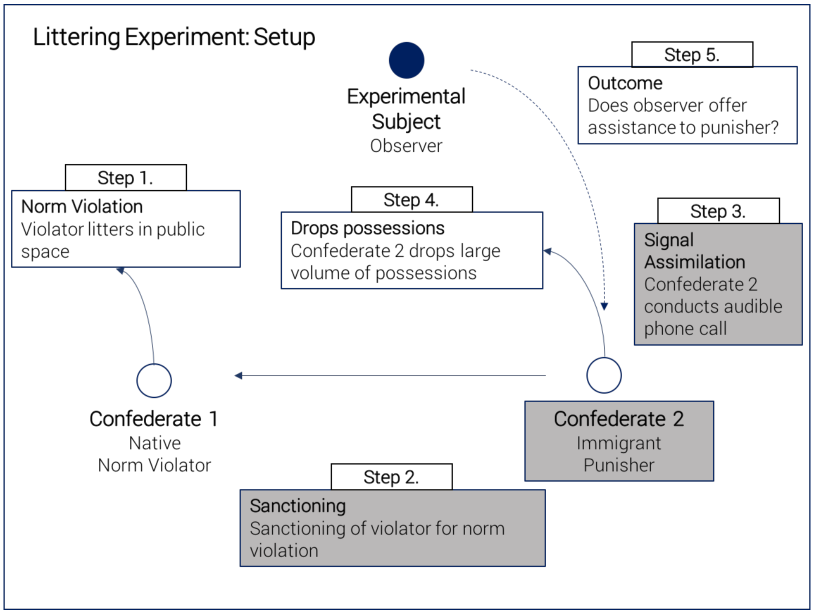

Littering Experiment: Setup
Experimental Subject
Observer
Step 5.
Outcome
Does observer offer assistance to punisher?
Step 1.
Norm Violation
Violator litters in public space
Step 4.
Drops possessions
Confederate 2 drops large volume of possessions
Step 3.
Signal Assimilation
Confederate 2 conducts audible phone call
Confederate 1
Native
Norm Violator
Confederate 2
Immigrant
Punisher
Step 2.
Sanctioning
Sanctioning of violator for norm violation

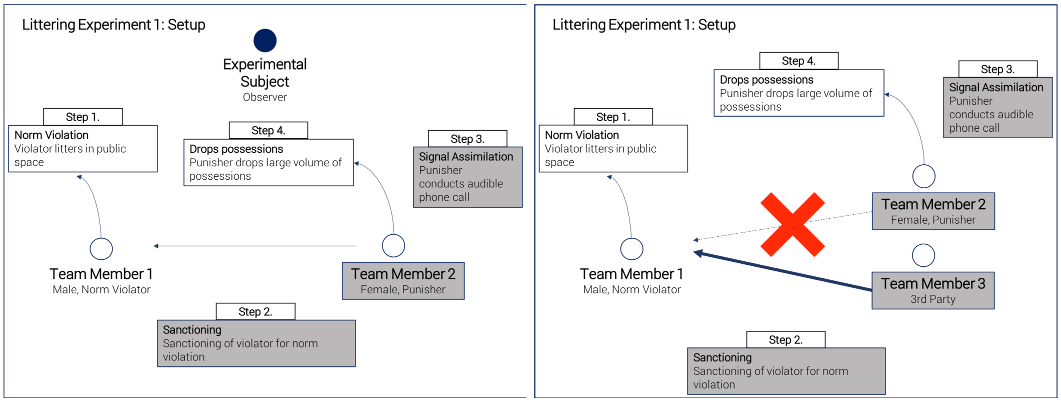
Fig. S1. Experimental setup

- **Step 1**: A German male confederate (the "violator") is instructed to violate a widely shared norm against littering in a train station platform in front of unknowing experimental subjects, as in the closely related experiment by Balafoutas et al. (2016).
- **Step 2**: A second female confederate sanctions the violator by politely, albeit firmly, asking the violator to pick up his trash. The violator picks up his trash and leaves the scene.
- **Step 3**: The female confederate conducts an audible phone call within earshot of the experimental subject in either German or their mother tongue.
- **Step 4**: In the midst of the phone call, the female confederate drops her possessions (a large volume of groceries that disperse and are hard to pick up) and appears to be in need of assistance.
- **Step 5**: We observe in step 5 whether the punisher receives assistance from experimental subjects who have observed the sequence of events. The main behavioral outcomes of the study are (a) whether the female confederate receive *any* assistance from bystanders; and (b) the *proportion* of bystanders who offered assistance.

Treatment manipulation. We experimentally manipulated two core dimensions of the intervention.

- **Dimension 1**: Ascriptive characteristics of female confederate (punisher).
  1. Immigrant confederate wearing a hijab
  2. Immigrant confederate wearing plain clothing without hijab
  3. Immigrant confederate wearing plain clothing with a Christian cross
  4. Native confederate (German)
- **Dimension 2**: Enforcement of anti-littering norm. Figure S2 provides a diagrammatic representation of how treatment dimension 2 was manipulated.

1. Anti-littering norm is enforced by the female confederate (punisher) who is later in need of assistance.
2. Anti-littering norm is enforced by a different confederate (third party).


Littering Experiment 1: Setup
Experimental Subject
Observer
Step 1.
Norm Violation
Violator litters in public space
Step 4.
Drops possessions
Punisher drops large volume of possessions
Step 3.
Signal Assimilation
Punisher conducts audible phone call
Team Member 1
Male, Norm Violator
Team Member 2
Female, Punisher
Step 2.
Sanctioning
Sanctioning of violator for norm violation
Littering Experiment 1: Setup
Step 4.
Drops possessions
Punisher drops large volume of possessions
Step 3.
Signal Assimilation
Punisher conducts audible phone call
Step 1.
Norm Violation
Violator litters in public space
Team Member 2
Female, Punisher
Team Member 1
Male, Norm Violator
Team Member 3
3rd Party
Step 2.
Sanctioning
Sanctioning of violator for norm violation

(a) Female confederate enforces norm

(b) Third party enforces norm

Fig. S2. Manipulation of treatment dimension 2: Norm enforcement

Pre-analysis plan. We filed a pre-analysis plan (PAP) for this paper with Evidence in Governance and Politics (ID 20180725AB at www.egap.org) on July 30, 2018. The date of filing *preceded* the commencement of data collection for the project other than the pilot test runs (rehearsals), which were conducted to acquaint the enumerators with the procedure and details of the intervention. None of the pilot test run data are used for the purpose of the analysis. We note that in line with the registered preanalysis plan, one additional treatment dimension (linguistic assimilation) was manipulated as a part of the experimental intervention. However, since the focus of this paper is on the effect of civic norms on discriminatory behavior, and we face length constraints in this manuscript, we omit discussion of the results on the additional treatment dimension and reserve them for another publication (in progress).

Outcomes. We are interested in measuring the level of assistance offered to the female confederate who drops her possessions (bag of oranges) in the intervention, as specified in our pre-analysis plan. Enumerators observing each iteration of the intervention collected the following information regarding the reaction of bystanders. *This information was collected at the level of the iteration, which constitutes our unit of analysis.*

- *bystander*: Total number of bystanders within a 3 meter radius of where the iteration is taking place (count)
- *bystander_fem*: Total number of female bystanders within the 3 meter radius (count)
- *bystanderHP*: Total number of bystanders with headphones or earphones (count)
- *help*: Whether any bystander offered assistance to the female confederate (dichotomous)
- *help_count*: The number of bystanders who offered assistance (count)
- *help_count_fem*: The number of female bystanders who offered assistance (count)

Using this information, we construct one main outcome and additional auxiliary outcomes that will be used for the empirical analyses. These outcomes are calculated at the iteration level.

- *help*: Did *any* bystander offer assistance by moving to pick up possessions that the confederate has dropped? (**main**)
- *pcthelp*: The *proportion* of bystanders who offered assistance by moving to pick up possessions that the confederate has dropped (**auxiliary**)
- *womenhelp*: Did *any female* bystander offer assistance? (**auxiliary**)
- *menhelp*: Did *any male* bystander offer assistance? (**auxiliary**)

Data was collected for additional treatments in this manuscript, in accordance with our pre-analysis plan. In this paper, we analyze only the set of outcomes that focus on the effect of civic norms on discriminatory behavior. We reserve the other results for discussion in other publications.

2. Logistics and Procedures

Site selection. The interventions were conducted at train stations across 31 medium to large-sized cities/towns in the German states of North Rhine-Westphalia (NRW), Brandenburg, and Saxony. These states were not chosen at random; rather, we arrived at the decision to conduct these interventions in the three states after carefully weighing a combination of state and region-level sociodemographic factors that we believed would be of interest. The most obvious difference between North Rhine-Westphalia and the two other states (Brandenburg, and Saxony) is that they fell under West and East Germany prior to reunification. In addition, these two areas have been traditionally been exposed to very different levels of immigration in Germany's post war history. Whereas NRW is considered one of the most ethnically diverse federal states, with the highest proportion of foreign born populations in the country, the two other states have remained relatively ethnically homogeneous. Furthermore, the recent refugee crisis rising as result of the protracted conflict in the Middle East has also had a differential impact on the three states. The Königstein quota system, which combines state level tax revenues and population to assign asylum seekers, has naturally resulted in a high influx of refugees into NRW, which also happens to be one of the most populous and affluent states in Germany, and a low influx of refugees to Brandenburg and Saxony, which are sparsely populated and lag behind western German states in terms of tax revenue. But perhaps most importantly, there is ample reason to suggest that the level of racial resentment might vary significantly across the west (NRW) and the east (Saxony, Brandenburg); the level of electoral support for the far-right Alternative für Deutschland (AfD), which primarily campaigned on an anti-immigration agenda, in state and federal elections has been markedly higher in the East in comparison to the west. In some parts of Saxony, the AfD managed to secure the party vote share.

The list of cities and the number of train platforms (in parentheses) at each of the train stations where data collection was implemented is presented below.

- **North Rhine-Westphalia**: Münster (9), Bielefeld (8), Minden (5), Rheine (6), Köln (11), Köln Messe/Deutz (12), Mönchengladbach (9), Neuss (8), Siegen (6), Bonn (5), Düsseldorf (20), Wuppertal (5), Dortmund (31), Duisburg (12), Bochum (8), Gelsenkirchen (6), Hagen (16), Essen (13), Wanne-Eickel (8)
- **Saxony**: Leipzig (21), Görlitz (6), Chemnitz (14), Dresden (16), Zwickau (8)
- **Brandenburg**: Potsdam (7), Forst-Lausitz (5), Cottbus (10), Frankfurt-Oder (12), Brandenburg (6)

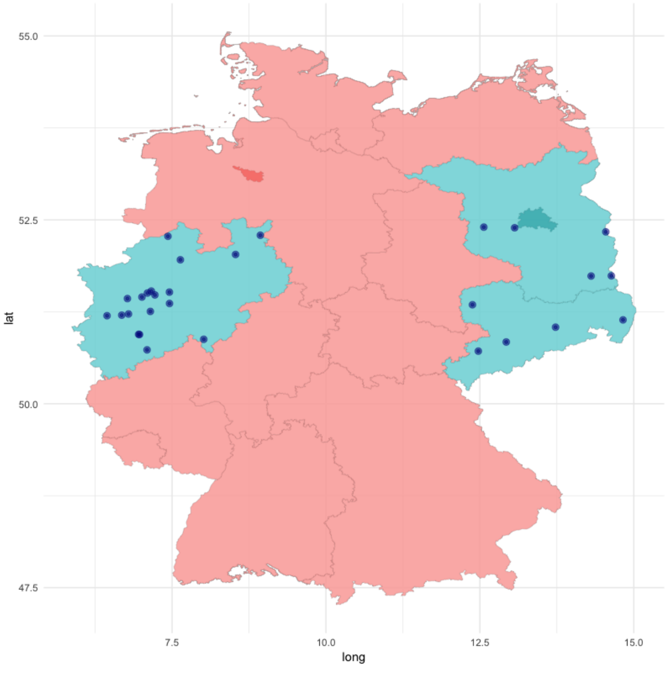


Fig. S3. Study sites – 29 train stations across 3 states in North Rhine-Westphalia, Saxony, and Brandenburg

Team constitution. We formed a total of seven confederate teams (three in North Rhine-Westphalia, two each in Saxony and Brandenburg), with four to five confederates constituting each team (total of 34 confederates). In order to make sure that we could cover all the roles required to implement the 14 different treatment conditions, we made sure that each team had at least one white German male confederate (playing the violator), at least one female confederate of a immigrant minority background (playing the female confederate), at least one white German female confederate (playing the control condition to the immigrant female confederate). When recruiting the confederate to play the immigrant punisher, we took extra care to hire women who were clearly identifiable by Germans as a member of the immigrant minority community based on their skin tone and phenotype; we deliberately excluded those with the possibility of being mistaken as a native German. Of the seven women recruited to play these roles, three were Turkish, two Egyptian, one Syrian, and one Kurdish in origin. We also made sure that the confederate playing the violator were clearly identifiable as a white German male. In addition to filling these key roles, we also hired at least one or two additional confederates who could play the role of third-party enforcer as well as serve as the outcome coders. In order to mitigate ethnicity-independent characteristics of the actors to influence bystanders' behavior, we decided to use a rather large number of actors of similar age and with similar attire for each confederate role. The make-up of each team with key roles highlighted are listed below.

- **NRW 1** (5): Gesika (immigrant female), Tobias (norm violator), Martina, Helena, Merlyn
- **NRW 2** (5): Bahar (immigrant female), Mirko (norm violator), Maria, Rudolph, Fulden
- **NRW 3** (5): Nilay (immigrant female), Tassilo (norm violator), Stefanie, Florence, Emine
- **Saxony 1** (4): Zeynep (immigrant female), Franz (norm violator), Juliane, Luzie
- **Saxony 2** (5): Mirna (immigrant female), Timon (norm violator), Sarah, Charlotte, Vatan
- **Brandenburg 1** (5): Emel (immigrant female), Moritz (norm violator), Damla, Louise, Koray
- **Brandenburg 2** (5): Yasmin (immigrant female), David (norm violator), Helin, Judith, Kitty

Training. Before the beginning of the intervention in each state, the confederates and enumerators that would observe and code the behavior of the bystanders participated in day-long training workshops led by the authors to ensure a consistently high quality in the delivery of the intervention. These trainings focused on how to select the settings for the intervention, how to play the different roles, how to ensure consistent performances across actors and across teams, and how to code bystander behavior consistently. For the main outcome of the study, whether a bystander provided assistance, enumerators were instructed to code any attempt to offer help in picking up oranges that consisted of a clear physical movement towards the oranges in an effort to help as provision of help, i.e. a clear movement to signal willingness to provide help in picking up oranges was necessary. In order to ensure consistent coding across enumerators and teams, different scenarios were discussed through role-playing activities during the training sessions. These training workshops were followed by extensive test runs in actual train stations with the authors. During the actual data collection, two enumerators independently observed the bystanders from different positions in an effort to minimize measurement error.

We took numerous precautions and trained the confederates and enumerators extensively in procedures to select the sites for the iterations in a way that minimizes the potential for bystanders to witness more than one iteration. First, the specific sites on each train platform were chosen such that it was hard to see the interaction from other platforms (e.g., by making use of walls and signs on the platform, timing the interaction such that stationary trains would block the sight). Second, platforms and the specific sites on those platforms were selected to minimize the chance of repeated participation by the same bystanders. After concluding one iteration on one platform, teams would switch to the platform farthest away from this one that had passengers waiting on it (only train stations with at least four tracks were used). Furthermore, the specific site on that new platform would be chosen to maximize the distance from the previous iteration (e.g., by going to the other end/side). Third, the enumerators tasked with observing the bystanders and coding their behavior were trained to make note of the bystanders for each iteration in order to avoid that—despite the other precautions—bystanders might witness more than one iteration (e.g., if passengers had stayed around after the departure of the train from that platform or had switched platforms). In the very limited instances where the same team conducted interventions at the same train station on more than one day, we conducted field work on different days of the week, choosing a business day and a weekend day in order to minimize chances of commuters being exposed to more than one iteration. Furthermore, enumerators were instructed to begin on the opposite track/side of the train station that during the prior day.

A note on enumerator "blinding" as to the purpose of the project. It was not possible to blind confederates to the general purpose of the experiment. All the coders were intelligent students who were interested in learning about research, thus after a few iterations the coders would have figured out that we were collecting data on bystander behavior across the different treatment conditions. However, we took steps to reduce the risk that coding reflected demand effects and confederates who acted out parts of the scene were expressly told to follow the script and to avoid behaviors that might be designed to elicit specific responses from the bystanders. We did not share the PAP with the actors or coders so they did not know what our prior expectations were for this experiment. They were given a script to follow during the intervention, were given detailed instructions on how to act (e.g. they were told to be polite albeit firm when enforcing the norm; to speak in a normal voice; and not to appear aggressive), and monitored during the iterations. Furthermore, most iterations were coded separately by two enumerators. Finally, there was no normative content in the material we used for the training of confederates (e.g. we referred to measuring assistance to confederates, rather than measuring discrimination and did not use loaded terms such as "bias" or "racism").

Ethical and safety considerations. We took great care to minimize the potential risk to study participants. For a full discussion of these measures, see the research protocol that was reviewed and approved by University of Pennsylvania's Institutional Review Board (IRB Protocol #829824). Beyond our efforts to minimize potential risks to subjects participating in the study, we also took a number of steps to ensure the safety of our research assistants (confederates and enumerators) during the study. Prior to the onset of data collection, we consulted a number of German experts on how to minimize potential risks to our RAs, esp. the norm violating confederates and the norm enforcing confederates. For example, we decided to pick only female confederates for the role of norm enforcer in order to minimize the risk of a physical conflict between bystanders and the confederate. Furthermore, the other confederates and the enumerators within each team closely monitored the bystanders and stood by, ready to intervene, if necessary. During the training sessions, we discussed potential risks and safety strategies extensively with the research assistants. RAs were instructed to stop the intervention if they felt unsafe at any point. The authors were in constant contact with all teams during the data collection, monitoring their progress and potential safety issues early-on. Last, the German train company, Deutsche Bahn, was instructed about research activities taking place at any given train station on any given day.

3. Bystander Composition and Scene Characteristics

In this subsection, we present descriptive statistics and additional information on the composition of the bystanders and other iteration characteristics. A minimum of 3 bystanders were required for each iteration. As discussed above, treatment assignment was orthogonal to all bystander characteristics. Therefore, we should not expect these characteristics to affect the results. To further demonstrate this empirically that, for example, the number of bystanders does not systemically affect the results, we also report specification that have number of bystander fixed effects, where the proportion outcome is used in the analysis. The estimates are virtually the same as without the fixed effects. We also include the full set of bystander composition and scene characteristics in our regression based analyses reported in Table S5 and S6. As expected, the inclusion of these additional covariates also do not change our original findings.

Table S1. Bystander composition and scene characteristics

Statistic	N	Mean	St. Dev.
Number of bystanders	1,614	4.428	1.449
Proportion of female bystanders	1,614	0.542	0.258
Proportion of bystanders w/ headphones	1,614	0.071	0.130
Hour of iteration	1,614	12.887	2.753
Iteration during rush hour (binary)	1,614	0.170	0.376
Temperature during iteration	1,614	29.053	3.708

Unfortunately, we were not able to collect information about bystander immigration status or ethnicity, given the already elaborate design. We do not think that poses a problem for our inferences. If bias is driven by ethnic or religious differences, as previous literature suggests, then the larger number of immigrant bystanders, the smaller the degree of discrimination that we should find. It follows that we could view our estimates as lower bounds of the true extent of native-immigrant discrimination, which would have been higher if all bystanders were native. Furthermore, the research teams were instructed to avoid bystander groups that were speaking in a foreign language or were clearly perceived as immigrants. These instructions were uniformly applied across all treatment conditions, and therefore have no reason to believe that there are systematic differences in the composition of the bystander pool in terms of their ethnicity or immigration status.

4. Covariate Balance

In this subsection, we present covariate balance statistics for our experimental treatment conditions. While covariate imbalance can arise due to chance, the randomization seems to have successfully obtained balance on each of the 6 pretreatment covariates we collected, both in the full sample as well as the samples disaggregated by state. Figures S2 and S3 present balance statistics for all statistical tests included in Figures 3 and 4 of the main text. Figure S4 presents the balance statistics for the hijab and native comparison by federal state. We include this balance table because we include analysis in the Supplementary Information regarding the hijab and native comparison in particular, disaggregated by state and region.

Table S2. Covariate balance for comparisons in Figure 3

	Mean Treated	Mean Control	T test p-value	KS test p-value
Native vs. immigrant with cross: column (1) vs (2)				
Number of bystanders	4.4301075	4.4625850	0.7807861	0.7854
Proportion of female bystanders	0.5431084	0.5293897	0.4679242	0.3018
Proportion of bystanders w/ headphones	0.0571796	0.0736300	0.0981795	0.1814
Hour of iteration	12.8064516	12.9551020	0.5227605	0.1472
Iteration during rush hour (binary)	0.1751152	0.1571429	0.5579957	-
Temperature during iteration	28.8234255	28.9428571	0.7041512	0.3412
Joint F-statistic: 0.6241 (p-value = 0.7111)				
Immigrant with cross vs. immigrant control: column (2) vs (3)				
Number of bystanders	4.4625850	4.3244980	0.1626622	0.0844
Proportion of female bystanders	0.5293897	0.5600671	0.0921786	0.2592
Proportion of bystanders w/ headphones	0.0736300	0.0698276	0.6593346	0.8978
Hour of iteration	12.9551020	12.9686747	0.9404746	0.8550
Iteration during rush hour (binary)	0.1571429	0.1855422	0.2603880	-
Temperature during iteration	28.9428571	28.9612490	0.9384094	0.9248
Joint F-statistic: 1.042 (p-value = 0.3965)				
Immigrant with cross vs. immigrant with hijab: column (2) vs (4)				
Number of bystanders	4.4625850	4.4243318	0.6961367	0.6902
Proportion of female bystanders	0.5293897	0.5398469	0.5211604	0.6566
Proportion of bystanders w/ headphones	0.0736300	0.0757804	0.8083094	0.9794
Hour of iteration	12.9551020	12.7075472	0.1767195	0.0404
Iteration during rush hour (binary)	0.1571429	0.1650943	0.7448717	-
Temperature during iteration	28.9428571	28.8490566	0.6984514	0.9398
Joint F-statistic: 0.4641 (p-value = 0.8352)				
Immigrant control vs. immigrant with hijab: column (3) vs (4)				
Number of bystanders	4.3244980	4.4243318	0.3001348	0.5102
Proportion of female bystanders	0.5600671	0.5398469	0.2872883	0.8126
Proportion of bystanders w/ headphones	0.0698276	0.0757804	0.5115495	0.8872
Hour of iteration	12.9686747	12.7075472	0.1692260	0.2560
Iteration during rush hour (binary)	0.1855422	0.1650943	0.4368234	-
Temperature during iteration	28.9612490	28.8490566	0.6511535	0.9642
Joint F-statistic: 0.8374 (p-value = 0.5411)				
Native vs. immigrant with hijab: column (1) vs (4)				
Number of bystanders	4.4301075	4.4243318	0.9597913	0.6530
Proportion of female bystanders	0.5431084	0.5398469	0.8682467	0.7282
Proportion of bystanders w/ headphones	0.0571796	0.0757804	0.0719134	0.1730
Hour of iteration	12.8064516	12.7075472	0.6789056	0.4798
Iteration during rush hour (binary)	0.1751152	0.1650943	0.7508337	-
Temperature during iteration	28.8234255	28.8490566	0.9365924	0.4436
Joint F-statistic: 0.5481 (p-value = 0.7716)				

Table S3. Covariate balance for comparisons in Figure 4

	Mean Treated	Mean Control	T test p-value	KS test p-value
Native enforcer vs. native non-enforcer: column (1) vs (2)				
Number of bystanders	4.4466667	4.4159544	0.8723780	0.1078
Proportion of female bystanders	0.5343120	0.5506267	0.6030468	0.2386
Proportion of bystanders w/ headphones	0.0472388	0.0656761	0.2346577	0.4222
Hour of iteration	12.7100000	12.8888889	0.6523117	0.8228
Iteration during rush hour (binary)	0.1500000	0.1965812	0.3664944	-
Temperature during iteration	29.1793333	28.5192308	0.2186675	0.1334
Joint F-statistic: 0.9079 (p-value = 0.4901)				
Native non-enforcer vs. immigrant with hijab enforcer: column (2) vs (3)				
Number of bystanders	4.4159544	4.4802956	0.6446874	0.5332
Proportion of female bystanders	0.5506267	0.5633615	0.6368757	0.2036
Proportion of bystanders w/ headphones	0.0656761	0.0860165	0.1947927	0.0446
Hour of iteration	12.8888889	12.7931034	0.7667732	0.3188
Iteration during rush hour (binary)	0.1965812	0.1477833	0.2745965	-
Temperature during iteration	28.5192308	28.8801314	0.4276426	0.2440
Joint F-statistic: 0.7331 (p-value = 0.6232)				
Immigrant with hijab enforcer vs. Immigrant with hijab non-enforcer: column (3) vs (4)				
Number of bystanders	4.4802956	4.3729261	0.4248084	0.1332
Proportion of female bystanders	0.5633615	0.5182475	0.0642700	0.1822
Proportion of bystanders w/ headphones	0.0860165	0.0663781	0.1352747	0.1144
Hour of iteration	12.7931034	12.6289593	0.5427395	0.4346
Iteration during rush hour (binary)	0.1477833	0.1809955	0.3570004	-
Temperature during iteration	28.8801314	28.8205128	0.8674564	0.9682
Joint F-statistic: 1.325 (p-value = 0.2446)				

Table S4. Covariate balance for hijab vs native comparison, by state

	Mean Treated	Mean Control	T test p-value	KS test p-value
Immigrant hijab vs native, North-Rhine Westfalia:				
Number of bystanders	4.6979167	4.5361635	0.3336713	0.3062
Proportion of female bystanders	0.5181347	0.5468327	0.3051481	0.8136
Proportion of bystanders w/ headphones	0.1018057	0.0717577	0.0496565	0.0880
Hour of iteration	12.9776786	13.0471698	0.8402304	0.7416
Iteration during rush hour (binary)	0.1741071	0.2075472	0.4778560	-
Temperature during iteration	28.6406994	28.8407233	0.6771188	0.4688
Joint F-statistic: 1.02 (p-value = 0.4124)				
Immigrant hijab vs native, Saxony:				
Number of bystanders	4.3011551	4.5087719	0.3579043	0.1188
Proportion of female bystanders	0.5718711	0.5777436	0.8724649	0.4322
Proportion of bystanders w/ headphones	0.0370442	0.0307018	0.6709330	0.8744
Hour of iteration	12.6336634	12.5964912	0.9402578	0.6668
Iteration during rush hour (binary)	0.1881188	0.1578947	0.6293403	-
Temperature during iteration	29.6580858	29.3877193	0.5790738	0.7930
Joint F-statistic: 0.2955 (p-value = 0.9383)				
Immigrant hijab vs native, Brandenburg:				
Number of bystanders	3.9309764	4.1388889	0.3185076	0.5840
Proportion of female bystanders	0.5563023	0.4992384	0.1731315	0.3290
Proportion of bystanders w/ headphones	0.0564137	0.0565122	0.9965654	0.8542
Hour of iteration	12.1717172	12.5555556	0.3835939	0.8012
Iteration during rush hour (binary)	0.1212121	0.1296296	0.8822794	-
Temperature during iteration	28.4951178	28.1938272	0.6742378	0.5794
Joint F-statistic: 0.7176 (p-value = 0.636)				

5. Regression-based Presentation of Treatment Effects

Table S5. Hijab versus native comparisons 1: Discrimination is consistently observed using both a binary measure of help and the share of bystanders offering help

	Hijab versus native									
	Any help?					% of bystanders helped?				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Hijab (vs. Native)	-0.120^{***}	-0.123^{***}	-0.119^{***}	-0.118^{***}	-0.124^{***}	-0.065^{***}	-0.067^{***}	-0.064^{***}	-0.068^{***}	-0.066^{***}
	(0.036)	(0.035)	(0.035)	(0.036)	(0.038)	(0.020)	(0.020)	(0.019)	(0.019)	(0.020)
Constant	0.783^{***}					0.316^{***}				
	(0.027)					(0.016)				
State FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
Team FE	No	No	Yes	No	No	No	No	Yes	No	No
Bystander FE	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Other Controls	No	No	No	No	Yes	No	No	No	No	Yes
Observations	666	666	666	666	641	666	666	666	666	641
R^2	0.015	0.029	0.072	0.058	0.066	0.016	0.027	0.084	0.110	0.115
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$									

Comparisons between immigrant hijab condition and native condition, pooling across norm enforcement dimension. Outcomes examined are our dichotomous measure of whether any bystander helped (Columns (1)–(5)) and the percentage of bystanders who helped (Columns (6)–(10)). Columns (1) and (6) report the average treatment effect (ATE) without any controls, while columns (2) and (7) report the ATE with state fixed effects. Columns (3) and (8) report the ATE with team fixed effects. Columns (4) and (9) report the ATE with both state and number of bystanders fixed effects. Columns (5) and (10) report the ATE with state and number of bystander fixed effects, as well as the full set of pretreatment controls (proportion of female bystanders, proportion of bystanders with headphones, hour of day, rush hour dummy, temperature at time of iteration). Constant terms for columns (1) and (6)—the baseline specifications—are the means for the control group (native category). Robust standard errors are reported in parentheses.

Table S6. Hijab versus native comparisons 1, clustered standard errors

	Hijab versus native									
	Any help?					% of bystanders helped?				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Hijab (vs. Native)	-0.120^{***}	-0.123^{***}	-0.119^{***}	-0.118^{***}	-0.124^{***}	-0.065^{**}	-0.067^{**}	-0.064^{**}	-0.068^{**}	-0.066^{**}
	(0.043)	(0.043)	(0.042)	(0.043)	(0.045)	(0.030)	(0.029)	(0.027)	(0.030)	(0.031)
Constant	0.783^{***}					0.316^{***}				
	(0.027)					(0.022)				
State FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
Team FE	No	No	Yes	No	No	No	No	Yes	No	No
Bystander FE	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Other Controls	No	No	No	No	Yes	No	No	No	No	Yes
Observations	666	666	666	666	641	666	666	666	666	641
R^2	0.015	0.029	0.072	0.058	0.066	0.016	0.027	0.084	0.110	0.115
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$									

Table S6 replicates Table S5 with *robust standard errors clustered at the study site level (train station).*

Table S7. Hijab versus native comparison, by region: Discrimination is larger in former East Germany

	Hijab versus native					
	Any help?		% of bystanders helped?			
	(1)	(2)	(3)	(4)	(5)	(6)
Hijab (vs. Native)	−0.162***	−0.087*	−0.082***	−0.093***	−0.052*	−0.045*
	(0.053)	(0.047)	(0.029)	(0.029)	(0.027)	(0.027)
Constant	0.759***	0.807***	0.302***		0.330***	
	(0.041)	(0.037)	(0.024)		(0.022)	
Region	East	West	East	East	West	West
Bystander FE	No	No	No	Yes	No	Yes
Observations	313	353	313	313	353	353
R^2	0.027	0.009	0.026	0.109	0.010	0.106
Note:						*p<0.1; **p<0.05; ***p<0.01

Comparisons between immigrant hijab condition and native condition, pooling across norm enforcement dimension, but disaggregated by region (Former East Germany and West Germany). Outcomes examined are 1) our dichotomous measure of whether any bystander helped and 2) the percentage of bystanders who helped. Columns (1) and (2) report the average treatment effect (ATE) on our dichotomous main outcome, while columns (3) – (6) report the ATE using the percentage of bystanders who helped. Columns (4) and (6) report specifications with number of bystanders fixed effects. Constant terms for columns (1), (2), (3), and (5)—the baseline specifications—are the means for the control group (native category). Robust standard errors are reported in parentheses.

Table S8. Hijab versus native comparison, by region, clustered standard errors

	Hijab versus native					
	Any help?		% of bystanders helped?			
	(1)	(2)	(3)	(4)	(5)	(6)
Hijab (vs. Native)	−0.162***	−0.087	−0.082***	−0.093***	−0.052	−0.045
	(0.049)	(0.070)	(0.030)	(0.033)	(0.050)	(0.050)
Constant	0.759***	0.807***	0.302***		0.330***	
	(0.030)	(0.045)	(0.022)		(0.038)	
Region	East	West	East	East	West	West
Bystander FE	No	No	No	Yes	No	Yes
Observations	313	353	313	313	353	353
R^2	0.027	0.009	0.026	0.109	0.010	0.106
Note:						*p<0.1; **p<0.05; ***p<0.01

Table S8 replicates Table S7 with *robust standard errors clustered at the study site level (train station).*

Table S9. Hijab versus native comparison, by state: Discrimination is largest in the state of Saxony

	Hijab versus native								
	Any help?			% of bystanders helped?					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Hijab(vs. Native)	−0.087*	−0.217***	−0.105	−0.052*	−0.045*	−0.119***	−0.148***	−0.044	−0.049
	(0.047)	(0.070)	(0.080)	(0.027)	(0.027)	(0.041)	(0.041)	(0.040)	(0.044)
Constant	0.807***	0.825***	0.691***	0.330***		0.337***		0.266***	
	(0.037)	(0.051)	(0.063)	(0.022)		(0.034)		(0.033)	
State	NRW	Sachsen	Bburg	NRW	NRW	Sachsen	Sachsen	Bburg	Bburg
Bystander FE	No	No	No	No	Yes	No	Yes	No	Yes
Observations	353	159	154	353	353	159	159	154	154
R^2	0.009	0.050	0.011	0.010	0.106	0.054	0.174	0.008	0.084
Note:									*p<0.1; **p<0.05; ***p<0.01

Comparisons between immigrant hijab condition and native condition, pooling across norm enforcement dimension, but disaggregated by federal state (North Rhine-Westphalia, Brandenburg, and Saxony). Outcomes examined are 1) our dichotomous measure of whether any bystander helped, and 2) the percentage of bystanders who helped. Columns (1)–(3) report the average treatment effect (ATE) on our dichotomous main outcome, while columns (4)–(9) report the ATE using the percentage of bystanders who helped. Constant terms for columns (1), (2), (3), (4), (6), and (8)—the baseline specifications—are the means for the control group (native category). Robust standard errors are reported in parentheses.

Table S10. Hijab versus native comparison, by state, clustered standard errors

	Hijab versus native								
	Any help?			% of bystanders helped?					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Hijab(vs. Native)	−0.087	−0.217***	−0.105	−0.052	−0.045	−0.119***	−0.148***	−0.044	−0.049
	(0.070)	(0.054)	(0.070)	(0.050)	(0.050)	(0.039)	(0.042)	(0.034)	(0.035)
Constant	0.807***	0.825***	0.691***	0.330***		0.337***		0.266***	
	(0.045)	(0.031)	(0.034)	(0.038)		(0.028)		(0.017)	
State	NRW	Sachsen	Bburg	NRW	NRW	Sachsen	Sachsen	Bburg	Bburg
Bystander FE	No	No	No	No	Yes	No	Yes	No	Yes
Observations	353	159	154	353	353	159	159	154	154
R^2	0.009	0.050	0.011	0.010	0.106	0.054	0.174	0.008	0.084
Note:									*p<0.1; **p<0.05; ***p<0.01

Table S10 replicates Table S9 with *robust standard errors clustered at the study site level (train station).*

Table S11. Immigrant (hijab + control) versus native comparisons

	Immigrants (hijab + control) versus native				
	Any help?		% of bystanders helped?		
	(1)	(2)	(3)	(4)	(5)
Immigrants (vs. Natives)	−0.070**	−0.070**	−0.050***	−0.051***	−0.053***
	(0.031)	(0.031)	(0.018)	(0.018)	(0.018)
Constant	0.783***		0.316***		
	(0.027)		(0.016)		
State FE	No	Yes	No	Yes	Yes
Bystander FE	No	No	No	No	Yes
Observations	1,098	1,098	1,098	1,098	1,098
R^2	0.004	0.018	0.008	0.019	0.092

Note: *p<0.1; **p<0.05; ***p<0.01

Comparisons between immigrant hijab and immigrant control conditions versus native condition, pooling across norm enforcement dimension. Outcomes examined are 1) our dichotomous measure of whether any bystander helped (our main outcome), and 2) the percentage of bystanders who helped. Columns (1) and (2) report the average treatment effect (ATE) on our dichotomous main outcome, while columns (3)–(5) report the ATE using the percentage of bystanders who helped. Columns (1) and (3) report the average treatment effect (ATE) without state fixed effects, while columns (2) and (4) report the ATE with state fixed effects. Column (5) includes state fixed effects and number of bystanders fixed effects. Constant terms for columns (1) and (3)—the baseline specifications—are the means for the control group (native category). Robust standard errors are reported in parentheses.

Table S12. Immigrant (hijab + control) versus native comparisons, clustered standard errors

	Immigrants (hijab + control) versus native				
	Any help?		% of bystanders helped?		
	(1)	(2)	(3)	(4)	(5)
Immigrants (vs. Natives)	−0.070*	−0.070**	−0.050**	−0.051**	−0.053**
	(0.036)	(0.035)	(0.025)	(0.025)	(0.026)
Constant	0.783***		0.316***		
	(0.027)		(0.022)		
State FE	No	Yes	No	Yes	Yes
Bystander FE	No	No	No	No	Yes
Observations	1,098	1,098	1,098	1,098	1,098
R^2	0.004	0.018	0.008	0.019	0.092

Note: *p<0.1; **p<0.05; ***p<0.01

Table S12 replicates Table S11 with *robust standard errors clustered at the study site level (train station).*

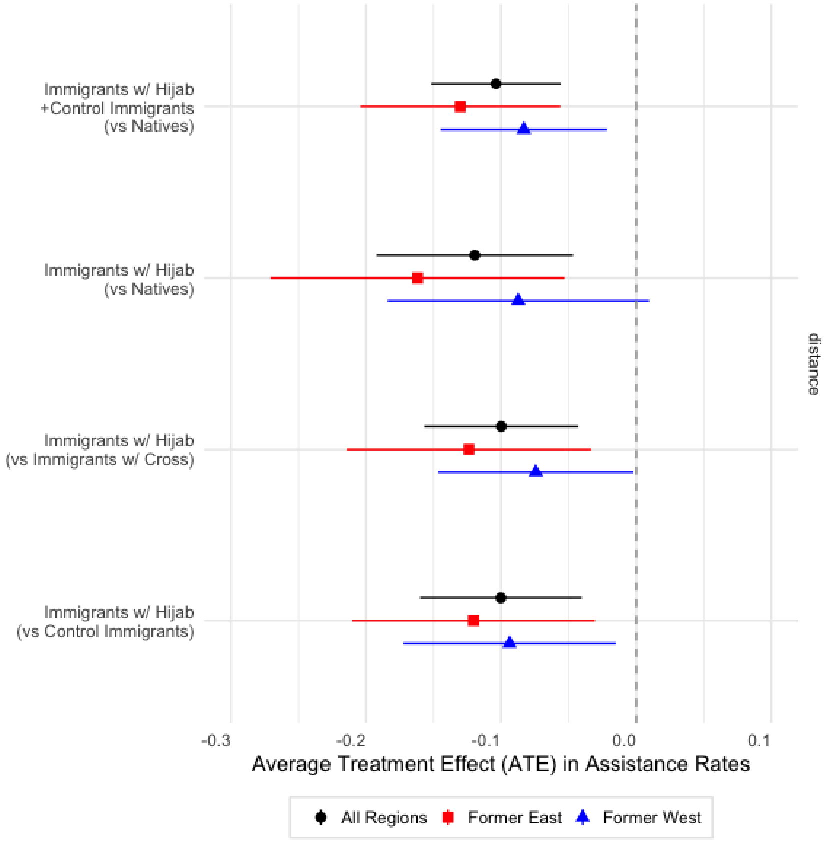


Fig. S4. ATEs for ascriptive differences

Figure S4 reports the average treatment effects (ATE) for ascriptive characteristics. The circle, square and triangle correspond to the point estimate of the ATE in the full sample and the iterations conducted in former East and West Germany respectively. The lines represent 95 percent confidence intervals for the point estimates. The vertical axis reports the treatment conditions compared.

Table S13. Norm enforcement effects among immigrants

	Norm enforcer vs non-enforcer				
	Any help?		% of bystanders helped?		
	(1)	(2)	(3)	(4)	(5)
Norm enforcer (vs. Non-enforcer)	0.052**	0.052**	0.023*	0.023*	0.027**
	(0.024)	(0.024)	(0.012)	(0.012)	(0.012)
Constant	0.707***		0.258***		
	(0.017)		(0.008)		
State FE	No	Yes	No	Yes	Yes
Bystander FE	No	No	No	No	Yes
Observations	1,388	1,388	1,388	1,388	1,388
R^2	0.003	0.015	0.003	0.014	0.078

Note: *p<0.1; **p<0.05; ***p<0.01

Comparison of the level of assistance offered to immigrants who enforce the anti-littering norm and immigrants who do not enforce the norm, pooling across ascriptive differences dimension. Outcomes examined are 1) our dichotomous measure of whether any bystander helped (our main outcome), and 2) the percentage of bystanders who helped. Columns (1) and (2) use the dichotomous measure as the outcome, whereas (3)–(5) use the percentage measure. Columns (1) and (3) are specifications without state fixed effects, while columns (2) and (4) are specifications with state fixed effects. Column (5) report specifications with both state and bystander fixed effects. Constant terms for columns (1) and (3)—the baseline specifications—are the means for the control group (non-enforcers). Robust standard errors are reported in parentheses.

Table S14. Norm enforcement effects among immigrants, clustered standard errors

	Norm enforcer vs non-enforcer				
	Any help?		% of bystanders helped?		
	(1)	(2)	(3)	(4)	(5)
Norm enforcer (vs. Non-enforcer)	0.052**	0.052**	0.023*	0.023*	0.027**
	(0.023)	(0.023)	(0.013)	(0.012)	(0.012)
Constant	0.707***		0.258***		
	(0.022)		(0.015)		
State FE	No	Yes	No	Yes	Yes
Bystander FE	No	No	No	No	Yes
Observations	1,388	1,388	1,388	1,388	1,388
R^2	0.003	0.015	0.003	0.014	0.078

Note: *p<0.1; **p<0.05; ***p<0.01

Table S14 replicates Table S13 with *robust standard errors clustered at the study site level (train station).*

Table S15. Norm enforcement effects by region

	Dependent variable			
	Any help?		% of bystanders helped?	
	(1)	(2)	(3)	(4)
Norm enforcer (vs. Non-enforcer)	0.080^{**}	0.028	0.049^{***}	0.001
	(0.037)	(0.030)	(0.017)	(0.017)
Constant	0.643^{***}	0.762^{***}	0.220^{***}	0.291^{***}
	(0.026)	(0.022)	(0.011)	(0.012)
Region	East	West	East	West
Observations	639	749	639	749
R^2	0.007	0.001	0.012	0.00000
Note:				*p<0.1; **p<0.05; ***p<0.01

Comparison of the level of assistance offered to immigrants who enforce the anti-littering norm and immigrants who do not enforce the norm, pooling across ascriptive differences dimension, disaggregated by region. Outcomes examined are 1) our dichotomous measure of whether any bystander helped (our main outcome), and 2) the percentage of bystanders who helped. Columns (1) and (2) report the average treatment effect (ATE) on our dichotomous main outcome, while columns (3) and (4) report the ATE using the percentage of bystanders who helped. Robust standard errors are reported in parentheses.

Table S16. Norm enforcement effects by region

	Dependent variable			
	Any help?		% of bystanders helped?	
	(1)	(2)	(3)	(4)
Norm enforcer (vs. Non-enforcer)	0.080^{*}	0.028	0.049^{***}	0.001
	(0.041)	(0.024)	(0.007)	(0.018)
Constant	0.643^{***}	0.762^{***}	0.220^{***}	0.291^{***}
	(0.021)	(0.025)	(0.005)	(0.022)
Region	East	West	East	West
Observations	639	749	639	749
R^2	0.007	0.001	0.012	0.00000
Note:				*p<0.1; **p<0.05; ***p<0.01

Table S16 replicates Table S15 with *robust standard errors clustered at the study site level (train station).*

Table S17. Language effects among immigrants

	Dependent variable			
	Any help?		% of bystanders helped?	
	(1)	(2)	(3)	(4)
German(vs Foreign Language)	−0.016	−0.011	0.004	0.006
	(0.024)	(0.024)	(0.012)	(0.012)
Constant	0.740***		0.267***	
	(0.017)		(0.008)	
State FE	No	Yes	No	Yes
Observations	1,388	1,388	1,388	1,388
R^2	0.0003	0.011	0.0001	0.011
Note:				*p<0.1; **p<0.05; ***p<0.01

In addition to the two main treatment dimensions—ascriptive characteristics and norm enforcement—our research design manipulated a third dimension—language used by the confederate in the phone call. This was based on the theoretical discussion presented by Hopkins (1), which argued that language would be a salient dimension through which ingroup outgroup differences are perceived. The intuition for the analysis presented in this Table S18 is to compare the level of assistance offered to immigrants who speak German during the phone call versus those that use a foreign language unintelligible to the host population. Columns (1) and (2) report the average treatment effect (ATE) on our dichotomous main outcome, while columns (3) and (4) report the ATE using the percentage of bystanders who helped. The findings reported in columns (1)-(4) suggest that linguistic assimilation has no discernible impact on how immigrants are treated by the host population. Robust standard errors are reported in parentheses.

Table S18. Language effects among immigrants

	Dependent variable			
	Any help?		% of bystanders helped?	
	(1)	(2)	(3)	(4)
German(vs Foreign Language)	−0.016	−0.011	0.004	0.006
	(0.018)	(0.018)	(0.010)	(0.010)
Constant	0.740***		0.267***	
	(0.022)		(0.012)	
State FE	No	Yes	No	Yes
Observations	1,388	1,388	1,388	1,388
R^2	0.0003	0.011	0.0001	0.011
Note:				*p<0.1; **p<0.05; ***p<0.01

Table S18 replicates Table S17 with *robust standard errors clustered at the study site level (train station).*

6. Additional manipulation checks

Manipulation checks regarding the perception of confederate ethnicity. In order to support our claim that discrimination against our immigrant confederates is driven by religious but *not* ethnoracial (phenotypical) differences, we must show that German host populations perceive our confederates to be of immigrant minority background (in the control condition when they are not wearing a hijab). We therefore conducted a new follow-up survey on Clickworker.com, an online crowdsourcing work platform similar to Amazon's M-Turk to recruit adult German respondents to evaluate our confederate's photos and report their perceived country of origin. We conducted this survey on a sample of 208 German adults above 19 years of age. Each evaluation question presented a photo of our confederate, and then asked "in your best guess, where do you think this person is from?" Respondents were then asked to choose from "German" versus four other countries (Turkey, Egypt, Iraq, and Syria), which were the real countries of origin for our immigrant confederates. All respondents evaluated a total of 15 confederate photographs (all seven of our immigrant confederates, and roughly 1/2 of the total German native confederates that participated in the intervention of the experiment). This yields a total of 3,120 evaluations across all photos.

Table S19. Proportion of respondents identifying confederate as a German native

	Native Confederates	Immigrant Confederates	Difference	P-Value
Experimental weights	82.97%	15.38%	67.59%p	< 0.001

It is clear that respondents are able to draw stark distinctions in the country of origin of our German native confederates versus immigrant confederates. On average, respondents correctly identify German native confederates as Germans between 82–83% of the time. In stark contrast, only 15–16% of respondents mistakenly categorize our immigrant minority confederates' country of origin as Germany. The difference is consistently in excess of 65% points, and is statistically distinguishable at p<0.001. These manipulation checks provide strong evidence that our immigrant confederates were sufficiently different in terms of their ethnic attributes (phenotype, skin tone) to German native confederates, and bystanders in our main experiment are highly likely to have perceived our immigrant control confederates as immigrants or Germans with an immigrant background. As with every survey, it is possible to consider different ways of presenting the survey questions. For example, a longer list of countries could have been provided to respondents to choose from; other countries (beyond Germany) with majority Christian population could have been included; or responses could have been left open-ended. Nonetheless, the evidence in this survey is so stark as to suggest that these slight modifications would not impact our conclusions from the manipulation checks.

7. Additional survey evidence on perceptions regarding the anti-littering norm in Germany

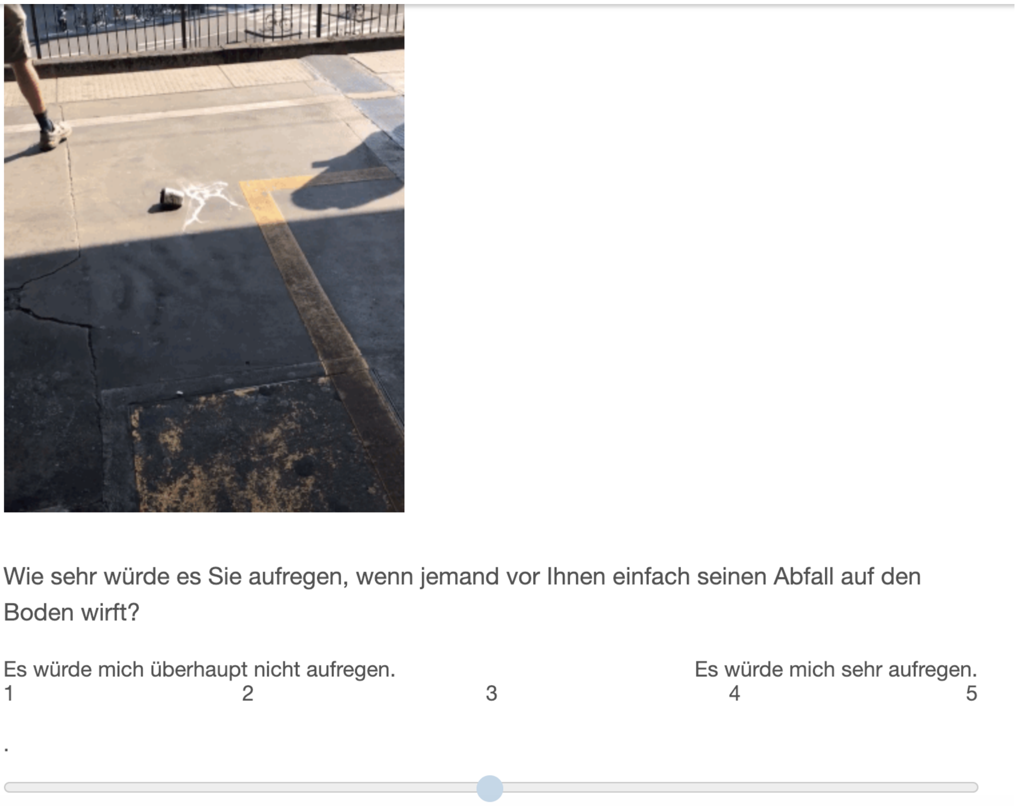
In this section, we present results from a survey that was conducted on a sample of 316 German respondents across Germany regarding their attitudes towards littering. Online samples have been used extensively in political science research in American Politics and other areas of the discipline. We used a stratified sample to ensure representation from the cities where the experiment was fielded. The survey is not intended to provide definitive results that are representative of public opinion in Germany. We could not identify an existing survey-based source on the question of interest, so we decided to pursue a triangulation strategy and conducted a media analysis using publicly available information (see results below) as well as a new online survey that we designed specifically to collect information on whether Germans care about the norm of non-littering (an uncontroversial assumption in our view).

The survey allows us to test the premise that Germans share strong norms against littering and that they believe that immigrants, especially those who are not culturally integrated in German society, would be more likely to litter than German natives. We provide suggestive evidence in support of these premises via a survey administered on an online sample recruited through Clickworker.com. The survey included a battery of questions designed to probe the *strength* of the norms against littering amongst German host populations, as well as their perceptions regarding which demographic groups are more likely to violate the norm.

Norms against littering in German populations are strongly held among German host populations. In order to establish that norms against littering are strongly held and shared by a broad majority of Germans, we presented a short three second video clip of a person throwing litter on a train platform. We followed by asking two questions to the respondents regarding their reactions to the video clip. First, we asked the respondents to evaluate the extent to which they would find it upsetting if they saw someone littering in a public space. Respondents were asked to respond on a five point scale, ranging from 1 ("it would not upset me at all"), and 5 ("it would upset me very much"). Samples of the screen presented to respondents are shown below in Figure S5.

Responses to this survey item demonstrate that norms against littering are widely held. On a five point Likert scale (1-5), 86% of responses were either 4 or 5, meaning that Germans find violations of the anti-littering norm to be highly upsetting. A mere 0.6% responded that they do not find littering to be upsetting at all.

We followed this question with a survey item that asked what actions respondents would take when confronted with a situation in which they observed someone littering in a public space. The options presented included "I would tell the person to pick up the trash", "I would pick up the trash myself", "I would see how other people near me respond and would point it out to them, where appropriate", "I would call the police", and "I would not care." As presented in the fourth bar (row) in



Wie sehr würde es Sie aufregen, wenn jemand vor Ihnen einfach seinen Abfall auf den Boden wirft?

Es würde mich überhaupt nicht aufregen. Es würde mich sehr aufregen.

1 2 3 4 5

.

Fig. S5. Screen capture of survey item on how much littering would be upsetting

Figure S6, of the 316 respondents, only 4.7% said that they "would not care." This means that 95.3% of all respondents replied that they would take some for of action to sanction and correct the norm violation.

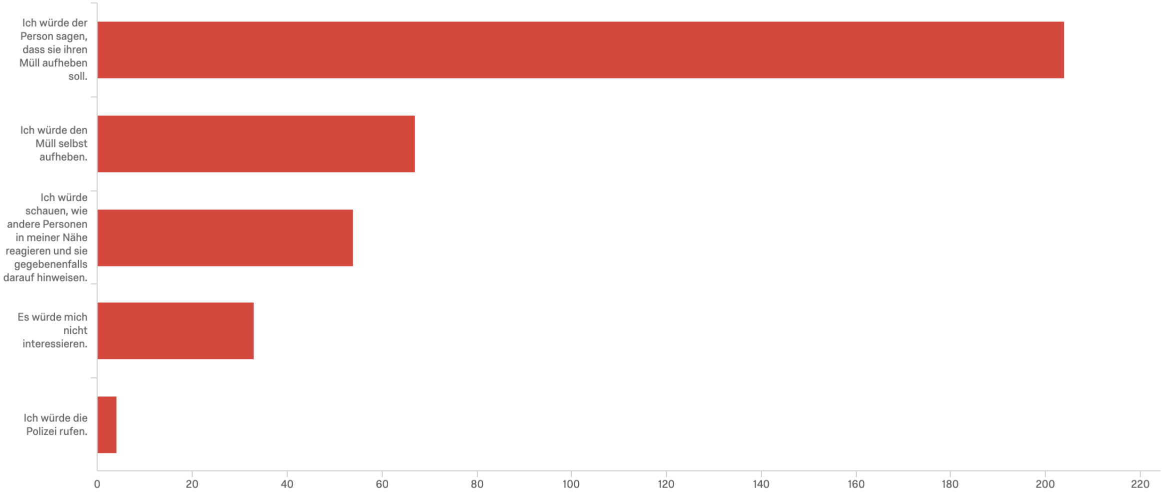


Fig. S6. Responses to "what would you do in a situation in which someone litters?"

Germans expect immigrants and foreigners to litter more than Germans. In addition to the items to probe the strength of the anti-littering norm, we also included items in the survey aimed at understanding whether German host populations expect immigrant minorities to be less respectful of the norm, and hence litter more frequently than native Germans. Specifically, we presented respondents with a photo of a littered street, and asked "In many German cities, people simply discard waste (such as coffee mugs, empty bottles, or packaging material) onto the street. Who do you think does this most often, Germans or immigrants and refugees?" We phrased the question item in a direct manner, fully acknowledging the possibility of social desirability bias to work against respondents answering “immigrants and refugees.”

Table S20. "Germans versus immigrants/refugees litter more"

	“Immigrants and refugees litter more”	“Germans litter more”	Difference	P-Value
Experimental weights	61.99%	38.01%	23.98%p	0.0011

Responses to this item are presented in Table S20. In calculating the means of responses, we apply the same approach we used for the manipulation checks and use weights based on the distribution of the observations in our main experimental sample, although the results remain substantively unchanged without the weights. Despite the concern that social desirability would bias against respondents' choosing the “immigrants and refugees” answer, 62% of respondents said that immigrants are more likely to litter than Germans. This means that only 38% of respondents said that Germans are more likely to litter than immigrants. This difference is statistically significant at the $P<0.01$ level. Given that social desirability bias is likely to work against there being a difference, we see this differential to be a lower bound.

This expectation that immigrants and foreigners litter more than Germans is also often expressed by politicians in the public discourse. In fact, newspapers regularly cover complaints about immigrants littering in public spaces. The mayor of Duisburg Sören Link, for example, claims that the increase in immigration in recent years has led neighbors to feel “strongly bothered by piles of garbage, noise, and rat infestation”[1].

In a similar vein, the prominent former Senator for Finance for Berlin, Thilo Sarrazin, claims that “the [city's] cleaning department clears up 20 tons of mutton leftovers from the Tiergarten [park] every Monday left by the Turkish community” [2]. Such perceptions are shared by politicians across the political spectrum: even politicians from the progressive Green party, such as the former Berlin state assembly member Claudia Hämmerling, who concludes that “this is how people behave who have never fully arrived here.”[3]

The crucial importance of complying with the anti-littering norm for the integration of immigrants is a common theme in the rhetoric of German politicians. For example, the former mayor of Neuköln, the Berlin borough with the highest concentration of immigrants, Heinz Buschkowsky claims: “A man with Turkish background does not have to prove his willingness to integrate by wearing lederhosen, drinking beer only by the liter or eating weisswurst for breakfast. Accepting the principles of our constitution as elements for his life and the life of his family is enough. ... [It is enough,] if he sends his children to school and if he carries his trash to the trashcan instead of throwing it from the balcony.”[4]

While such positions are expressed by politicians from all major parties, they are particularly common on the far right. The president of the far-right NPD party in North Rhine-Westphalia, Claus Cremer, for example, provocatively asks, “What do you say to such “cultural enrichers,” [immigrants “enriching the German culture”] who first need to be taught not to poop on other people's properties and to throw garbage in trash cans and not simply on the street?”[5]. The same party warns residents in Berlin (in the Rudow neighborhood) that, if asylum seeker accommodations are to open in their neighborhood, they will have to prepare for “being long-term neighbors with asylum seekers, with all the negative side effect, such as frequent noise, litter, and criminality.”[6] Similarly, AfD politician Matthias Niebel goes as far as saying that proper handling of trash “belongs ... to the core area of good German culture.”[7].

Why Germans expect immigrants to litter more than Germans. As a follow up to the previous survey item, we asked respondents who said that immigrants are more likely to litter than Germans to provide an open-ended justification for their answer. We present a collection of these comments, after translation into English, through a wordcloud in Figure S7.

Respondents most frequently cited the “lack of norms or rules regarding littering in the home country” of the immigrants as the reason why immigrants are likely to litter more than Germans. For example, one respondent explicitly mentioned that “there are no rules on waste disposal in their homelands”. Another respondent claimed that immigrants and refugees “may come from a country where the rules (against littering) are less strict. All in all, out of a total of 100 meaningful recorded responses, 22 invoked the differences in home country norms and rules, with some respondents invoking a “lack of culture” against littering in immigrant home countries.” Including the number of respondents who claimed that immigrants litter more than Germans because of their “habit,” this number increases to 30. A relatively substantial number of respondents attributed their expectations to what they perceived as a “lack of respect among immigrants for Germany and German traditions.” There were a total of 11 responses that invoked the term “respect”, making up the second largest category of responses.

[1] “Rasanter Anstieg beim Kindergeld alarmiert Städte”, T-Online, August 10, 2018

[2] “Sarrazin ist nah dran und doch daneben”, Tagesspiegel, Oct. 8, 2009

[3] “Die Affäre Hammelbein”, Zeit, August 20, 2009

[4] “Die bittere Wahrheit über unsere Schulen”, Bild, September 19, 2012

[5] “Kapitulationserklärung: Polizisten aus Rumänien und Bulgarien sollen in NRW für Ordnung sorgen”, NPD Bochum, October 22, 2013

[6] “Ein Asylbewerberheim in Rudow? Nicht mit uns!”, NPD Neukölln, October 16, 2012

[7] “Presseerklärung Müllentsorgung tägliche PHV. Stadtrat Matthias Niebel wundert sich”, Alternative-heidelberg.de, November 25, 2015

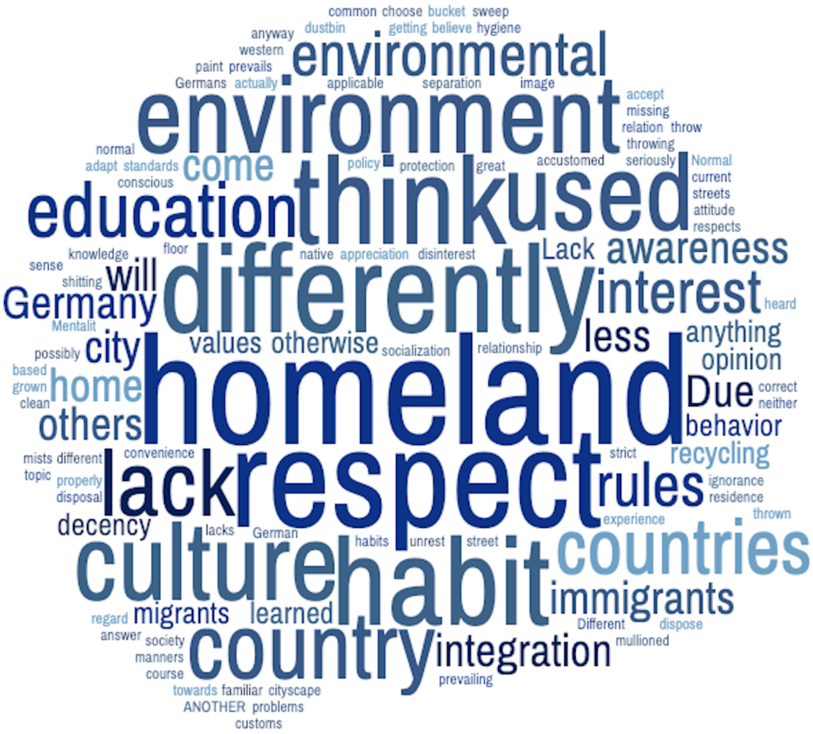


Fig. S7. Wordcloud of open-ended justifications for why respondents believe immigrants litter more than Germans

References

1. Hopkins DJ (2014) The upside of accents: Language, inter-group difference, and attitudes toward immigration. *British Journal of Political Science* 45:531–557.