



# Ethnic diversity and cooperation: evidence from a lost letter experiment

Nan Zhang & Alexandra Kommol

To cite this article: Nan Zhang & Alexandra Kommol (28 Mar 2024): Ethnic diversity and cooperation: evidence from a lost letter experiment, Journal of Ethnic and Migration Studies, DOI: [10.1080/1369183X.2024.2334909](https://doi.org/10.1080/1369183X.2024.2334909)

To link to this article: <https://doi.org/10.1080/1369183X.2024.2334909>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 28 Mar 2024.



Submit your article to this journal [↗](#)



Article views: 75



View related articles [↗](#)



View Crossmark data [↗](#)

## Ethnic diversity and cooperation: evidence from a lost letter experiment

Nan Zhang and Alexandra Kommol

Mannheim Center for European Social Research, Mannheim, Germany

### ABSTRACT

Are residents of ethnically diverse communities less likely to encounter acts of cooperation and kindness from strangers? Although a vast literature has arisen examining the relationship between ethnic diversity and popular attitudes such as social trust, we know little about how prosocial *behaviors* vary across neighborhoods. This study presents evidence from a ‘lost letter’ experiment conducted across 77 neighborhoods in 13 German cities. We experimentally vary the ethnic identity of letter senders and recipients, and also carefully select our experimental field sites to create ‘matched sets’ of neighborhoods that differ only with respect to the percentage of foreign residents. We find no difference in return rates across our ethnicity treatments, and no overall relationship between return rates and neighborhood diversity. However, comparing *across* ‘matched sets,’ we do detect significantly lower return rates in areas featuring higher unemployment. Taken together, our findings support the view that it is not diversity *per se*, but rather associated socioeconomic deprivation, which is most detrimental to cooperative neighborhood interactions.

### ARTICLE HISTORY

Received 9 October 2023  
Accepted 17 March 2024

### KEYWORDS

Ethnic diversity; social cohesion; discrimination; field experiment; Germany

Continuing immigration to Western societies has raised critical questions surrounding the impact of ethnic diversity on social cohesion (Putnam 2007). Reviewing a vast accumulated literature, recent meta-analyses point to an overall negative relationship between residential ethnic diversity and social trust and related sentiments (Dinesen, Schaeffer, and Sønderskov 2020; Schaeffer 2016). Prior research, however, has relied almost exclusively upon *attitudinal* measures of social cohesion as captured in opinion surveys. By contrast, only a handful of studies have examined the relationship between residential ethnic diversity and prosocial *behaviors* at the neighborhood level, with mixed results (Baldassarri 2020; Holland, Silva, and Mace 2012; Koopmans and Veit 2014a; Sampson 2012; Volker et al. 2016)<sup>1</sup>.

To address this gap, we present evidence from a ‘lost letter’ experiment conducted across 77 neighborhoods in 13 German cities. In our experiment, roughly 1150 sealed, stamped, and pre-addressed envelopes were dispersed on public sidewalks in residential areas for passersby to find. The rate at which such ‘lost’ letters are returned provides an

**CONTACT** Nan Zhang  n.zhang@uni-mannheim.de

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

unobtrusive measure of prosocial behavior at the local level (Baldassarri 2020; Holland, Silva, and Mace 2012; Koopmans and Veit 2014a; Sampson 2012). We further experimentally vary both (i) the ethnicity (German vs. Turkish) of the senders and recipients as well as (ii) the type of correspondence (a personal vs. an official letter). This design thus allows us to study prosociality separately for ethnic majority and minority beneficiaries and examine the robustness of our findings across different kinds of letters.

A key innovation of our study concerns the procedure for selecting field sites. Specifically, we sought to account for the fact that diverse neighborhoods also tend to be poor neighborhoods (Abascal and Baldassarri 2015; Dancygier 2018; Dinesen, Schaeffer, and Sønderskov 2020; Gereke, Schaub, and Baldassarri 2018), and that, for a variety of reasons, poorer individuals may be less prosocial, trusting, or civically engaged (Andreoni, Nikiforakis, and Stoop 2021; Haushofer and Fehr 2014; Schaub 2021). A simple comparison of lost letter return rates across ethnically homogeneous and heterogeneous neighborhoods therefore risks confounding the impacts of poverty and diversity. Accordingly, we adopt a matching strategy wherein sets of neighborhoods are selected based on similar levels of unemployment (as well as population size, residential turnover, and other characteristics), but differing shares of non-German residents. Letter drops were conducted on the same day (and at similar times of day) within each ‘matched set’ of neighborhoods. Our procedures thus help to ensure that neighborhood diversity is statistically independent of other potential correlates of letter return rates, thereby allowing us to isolate the contextual ‘effect’ of ethnic composition<sup>2</sup>.

Our experiment yields three main findings. First, although personal cards were returned at higher rates than official letters, we uncover no evidence of ethnic discrimination for either type of correspondence. Secondly, return rates for all letters are unrelated to neighborhood diversity. Finally, comparing *across* ‘matched sets’ of neighborhoods, we do detect significantly lower return rates in areas featuring higher unemployment. Our evidence thus supports the view that diversity *per se* is not detrimental to social cohesion (Abascal and Baldassarri 2015; Kustov and Pardelli 2018; Letki 2008). Rather, our findings underscore the importance of addressing socioeconomic inequalities and structural disadvantages in fostering cooperative neighborhood interactions in ethnically diverse settings.

### **Theoretical Mechanisms: Diversity and Social Cohesion**

The literature has put forward several explanations linking ethnic diversity to prosocial behavior. A first general mechanism is rooted in *Social Identity Theory* (Tajfel 1982), and posits that individuals are psychologically predisposed to treat in-group members more positively than out-group members (Balliet, Wu, and Dreu 2014; Bernhard, Fischbacher, and Fehr 2006). Further, in modern societies, group boundaries are often constructed along ethnic or national lines. From this perspective, ethnic diversity can be expected to affect collective outcomes by either increasing or decreasing the number of in-group members to whom individuals are ‘parochially altruistic’<sup>3</sup>.

Another prominent perspective draws from theories of *Group Threat* (e.g. Blalock 1967), which hold that ethnic intermingling breeds conflict by fueling competition for real and symbolic resources<sup>4</sup>. To the extent that ethnic threats induce out-group hostility (Enos 2014) and in-group solidarity (Jardina 2019), prosocial behavior

is eroded across ethnic lines. Importantly, while ethnic threats are sometimes couched in ‘color-blind’ terms (i.e. without reference to specific in-groups and out-groups), *Group Threat* theories are most often invoked to explain the negative reactions of the dominant majority (i.e. ‘natives’) to the growth of minority populations<sup>5</sup>.

Both the *Social Identity* and *Group Threat* perspectives highlight the boundary between (‘native’) in-groups vs. (minority) out-group members. In contrast, alternative viewpoints focus on explaining prosocial behavior toward strangers in general (regardless of ethnicity). Popularized by Putnam’s seminal study of social capital, *Constrict Theory* argues that diversity induces community residents to ‘hunker down’, such that ‘trust even of one’s own race is lower, altruism and community cooperation rarer, friends fewer’ (Putnam 2007, 137). While Putnam himself is silent about why this process occurs, subsequent work has linked this phenomenon to feelings of anomie and social isolation which hamper trust and cooperation both within and across ethnic groups (Dinesen, Schaeffer, and Sønderskov 2020).

*Group Threat* and *Constrict* theories are both theories of *contextual effects* – that is, the behavior of individuals is posited to change as a function of the larger social context. In contrast, a relationship between diversity and collective outcomes may also arise due to *compositional* differences across neighborhoods. For example, in their re-analysis of Putnam (2007), Abascal and Baldassarri (2015) find that a large part of the negative association between diversity and trust across US communities can be attributed to the fact that (i) non-Whites report lower trust than Whites, and (ii) by definition, non-Whites are more likely to live in diverse communities. Similar *compositional effects* may be at work in the European context if ‘natives’ and immigrant display different baseline levels of prosociality.

Finally, an altogether different interpretation for the negative relationship between diversity and social cohesion focuses not so much on ethnicity *per se*, but rather on social disadvantage and other important correlates of diversity (Abascal and Baldassarri 2015; Gereke, Schaub, and Baldassarri 2018). Put simply, diverse communities also tend to be poor communities, and poverty itself – or more precisely, its absence – is an important driver of prosociality and civic participation (Andreoni, Nikiforakis, and Stoop 2021; Schaub 2021). Deprived communities may also be characterized by more frequent environmental and physical depredations which can ‘crowd out’ civic behaviors (Keizer, Lindenberg, and Steg 2008). Thus, under this view, socioeconomic inequalities do the ‘heavy lifting’ in explaining variation in collective outcomes between homogeneous and heterogeneous communities.

### **Attitudes vs. Behavior**

Methodologically, the vast majority of previous studies have sought to evaluate hypotheses derived from these perspectives using public opinion data on trust (e.g. ‘Do you believe that most people can be trusted?’) and related attitudes (Dinesen, Schaeffer, and Sønderskov 2020). However, such attitudes may only imperfectly capture the actual quality of social interactions within a locality. Indeed, prior studies have uncovered a poor correspondence between stated attitudes and actual behaviors (Pager and Quillian 2005; Winter and Zhang 2018)<sup>6</sup>.

This mismatch between attitudes and behavior can arise for a number of reasons. First, attitudinal responses may be shaped by social desirability concerns. For instance,

few individuals nowadays would readily admit to harboring ethnic prejudices, although such prejudices may influence everyday behavior, subtly or otherwise. The same logic may apply to answers to the trust question if, for example, residents in certain areas feel that they *ought* to say that their neighbors can be trusted.

A second issue relates to the distinction between ‘fast’ and ‘slow’ thinking (Kahneman 2011). Here, survey questions are often presumed to elicit consciously-held attitudes and beliefs which emerge from intentional deliberation. In contrast, many everyday behaviors may arise ‘spontaneously’ from implicit feelings and judgments that operate below the level of conscious awareness (Zhang, Gereke, and Baldassarri 2022).

Finally, to the extent that survey prompts emphasize abstract or hypothetical situations, they necessarily gloss over more specific interpersonal or situational factors which are relevant in decision-making. For instance, regardless of whether one (hypothetically) trusts ‘most people’ in the abstract, an altogether different set of considerations may apply in encounters with real flesh-and-blood individuals. Indeed, as Pager and Quillian (2005, 372) note, ‘it is difficult to anticipate how any individual, including oneself, may react to a situation previously encountered only in hypothetical terms’.

Taken together, these considerations call for more research on how prosocial *behaviors* vary across neighborhoods featuring higher or lower ethnic diversity. While not denying that survey responses ‘tell us something meaningful about the attitudes of respondents [in these contexts]...we have no way of anticipating the degree to which these expressed attitudes will be reflected in any particular set of behaviors’ (ibid).

### **The Lost Letter Experiment**

To address these limitations, a handful of studies have employed the ‘lost letter’ technique as a *behavioral* measure of social cohesion (Milgram, Mann, and Harter 1965). This method typically consists of dispersing sealed, stamped, and pre-addressed envelopes in public spaces. Passersby who find the letters can ignore, return, or even destroy them. Because returning a letter is costly in terms of time and effort, return rates can be interpreted as a measure of prosocial behavior towards an unknown sender and receiver.

Such experiments have two features suitable to our purposes. First, by comparing rates at which letters are returned from different neighborhoods, we can assess the relationship between ethnic diversity and cooperative local interactions. Secondly, by varying the ethnic identities of the senders and receivers, we can measure the degree to which prosocial behavior is shaped by group boundaries.

Prior studies have generally yielded mixed results with respect to both of these dimensions. First, while some research finds a negative correlation between return rates and neighborhood diversity (Koopmans and Veit 2014a; Volker et al. 2016), other work has yielded null results (Baldassarri 2020; Sampson 2012). Secondly, some studies have produced evidence of ethnic discrimination in return rates (Ahmed 2010; Hellmann et al. 2015), while others show that the use of ethnic names makes little difference (Baldassarri 2020; Koopmans and Veit 2014a; Volker et al. 2016).

We present results from a lost letter study<sup>7</sup> in which approximately 1150 letters were dropped across 77 neighborhoods in 13 German cities<sup>8</sup>. Approximately half of our letters took the form of a personal card announcing the birth of a child (see Figure 1). The card



**Figure 1.** Materials employed in the lost letter experiment.

itself was placed within an envelope displaying sender's and receiver's addresses in a handwritten script<sup>9</sup>. To study whether ethnic discrimination contributes to lower cooperation in ethnically diverse areas, we experimentally manipulate the identity of the names displayed on the envelopes. Specifically, the cards were either addressed from the Weber family to M. Fischer (common German names), or from the Öztürk family to Y. Aktürk (common Turkish names)<sup>10</sup>. In all cases, the sender's address always indicated a non-existent street number located within the same city where the letter was ostensibly lost. The receiver's address (belonging to one of the authors) was identical across all letters<sup>11</sup>.

Beyond manipulating names, we also wished to assess the robustness of our results to different ‘types’ of letters. Accordingly, we implemented an additional set of treatments to examine ethnic discrimination in civic behavior. More specifically, we dispersed two additional letters designed to mimic official correspondence sent by a representative of a fictitious civic organization<sup>12</sup>. These letters were placed within envelopes carrying the logo of the organization and containing a plastic window displaying the receiver’s address. Half of these letters were addressed from the Citizens’ Initiative for Net Neutrality to Michael Fischer, while the other half were addressed from the organization Pro-Migration to Cem Aktürk<sup>13</sup>. The receiver’s address remained unchanged.

Research assistants (RAs) were instructed to drop the letters face-up on public sidewalks in residential neighborhoods<sup>14</sup>. Working from Google Maps and OpenStreetMap, we chose a set of 16 drop locations within each neighborhood<sup>15</sup>. We attempted to identify locations spaced at least 250 meters apart to minimize the chances that the same individual could encounter multiple letters. Additionally, letters were dropped in two waves spaced several days apart to further avoid crowding<sup>16</sup>. Finally, we sought to avoid choosing drop locations next to schools, public transport stops, and post boxes. With our list of drop locations in hand, we randomly shuffled the order and then assigned one of our four letters to each location. The assigned letters and their corresponding drop locations were printed on paper maps that we distributed to RAs.

Letter drops took place on rain-free days between June 2022 and June 2023. We instructed the RAs to drop the letters as close as possible to their assigned locations, but to avoid litter, trash bins, etc. Each letter contained a unique 4-digit code that allowed us to track where the letter was dropped. On official letters, this code was indicated in small font beneath the receiver’s address. On personal cards, this code was hidden in the birth weight. RAs recorded the code next to the street address where they dropped the letter, thus allowing us to match returned letters to neighborhoods<sup>17</sup>.

### Selecting Neighborhoods

In choosing our field sites, we sought to select residential neighborhoods that differed with respect to ethnic diversity, but which were otherwise similar along other dimensions that may be correlated with letter return rates<sup>18</sup>. To that end, we began by identifying a set of TARGET NEIGHBORHOODS in a metropolitan area where the percentage of non-German residents exceeded a pre-defined threshold (e.g. 20%)<sup>19</sup>. For each TARGET, our algorithm then searched for MATCHED NEIGHBORHOODS within the same metropolitan area using the following criteria:

- (i) the percentage of foreign residents should be at least 5pp *lower* than in the TARGET NEIGHBORHOOD
- (ii) the neighborhood population is within  $+/- 50\%$  of the TARGET’s population, and the absolute population difference cannot exceed 5000 residents
- (iii) the difference in unemployment rates (averaged over the past 3 years) cannot exceed  $+/- 1$  pp, and neither can the gap between German and non-German unemployment rates
- (iv) the difference in the percentage of long-term residents (living at least 5 years in the same residence) cannot exceed  $+/- 5$ pp

- (v) the difference in diversification trends (measured as the 10-year change in the percentage of foreign residents) cannot exceed  $\pm 2$  pp
- (vi) the difference in the population share of any specific national grouping<sup>20</sup> within the overall non-German population cannot exceed  $\pm 5$  pp

We refer to each TARGET NEIGHBORHOOD and the less-diverse neighborhoods matched to it as a MATCHED SET.

The matching procedure was conducted twice: once with TARGET NEIGHBORHOODS defined according to the 20% non-German residents threshold, yielding 53 neighborhoods grouped into 18 MATCHED SETS; and again according to a 30% threshold, yielding 27 neighborhoods across 11 MATCHED SETS<sup>21</sup>. Appendix Section 2 provides evidence on the success of this matching procedure in producing statistical independence between neighborhood diversity and the matching variables. Further, letters were dropped on the same day, and at roughly the same time of day, in all neighborhoods within the same MATCHED SET<sup>22</sup>. This ensured that factors such as weather conditions and proportions of certain ‘types’ of pedestrians (e.g. weekend shoppers, workday commuters, etc.) are constant across neighborhoods, and thus also statistically independent of diversity levels.

### Measuring Diversity

We utilize a simple measure of diversity operationalized as the percentage of foreign residents within a neighborhood. That said, we do not imply that the *concept* of diversity should be reduced to a simple dichotomy between ‘natives’ and ‘non-natives’. Rather, we acknowledge that diversity also encapsulates the idea of *heterogeneity* in terms of the variety of groups from different ethnic and cultural origins. Further, we recognize that our choice of measurement (percentage foreigners) appears to privilege *Group Threat* mechanisms and differences in baseline prosociality between ‘natives’ and ‘non-natives’, while alternative operationalizations such as the Hirschman-Herfindahl Index (HHI) instead emphasize the concept of fractionalization which features centrally in, for example, the anomie channel underlying *Constrict Theory*.

Without gainsaying these important conceptual distinctions, we note however that, as a practical matter, many diversity measures are observationally equivalent (Abascal, Ganter, and Baldassarri 2023; Kustov and Pardelli 2018; Schaeffer 2013). Indeed, Schaeffer (2013) compared five different diversity measures – each operationalizing putatively distinct theoretical mechanisms – and found that ‘the competing indices are [all] indistinguishable even from the mere percentage of minorities’ (p. 756). Across our three metro regions, the Pearson correlation between the HHI and the percentage of foreign residents is a nearly perfect 0.99. Given this feature of the data, we do not devote too much attention to adjudicating between various diversity measures. Instead, we utilize the percentage of foreign residents for its simplicity, and rely upon our treatments to derive testable implications of the various theoretical considerations described above (see next section).

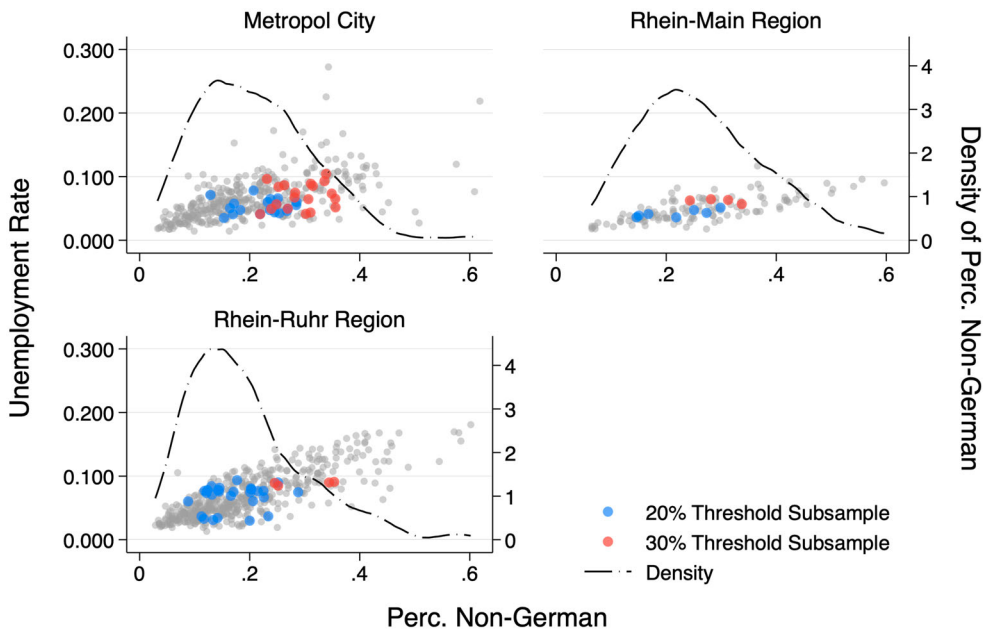
Critical readers may also object that our matching criteria are overly-restrictive in terms of the types of diverse neighborhoods we consider. For instance, important differences may exist between two neighborhoods with the same percentage of foreign



residents, but where the minority population either (a) is split between many national groups, or (b) mainly comprises a single ethnic ‘enclave’. Yet, such a comparison is eliminated by our sixth matching criterion. Similarly, some scholars have argued that it is not so much the size of the minority population which matters, but rather the *speed* of diversification (Hopkins 2010; Newman 2013), but this distinction is held constant in our fifth matching criterion.

Although we agree that these are important features of diverse neighborhoods that deserve further examination, as a practical matter, such neighborhoods are rarely found within our data (see also Schönwälder and Söhn 2009). For instance, we actually did conduct an alternative matching exercise where we tried to pair ethnic ‘enclaves’ with similarly-diverse areas containing a more balanced ‘mixture’ of ethnic minority groups. The results yielded only a handful of neighborhoods and proved insufficient for our purposes (since we did not want to ‘spam’ a large number of lost letters in the same geographic area). Given these constraints, we instead focused our matching efforts on *reducing* variance along these dimensions in order to more carefully isolate variation in diversity levels, broadly conceived.

Finally, readers may ask whether our matching procedure produces sufficient variation in ethnic diversity between matched neighborhoods. Figure 2 plots all (matched and un-matched) neighborhoods in our three metro areas by the percentage of foreign residents and the unemployment rate. The figure also shows how our 77 sample neighborhoods are situated within this full range. We can observe that our sample neighborhoods cover quite a lot of the distribution in terms of neighborhood diversity. We would therefore argue that our study is capturing the ‘normal’ difference in diversity levels between neighborhoods in these regions of Germany<sup>23</sup>.



**Figure 2.** Unemployment rates and shares of non-German residents across sample and non-sample neighborhoods in three metropol regions.

## Testable Implications

Given the parameters of our study design and the features of the underlying neighborhood data, what testable implications can we derive with respect to the theoretical mechanisms linking diversity to collective outcomes? We begin with *Constrict Theory*, which posits that diversity reduces generalized prosocial behavior – i.e. regardless of individuals' own ethnicity, as well as any in-group vs. out-group considerations. This leads to the simple prediction:

H1: There is a negative association between lost letter return rates and the percentage of foreign residents in the neighborhood.

Note that this prediction is also consistent with a compositional story wherein minorities demonstrate a lower baseline level of prosociality in comparison to natives (Abascal and Baldassarri 2015).

A different set of predictions obtain if we think about compositional effects from the perspective of *Social Identity Theory*. Specifically, 'parochial altruism' holds that 'natives' are more likely to return letters bearing German names (Weber/Fischer and NeuWeb) and, similarly, Turkish residents should be more likely to return letters bearing Turkish names. Further, it is important to note that, even the most diverse areas, Germans still constitute the absolute majority (if measured by nationality) or close to it (if measured by migration background) (see Appendix Table A1). In contrast, the percentage of Turkish residents in our sample neighborhoods never exceeds 11% by nationality<sup>24</sup>. Given these population shares (majority or near-majority German, minority Turkish), we can assume that, in comparison to German letters, Turkish letters are less likely to be encountered by an in-group member across all sample neighborhoods. This leads to the prediction:

H2: Letters bearing Turkish names will be returned at lower rates than letters bearing German names.

At the same time, the likelihood that a Turkish letter will be encountered by an in-group member increases with neighborhood diversity (Koopmans and Veit 2014b), while the corresponding 'encounter probability' for German letters decreases. As a result:

H3: The 'gap' in return rates between German and Turkish letters should narrow as the share of foreign residents in the neighborhood increases.

A third set of predictions can be derived from *Group Threat Theory*. Importantly, here we follow the predominant tendency in the literature and apply the *Threat* perspective to the behavior of 'natives' alone. The basic logic is that 'natives' are more threatened by large non-native populations. On the one hand, if threat manifests as discrimination, then 'natives' living in diverse neighborhoods should be less likely to return Turkish letters (and / or more likely to return German letters). On the other hand, there are fewer 'natives' (i.e. potential 'discriminators') living in diverse areas. Consequently, the relationship between neighborhood diversity and the aforementioned 'gap' in return rates will depend on the balance of these two dynamics. To the extent that threat-induced discrimination outweighs the compositional effect:

H4: The 'gap' in return rates between German and Turkish letters should grow as neighborhood diversity rises.

Otherwise, a predominant compositional effect yields a prediction consistent with H3.

We readily acknowledge that our testable implications can only partially distinguish between the various theoretical mechanisms discussed above. This is because we cannot observe the identity of the *letter finders*, and must therefore draw inferences based on differences in *aggregate* return rates<sup>25</sup>. At the same time, if we find *no support* for H1–H4, this would suggest that diversity is unrelated to collective outcomes through any of the theoretical perspectives outlined above. Indeed, this is precisely the position of scholars who argue that diversity functions largely as a proxy for socio-economic deprivation (Abascal and Baldassarri 2015; Abascal, Ganter, and Baldassarri 2023; Letki 2008). Thus our final prediction:

H5: There is a negative association between lost letter return rates and the percentage of unemployed residents in the neighborhood.

## Results

### Differences across Letter Types

We begin by examining differences in return rates depending on the ethnicity of the senders and receivers (H2). Such differences would be consistent with arguments rooted in ‘parochial altruism.’ Figure 3 displays the return rates across letter types for the entire experiment ( $N=1168$ ). We observe that roughly 75% of personal cards (Weber and Öztürk) were returned, compared to approximately 54% of official letters (NeuWeb and ProMigration). Further, 66% of ‘majority’ letters (Weber and NeuWeb) were returned, compared to 62% of letters (Öztürk and ProMigration) addressed to ethnic minority recipients.

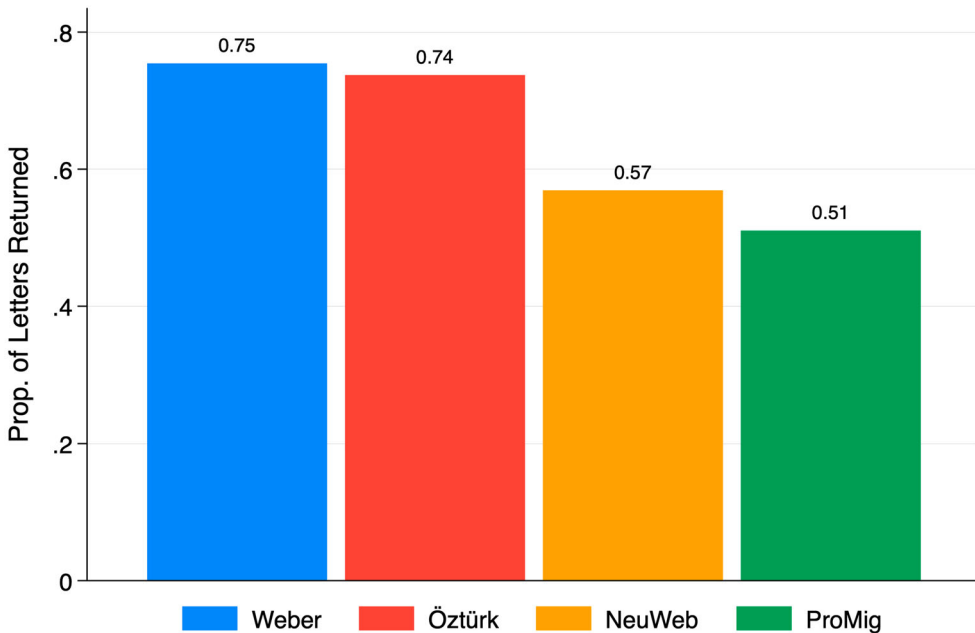


Figure 3. Aggregate return rates by letter treatment.

To assess the statistical significance of these patterns, we estimate the following linear probability model:

$$Y_{ic} = \beta_1 \ddot{O}ztu\ddot{r}k_{ic} + \beta_2 NeuWeb_{ic} + \beta_3 ProMigration_{ic} + Z_c \tag{1}$$

where  $Z$  represents a series of neighborhood-by-date fixed effects<sup>26</sup>. Our interest lies in the  $\beta$ s (and their linear combinations), which allow us to compare the return rates across letter treatments. Model (1) is estimated for the full sample, as well as separately for the 20%-threshold and 30%-threshold subsamples. Results are presented in columns (1), (2), and (4) of Table 1. Panel A compares the return rates for the Weber and Öztürk letters, while Panel B compares the NeuWeb and ProMigration treatments. Across all specifications, we can see that letters addressed to minorities were indeed less likely to be returned, although the differences are not statistically significant at even the  $p < 0.1$  level.

Beyond these main (non-)effects, we also assess whether the ‘gap’ in return rates between Turkish and German letters may change in more diverse neighborhoods (H3 and H4). Accordingly, we estimate the following interaction model:

$$\begin{aligned} Y_{ic} = & \delta_1 \ddot{O}ztu\ddot{r}k_{ic} + \delta_2 NeuWeb_{ic} + \delta_3 ProMigration_{ic} + Z_c \\ & + \delta_4 \ddot{O}ztu\ddot{r}k_{ic} \times Perc\_NonGerman_c \\ & + \delta_5 NeuWeb_{ic} \times Perc\_NonGerman_c \\ & + \delta_6 ProMigration_{ic} \times Perc\_NonGerman_c \end{aligned} \tag{2}$$

Our main interest is in  $\delta_4$ ,  $\delta_5$ , and  $\delta_6$  (and their linear combinations), which allow us to test whether neighborhood diversity moderates differences in return rates between treatments. Equation (2) is again estimated separately for each subsample, thus ensuring that variation in diversity is not proxying for other neighborhood differences. Results are presented in columns (3) and (5) of Table 1. The interaction coefficients are substantively small and statistically insignificant, indicating that the (non-)difference in return rates of letters addressed to ‘native’ majority vs. ethnic minority recipients is stable across neighborhoods<sup>27</sup>.

Overall, our results provide scant evidence that prosocial behavior is driven by ‘parochial altruism’ or *Group Threat*. Letters addressed to minorities are no less likely to be

**Table 1.** Regression results: differences in return rates across letter types.

	Full Sample	20% Threshold		30% Threshold	
	(1)	(2)	(3)	(4)	(5)
Panel A: reference group = Weber					
Öztürk	-0.017 (0.035)	-0.002 (0.044)	0.001 (0.042)	-0.047 (0.051)	-0.052 (0.051)
Öztürk × Perc. Non-German			0.004 (0.007)		-0.008 (0.011)
Panel B: reference group = NeuWeb					
ProMig	-0.057 (0.036)	-0.054 (0.043)	-0.055 (0.043)	-0.067 (0.073)	-0.069 (0.068)
ProMig × Perc. Non-German			-0.002 (0.007)		-0.002 (0.019)
N letters	1168	819	819	397	397

Note: Robust standard errors in parentheses. No differences are statistically significant.

returned than letters addressed to the ‘native’ majority, and these patterns do not differ across neighborhoods at varying levels of diversity.

### Differences across Neighborhoods

Although we find no differences in the return rates of letters addressed to majority vs. minority recipients, it may still be the case that the aggregate return rate declines as diversity increases (H1). Indeed, both *Constrict Theory* and some arguments about compositional effects predict precisely such a pattern. To examine this possibility, we estimate an OLS model regressing the average return rate of all letters in the neighborhood on the percentage of non-German residents<sup>28</sup>. Again, we conduct this analysis separately for the subsamples spanning the 20% non-German and 30% non-German thresholds. As shown by the coefficients presented in columns (1) and (3) of [Table 2](#), we find no relationship between residential ethnic diversity and the proportion of returned letters.

In columns (2) and (4) of [Table 2](#), we add neighborhood unemployment rates and the percentage of long-term residents (as a proxy for local residential turnover) as additional predictors in the model. Prior research has identified both factors as key explanations for lower levels of prosociality in ethnically diverse, disadvantaged neighborhoods (Sampson 2012; Volker et al. 2016). We additionally control for neighborhood population, as the chances that a passerby encounters a letter is higher in more populous neighborhoods. Importantly, although our protocols ensure that there is not much variation on these variables *within* MATCHED GROUPS, leverage for this analysis comes from variation *across* MATCHED GROUPS.

In support of H5, we find that unemployment rates are indeed strong, negative, and statistically significant predictors of lost letter return rates. Our estimates imply that a 1 pp increase in unemployment is associated with a 3pp – 4pp decrease in the proportion of letters returned. The patterns are remarkably similar across both subsamples (see [Figure 4](#)). Going from the least deprived to the most deprived neighborhood is associated with a decline in return rates from almost 80% to around 50%.

In summary, our lost letter experiment involving carefully-matched neighborhoods across 13 German cities uncovered neither (i) reduced prosociality towards ethnic minority beneficiaries, nor (ii) lower return rates in more ethnically-diverse neighborhoods,

**Table 2.** Regression results: predictors of neighborhood-level return rates.

	20% Threshold		30% Threshold	
	(1)	(2)	(3)	(4)
Perc. Non-German	0.000 (0.004)	-0.000 (0.004)	-0.005 (0.007)	-0.002 (0.006)
Unemployment Rate		-0.038*** (0.010)		-0.035* (0.014)
Perc. Longterm Residents		0.000 (0.003)		0.004 (0.006)
Total Population (in hundreds)		-0.000 (0.000)		0.000 (0.001)
Constant	0.658*** (0.088)	0.903** (0.281)	0.728*** (0.192)	0.623 (0.432)
N neighborhoods	53	53	27	27

Note: Robust standard errors in parentheses.<sup>+</sup> $p < 0.10$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

nor (iii) an interaction between (i) and (ii). Instead, we find that fewer letters are returned from neighborhoods with higher shares of unemployed residents. Taken together, these results underscore the importance of neighborhood disadvantage in shaping collective outcomes, while suggesting that – net of such disadvantages – diversity *per se* is not detrimental to local cooperation.

### Discussion and Conclusions

A central question in the social sciences asks whether ethnic diversity harms the quality of neighborhood relations. Theory offers contrasting predictions of the consequences of diversity for prosocial behavior. While some perspectives suggest a negative relationship between diversity and social cohesion, other scholars attribute lower cooperation not to diversity *per se*, but rather to the associated socioeconomic disadvantages prevailing in ethnically-heterogeneous neighborhoods. Turning to the evidence, prior research has generally uncovered a negative relationship between neighborhood diversity and *attitudinal* measures of social trust. However, the handful of studies comparing prosocial *behaviors* between diverse vs. homogeneous neighborhoods have yielded mixed results.

We report findings from a large scale lost-letter experiment in 77 neighborhoods across 13 cities across different regions in Germany. Our experimental materials feature both German or Turkish names in order to examine whether return rates are shaped by the ethnicity of the letter recipient (and sender). We also use both personal and official letters to assess the robustness of our findings across different types of correspondence. Finally, to isolate the impact of neighborhood diversity, we employ a matching strategy to create ‘matched sets’ of neighborhoods that differ little (e.g. in terms of local unemployment rates) *except* in the proportion of foreign residents.

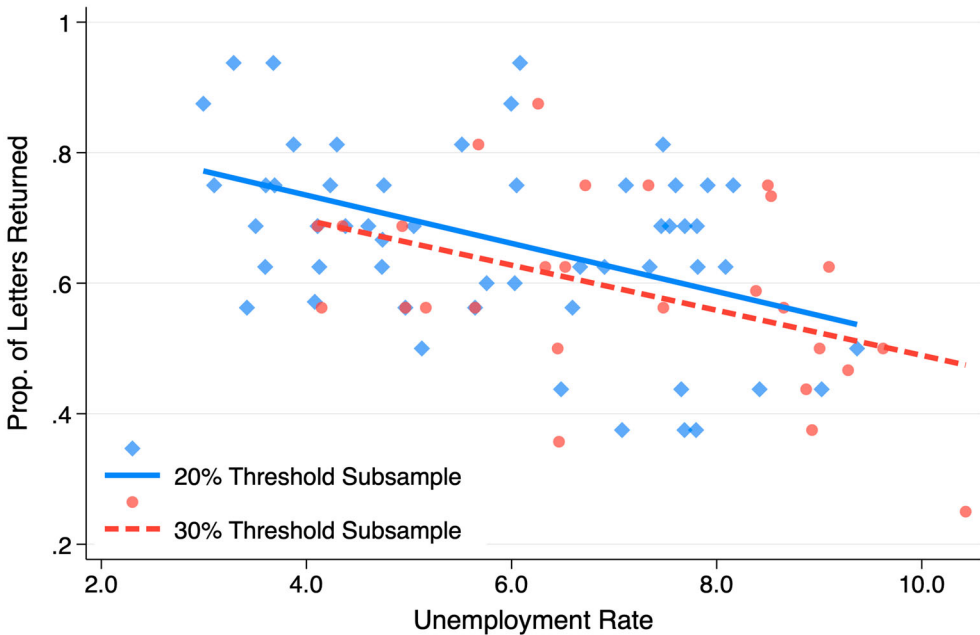


Figure 4. Overall return rates by neighborhood deprivation.

Overall, our results indicate that lost letter return rates are unrelated to neighborhood diversity, but strongly shaped by neighborhood deprivation (as measured in local unemployment rates). In terms of theory, our findings speak against broad generalizations about the negative effects of diversity on cooperation, and point instead to factors such as socioeconomic disadvantage in shaping collective outcomes.

How do our findings fit with prior research employing the lost letter technique? First, with regards to ethnic discrimination in return rates, while Hellmann et al. (2015) report that letters bearing minority names are returned less often, most other work fails to find a discriminatory effect (Baldassarri 2020; Koopmans and Veit 2014a; Volker et al. 2016)<sup>29</sup>. Our null results are in-line with this latter body of research.

Turning to the relationship between neighborhood characteristics and overall return rates, Baldassarri (2020) and Sampson (2012) find no ‘effect’ of ethnic diversity across Italian regions and Chicago neighborhoods. Our own findings also point in this direction. In contrast, Volker et al. (2016) document a significant correlation between return rates and the percentage of non-Western residents across Dutch neighborhoods, and Koopmans and Veit (2014a) also find a similar pattern in Berlin. It is worth noting, however, that the substantive size of the diversity ‘effect’ in Koopmans and Veit (2014a) is swamped by the influence of the local unemployment rate. More generally, while the relationship between diversity and return rates remains ambiguous, the detrimental impact of neighborhood deprivation emerges as a consistent and substantively important finding across all prior studies (including those employing attitudinal outcomes), as well as our own (see also Holland, Silva, and Mace 2012).

Why is the evidence for a negative diversity ‘effect’ so much weaker in the lost letter literature in comparison to attitudinal studies of social trust? Our prior discussion has already suggested some reasons for why behavioral and attitudinal results may differ. We raise here an additional consideration related to the lost letter design itself. Specifically, with our design, we never observe the identity or actions of the letter finder; nor do we know how many individuals encounter the letter before it is picked up. Instead, and in contrast to other studies where the attitudes and behavior of individuals can be directly measured, our experimental setup more resembles a “volunteer’s dilemma” (Volker et al. 2016) in which the presence of only a single ‘nice’ individual is sufficient to ensure a ‘good’ collective outcome.

This design feature may explain why lost letter studies typically find little evidence of ethnic discrimination, which is certainly present in other types of ‘everyday’ interactions where the behavior of *individual* passersby could be studied (Aidenberger and Doehne 2021; Aranguren, Madrisotti, and Durmaz-Martins 2023; Choi, Poertner, and Sambanis 2019; Kanitsar 2023; Mujcic and Frijters 2021; Winter and Zhang 2018; Zhang et al. 2019; Zhang, Gereke, and Baldassarri 2022). Note, however, that two different quantities are being estimated: whereas lost letter experiments study the likelihood that someone in need of assistance will ultimately find help, other field experimental work instead examines the propensity of passersby to offer help. This latter work demonstrates that minorities are helped less often, whereas our results (and those of other lost letter studies) suggest that – although it may require more effort – minorities are as successful as ‘natives’ in (eventually) locating a ‘volunteer’.

This interpretation of our results casts the impact of neighborhood deprivation in an even more striking light. Apparently, such ‘volunteers’ are significantly harder to find in

socioeconomically deprived neighborhoods. This highlights an important policy implication: while we have taken great pains to separate neighborhood diversity from socioeconomic disadvantage, these variables are in fact highly correlated across residential settings. Put simply: diverse neighborhoods tend to be poor neighborhoods. Our results, as well as recent individual-level evidence on the consequences of poverty (Andreoni, Nikiforakis, and Stoop 2021; Schaub 2021), thus further underscore the importance of policies for reducing concentrated disadvantage in enhancing cooperation in contemporary Western societies characterized by continuing immigration and growing ethnic diversity.

## Notes

1. A growing body of work studies prosocial behavior towards ethnic majorities and minorities in 'everyday settings' (Aidenberger and Doehne 2021; Aranguren, Madrisotti, and Durmaz-Martins 2023; Choi, Poertner, and Sambanis 2019; Kanitsar 2023; Mujcic and Frijters 2021; Winter and Zhang 2018; Zhang et al. 2019; Zhang, Gereke, and Baldassarri 2022). However, this work does not examine how behavior changes across field experimental contexts that differ with respect to ethnic diversity. Algan, Hémet, and Laitin (2016) utilize a natural experiment with exogenous allocation to public housing units in France to examine how diversity affects the perceived quality of public facilities (e.g. presence of graffiti, poor building maintenance), but do not observe prosocial behavior directly.
2. Our matching procedure also reduces model dependency when estimating the contextual 'effect' of diversity (Ho et al. 2007).
3. From the perspective of the 'native' majority, the in-group is likely to shrink (at least in relative terms) as diversity increases. In contrast, from a minority point-of-view, the size of the in-group is likely to increase in diverse settings (Koopmans and Veit 2014b).
4. In contrast, research invoking the *Contact Hypothesis* (Allport 1954) posits that, by offering greater opportunities for intergroup interactions, diverse settings may actually improve social cohesion. However, this research must contend with the fact that individuals can often *choose* to engage in (or avoid) interethnic contact (Zhang, Gereke, and Baldassarri 2022), thereby introducing the potential for selection bias. Indeed, more recent (field) experimental studies testing the causal effects of contact have yielded inconsistent results (Elwert, Keller, and Kotsadam 2023; Paluck, Green, and Green 2019).
5. There are good reasons to believe that the threat mechanism is less salient to immigrant minorities than to 'natives'. For one, while dominant groups are often threatened with the loss of their numerical status in diversifying societies (Abascal 2020; Craig, Rucker, and Richeson 2018; Jardina 2019), immigrants may already be accustomed to their minority status in host countries (Smith et al. 2016). Greater exposure to 'natives' should thus not lead to increased feelings of threat.
6. A classic sociological example of this problem is provided by LaPiere (1934), who toured the United States in the 1930s in the company of a Chinese couple. When stopping in restaurants and hotels, the couple were served in the overwhelming majority of cases; however, when these same establishments were surveyed several months later, over 90% unambiguously stated that they would *refuse* service to Chinese guests.
7. Ethics approval for this study was granted by the University of Mannheim. Since we do not know the identity of the letter finders, it was not possible to obtain informed consent. This requirement was therefore waived by the ethics commission.
8. A large number of our field sites were located in one of Germany's largest cities. The remaining field sites were located in multiple cities spread across the Rhein-Ruhr and Rhein-Main metropolitan regions. The Rhein-Ruhr metro area is a traditionally industrial region in Western Germany containing approximately 13 million inhabitants. It comprises numerous agglomerated cities of various sizes. Approximately 15% of residents have foreign



citizenship (country-wide average  $\approx 13\%$ ), and the unemployment rate (7.6%) is also higher than the national average (5.7%). Located further south, the Rhein-Main metro region contains approximately 6 million inhabitants, of which 18% are foreign citizens. The region is somewhat economically better off, with an unemployment rate of 5.2%. Source: <https://www.ikm-monitoring.de/> accessed: 11 Jan 2024. None of our field sites are located in the former East Germany (GDR). Given the history of limited immigration under the GDR, our matching algorithm was unable to find a sufficient number of diverse neighborhoods in the former East.

9. A handwriting machine was used to keep the handwriting consistent across all materials.
10. There are two main reasons why Turkish names were chosen to signal ethnic minority status. First, individuals of Turkish descent constitute the largest ethnic minority in Germany (Destatis 2023). Secondly, Muslim immigrants are more disliked than immigrants overall (Gusciute, Mühlau, and Layte 2021), and Turks make up the majority of Muslims in the country (Pfündel, Stichs, and Tanis 2021). Taken together, these considerations suggest that Turkish names signal a clear and commonly-encountered minority-group identity.
11. This address was not located in any of the cities in which the experiment was conducted.
12. The letter itself consisted of a short innocuous statement acknowledging an inquiry and promising a reply via telephone.
13. We included the ProMigration letters because we conjectured that an immigration organization would be perceived as political. We were therefore curious whether we might observe (greater) bias when immigration is framed as a political issue vs. the case of personal letters between individuals with a migration background.
14. Since we are interested in local patterns of prosociality, we avoided city centers where passersby would be less likely to actually reside in the neighborhood.
15. In a few geographically small neighborhoods, only 8 letters were dropped.
16. We adhered to this rule in most cases. However, in a few geographically large neighborhoods, all letter drops were carried out within a single wave. In addition, due to unforeseen RA unavailability, one wave was either canceled or rescheduled in a small number of neighborhoods.
17. In a small number of cases, the RAs wrote down the wrong code or failed to record all of their letter drops. As a result, we received a few letters with codes that could not be matched to an exact drop location. These are excluded from our analysis.
18. We relied upon official municipal definitions of neighborhood boundaries.
19. Harmonized neighborhood-level data are provided by the Inner-city Spatial Observatory (Innerstädtische Raumbewachung des BBSR 2020), hereafter: *IRB*. Our agreement with the *IRB* prevents us from publishing the names of individual cities or neighborhoods. It is also not possible to make our data public, since the dataset would contain identifying information on the cities and neighborhoods in which the experiment was conducted.
20. Appendix Section 1 lists the national groupings provided by the *IRB*.
21. Three neighborhoods appear twice: once as TARGET NEIGHBORHOODS in the 20%-threshold MATCHED SETS, and once as MATCHED NEIGHBORHOODS in the 30%-threshold MATCHED SETS.
22. We note two slight deviations from this general rule: (1) due to an RA oversight, 2 letters were dropped on the following day; (2) due to unforeseen RA unavailability, letter drops in 2 neighborhoods had to be rescheduled to a different day.
23. Of course, one could argue for comparing neighborhoods at the extremes of the x-axis, but then we run into the problem of confounding diversity with neighborhood deprivation since there are relatively few areas which are either (i) non-diverse and deprived, or (ii) highly diverse but not deprived.
24. Information on migration background is unfortunately not given for any national groups in the *IRB* data.
25. This is a key difference between lost letter studies and other field experiments on ‘everyday discrimination’ (Aidenberger and Doehne 2021; Aranguren, Madrisotti, and Durmaz-Martins 2023; Choi, Poertner, and Sambanis 2019; Kanitsar 2023; Mujcic and Frijters 2021;

Winter and Zhang 2018; Zhang et al. 2019; Zhang, Gereke, and Baldassarri 2022). At the same time, most of these ‘everyday discrimination’ studies are fielded in only a few locations and, to our knowledge, none attempts – as we do here – to examine how treatment effects are moderated by characteristics of the field setting.

26. The analyzes reported in Equations (1) and (2) were pre-registered in OSF <https://doi.org/10.17605/OSF.IO/VXG29>. Our pre-registration also indicated that we would estimate models that include statistical controls for factors such as the number of pedestrians / shops on the street, the presence of street-clean crews, etc. However, technical difficulties faced by some RAs during data collection prevented us from collecting this information for the entire dataset. We therefore omit these analyses from the paper under the assumption that, due to our randomization procedures, such variables are balanced across treatments in expectation.
27. We also pooled personal and official letters and re-estimated Equations (1) and (2), focusing on the differential return rates of ‘native’ majority vs. ethnic minority letters of all ‘types.’ Results (not shown) are unchanged.
28. Each observation is thus a neighborhood. Observations are weighted by the number of letters dropped in each area. Unlike the results presented in the previous section, our analyses here were not pre-registered and should therefore be treated as exploratory.
29. Interestingly, Ahmed (2010) shows that discrimination only occurs when the lost letters contain money (i.e. there is a financial gain from not posting the letter).

## Acknowledgments

We would like to thank Melike Kaplan for excellent research assistance.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

Generous financial support for this project was provided by the German Research Foundation (DFG)’s Emmy-Noether Program (ZH 613/1-1).

## Funding

Generous financial support for this project was provided by the German Research Foundation (DFG)’s Emmy-Noether Program (ZH 613/1-1).

## References

- Abascal Maria. 2020. “Contraction As a Response to Group Threat: Demographic Decline and Whites’ Classification of People Who are Ambiguously White.” *American Sociological Review* 85 (2): 298–322. <https://doi.org/10.1177/0003122420905127>.
- Abascal Maria, and Baldassarri Delia. 2015. “Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined.” *American Journal of Sociology* 121 (3): 722–782. <https://doi.org/10.1086/683144>.
- Abascal Maria, Ganter Flavien, and Baldassarri Delia. 2023. “Greater Diversity Or Fewer Whites? Disentangling Heterogeneity and Marginalized Group Share At Macro and Micro Levels.” *Socius*9:23780231231196507. <https://doi.org/10.1177/23780231231196507>.

- Ahmed Ali M. 2010. "Muslim Discrimination: Evidence From Two Lost-letter Experiments." *Journal of Applied Social Psychology* 40 (4): 888–898. <https://doi.org/10.1111/jasp.2010.40.issue-4>.
- Aidenberger Amelie, and Doehne Malte. 2021. "Unveiling Everyday Discrimination. Two Field Experiments on Discrimination Against Religious Minorities in Day-to-day Interactions." *The British Journal of Sociology* 72 (2): 328–346. <https://doi.org/10.1111/bjos.v72.2>.
- Algan Yann, Hémet Camille, and Laitin David D. 2016. "The Social Effects of Ethnic Diversity At the Local Level: A Natural Experiment with Exogenous Residential Allocation." *Journal of Political Economy* 124 (3): 696–733. <https://doi.org/10.1086/686010>.
- Allport Gordon W. 1954. *The Nature of Prejudice*. Reading MA: Addison-Wesley Publishing Company.
- Andreoni James, Nikiforakis Nikos, and Stoop Jan. 2021. "Higher Socioeconomic Status Does Not Predict Decreased Prosocial Behavior in a Field Experiment." *Nature Communications* 12 (1): 4266. <https://doi.org/10.1038/s41467-021-24519-5>.
- Aranguren Martin, Madrisotti Francesco, and Durmaz-Martins Eser. 2023. "Anti-Muslim Behaviour in Everyday Interaction: Evidence From a Field Experiment in Paris." *Journal of Ethnic and Migration Studies* 49 (3): 770–794. <https://doi.org/10.1080/1369183X.2021.1953378>.
- Baldassarri Delia. 2020. "Market Integration Accounts for Local Variation in Generalized Altruism in a Nationwide Lost-letter Experiment." *Proceedings of the National Academy of Sciences* 117 (6): 2858–2863. <https://doi.org/10.1073/pnas.1819934117>.
- Balliet Daniel, Wu Junhui, and De Dreu Carsten K. W. 2014. "Ingroup Favoritism in Cooperation: a Meta-analysis." *Psychological Bulletin* 140 (6): 1556. <https://doi.org/10.1037/a0037737>.
- Bernhard Helen, Fischbacher Urs, and Fehr Ernst. 2006. "Parochial Altruism in Humans." *Nature* 442 (7105): 912–915. <https://doi.org/10.1038/nature04981>.
- Blalock Hubert M. 1967. *Toward a Theory of Minority-group Relations*. New York: John Wiley.
- Choi Donghyun D., Poertner Mathias, and Sambanis Nicholas. 2019. "Parochialism, Social Norms, and Discrimination Against Immigrants." *Proceedings of the National Academy of Sciences* 116 (33): 16274–16279. <https://doi.org/10.1073/pnas.1820146116>.
- Craig Maureen A., Rucker Julian M., and Richeson Jennifer A. 2018. "Racial and Political Dynamics of An Approaching 'majority-minority' United States." *The ANNALS of the American Academy of Political and Social Science* 677 (1): 204–214. <https://doi.org/10.1177/0002716218766269>.
- Dancygier Rafaela M. 2018. *Dilemmas of Inclusion: Muslims in European Politics*. Princeton, NJ: Princeton University Press.
- Destatis. 2023. "Statistischer Bericht: Mikrozensus – Bevölkerung nach Migrationshintergrund. Erstergebnisse 2022."
- Dinesen Peter T., Schaeffer Merlin, and Sønderskov Kim M. 2020. "Ethnic Diversity and Social Trust: A Narrative and Meta-analytical Review." *Annual Review of Political Science* 23 (1): 441–465. <https://doi.org/10.1146/polisci.2020.23.issue-1>.
- Elwert Felix, Keller Tamás, and Kotsadam Andreas. 2023. "Rearranging the Desk Chairs: A Large Randomized Field Experiment on the Effects of Close Contact on Interethnic Relations." *American Journal of Sociology* 128 (6): 1809–1840. <https://doi.org/10.1086/724865>.
- Enos Ryan D. 2014. "Causal Effect of Intergroup Contact on Exclusionary Attitudes." *Proceedings of the National Academy of Sciences* 111 (10): 3699–3704. <https://doi.org/10.1073/pnas.1317670111>.
- Gereke Johanna, Schaub Max, and Baldassarri Delia. 2018. "Ethnic Diversity, Poverty and Social Trust in Germany: Evidence From a Behavioral Measure of Trust." *PloS One* 13 (7): e0199834. <https://doi.org/10.1371/journal.pone.0199834>.
- Gusciute Egle, Mühlau Peter, and Layte Richard. 2021. "All Welcome Here? Attitudes Towards Muslim Migrants in Europe." *International Migration* 59 (5): 149–165. <https://doi.org/10.1111/imig.v59.5>.
- Haushofer Johannes, and Fehr Ernst. 2014. "On the Psychology of Poverty." *Science* 344 (6186): 862–867. <https://doi.org/10.1126/science.1232491>.
- Hellmann Jens H., Berthold Anne, Rees Jonas H., and Hellmann Deborah F. 2015. "'A Letter for Dr. Outgroup': on the Effects of An Indicator of Competence and Chances for Altruism Toward

- a Member of a Stigmatized out-group.” *Frontiers in Psychology* 6:1422. <https://doi.org/10.3389/fpsyg.2015.01422>.
- Ho Daniel E., Imai Kosuke, King Gary, and Stuart Elizabeth A. 2007. “Matching As Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference.” *Political Analysis* 15 (3): 199–236. <https://doi.org/10.1093/pan/mpl013>.
- Holland Jo, Silva Antonio S., and Mace Ruth. 2012. “Lost Letter Measure of Variation in Altruistic Behaviour in 20 Neighbourhoods.” *PloS One* 7 (8): e43294. <https://doi.org/10.1371/journal.pone.0043294>.
- Hopkins Daniel J. 2010. “Politicized Places: Explaining Where and when Immigrants Provoke Local Opposition.” *American Political Science Review* 104 (1): 40–60. <https://doi.org/10.1017/S0003055409990360>.
- Innerstädtische Raumbearbeitung des BBSR. 2020. “Datengrundlage: Kommunalstatistiken der IRB-Städte / Statistik der Bundesagentur für Arbeit.”
- Jardina Ashley. 2019. *White Identity Politics*. New York: Cambridge University Press.
- Kahneman Daniel. 2011. *Thinking, Fast and Slow*. New York: Macmillan.
- Kanitsar Georg. 2023. “Leaving the Bike Unlocked: Trust Discrimination in Inter-ethnic Encounters.” *European Sociological Review*. <https://doi.org/10.1093/esr/jcad045>.
- Keizer Kees, Lindenberg Siegwart, and Steg Linda. 2008. “The Spreading of Disorder.” *Science* 322 (5908): 1681–1685. <https://doi.org/10.1126/science.1161405>.
- Koopmans Ruud, and Veit Susanne. 2014a. “Cooperation in Ethnically Diverse Neighborhoods: A Lost-letter Experiment.” *Political Psychology* 35 (3): 379–400. <https://doi.org/10.1111/pops.2014.35.issue-3>.
- Koopmans Ruud, and Veit Susanne. 2014b. “Ethnic Diversity, Trust, and the Mediating Role of Positive and Negative Interethnic Contact: A Priming Experiment.” *Social Science Research* 47:91–107. <https://doi.org/10.1016/j.ssresearch.2014.03.014>.
- Kustov Alexander, and Pardelli Giuliana. 2018. “Ethnoracial Homogeneity and Public Outcomes: The (non) Effects of Diversity.” *American Political Science Review* 112 (4): 1096–1103. <https://doi.org/10.1017/S0003055418000308>.
- LaPiere Richard T. 1934. “Attitudes Vs. Actions.” *Social Forces* 13 (2): 230–237. <https://doi.org/10.2307/2570339>.
- Letki Natalia. 2008. “Does Diversity Erode Social Cohesion? Social Capital and Race in British Neighbourhoods.” *Political Studies* 56 (1): 99–126. <https://doi.org/10.1111/j.1467-9248.2007.00692.x>.
- Milgram Stanley, Mann Leon, and Harter Susan. 1965. “The Lost-letter Technique: A Tool of Social Research.” *Public Opinion Quarterly* 29 (3): 437. <https://doi.org/10.1086/267344>.
- Mujcic Redzo, and Frijters Paul. 2021. “The Colour of a Free Ride.” *The Economic Journal* 131 (634): 970–999. <https://doi.org/10.1093/ej/ueaa090>.
- Newman Benjamin J. 2013. “Acculturating Contexts and Anglo Opposition to Immigration in the United States.” *American Journal of Political Science* 57 (2): 374–390. <https://doi.org/10.1111/ajps.2013.57.issue-2>.
- Pager Devah, and Quillian Lincoln. 2005. “Walking the Talk? What Employers Say Versus what they Do.” *American Sociological Review* 70 (3): 355–380. <https://doi.org/10.1177/000312240507000301>.
- Paluck Elizabeth Levy, Green Seth A., and Green Donald P. 2019. “The Contact Hypothesis Re-evaluated.” *Behavioural Public Policy* 3 (2): 129–158. <https://doi.org/10.1017/bpp.2018.25>.
- Pfündel Katrin, Sticks Anja, and Tanis Kerstin. 2021. “Muslimisches Leben in Deutschland 2020: Studie im Auftrag der Deutschen Islam Konferenz.” Nuremberg: Bundesamt für Migration und Flüchtlinge.
- Putnam Robert D. 2007. “E Pluribus Unum: Diversity and Community in the Twenty-first Century.” *Scandinavian Political Studies* 30 (2): 137–174. <https://doi.org/10.1111/scps.2007.30.issue-2>.
- Sampson Robert J. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago, IL: University of Chicago Press.
- Schaeffer Merlin. 2013. “Can Competing Diversity Indices Inform Us About why Ethnic Diversity Erodes Social Cohesion? A Test of Five Diversity Indices in Germany.” *Social Science Research* 42 (3): 755–774. <https://doi.org/10.1016/j.ssresearch.2012.12.018>.

- Schaeffer Merlin. 2016. *Ethnic Diversity and Social Cohesion: Immigration, Ethnic Fractionalization and Potentials for Civic Action*. New York: Routledge.
- Schaub Max. 2021. "Acute Financial Hardship and Voter Turnout: Theory and Evidence From the Sequence of Bank Working Days." *American Political Science Review* 115 (4): 1258–1274. <https://doi.org/10.1017/S0003055421000551>.
- Schönwälder Karen, and Söhn Janina. 2009. "Immigrant Settlement Structures in Germany: General Patterns and Urban Levels of Concentration of Major Groups." *Urban Studies* 46 (7): 1439–1460. <https://doi.org/10.1177/0042098009104575>.
- Smith Sanne, Van Tubergen Frank, Maas Ineke, and McFarland Daniel A. 2016. "Ethnic Composition and Friendship Segregation: Differential Effects for Adolescent Natives and Immigrants." *American Journal of Sociology* 121 (4): 1223–1272. <https://doi.org/10.1086/684032>.
- Tajfel Henri. 1982. "Social Psychology of Intergroup Relations." *Annual Review of Psychology* 33 (1): 1–39. <https://doi.org/10.1146/psych.1982.33.issue-1>.
- Volker Beate, Mollenhorst Gerald, Steenbeek Wouter, Schutjens Veronique, and Flap Henk. 2016. "Lost Letters in Dutch Neighborhoods: A Field Experiment on Collective Efficacy." *Social Forces* 94 (3): 953–974. <https://doi.org/10.1093/sf/sov106>.
- Winter Fabian, and Zhang Nan. 2018. "Social Norm Enforcement in Ethnically Diverse Communities." *Proceedings of the National Academy of Sciences* 115 (11): 2722–2727. <https://doi.org/10.1073/pnas.1718309115>.
- Zhang Nan, Aidenberger Amelie, Rauhut Heiko, and Winter Fabian. 2019. "Prosocial Behaviour in Interethnic Encounters: Evidence From a Field Experiment with High-and Low-status Immigrants." *European Sociological Review* 35 (4): 582–597. <https://doi.org/10.1093/esr/jcz030>.
- Zhang Nan, Gereke Johanna, and Baldassarri Delia. 2022. "Everyday Discrimination in Public Spaces: a Field Experiment in the Milan Metro." *European Sociological Review* 38 (5): 679–693. <https://doi.org/10.1093/esr/jcac008>.

## Appendices

### Appendix 1. Classification of Minority Groups

The IRB provides neighborhood counts of the non-German population grouped into the following national categories:

- Italy, Vatican, San Marino, Malta
- Greece, Cyprus
- Turkey
- Spain, Portugal, Andorra
- Estonia, Latvia, Lithuania, Poland, Czech Republic, Slovakia, Slovenia, Hungary
- Romania, Bulgaria, Croatia
- Russia, Ukraine, Belarus, Moldova
- Albania, Bosnia-Herzegovina, Serbia, Montenegro, Macedonia
- Belgium, France, Luxembourg, the Netherlands, Denmark, Ireland, Austria, Finland, Sweden, the UK
- Norway, Iceland, Switzerland, Lichtenstein, Monaco
- North Africa
- Sub-Saharan Africa
- USA and Canada
- Mexico, Central, and South America, and the Caribbean
- Central Asian states that were formerly part of the USSR
- Middle Eastern states
- Iran, Afghanistan, Pakistan, India, Bangladesh, and Sri Lanka

- East Asian states
- Southeast Asian states
- Australia, New Zealand and Oceania

Minority fractionalization is calculated based on these groupings using the following formula:

$$HHI = 1 - \sum_{i=1}^N S_i^2 \tag{A1}$$

where  $S_i$  refers to the population of minority group  $i$  divided by the total non-German population.

## Appendix 2. Balance Tables

The tables compare characteristics of neighborhoods above and below the thresholds for percentage non-German used in the matching algorithm. Means, differences-in-means, and t-test p-values are displayed.

As expected, the percentages of non-Germans and individuals with migration background are significantly different across the thresholds, but the neighborhoods are otherwise very similar on the matching variables.

**Table A1.** Balance on neighborhood characteristics around the 20% threshold.

	<20% non-Germans	>20% non-Germans	Diff in Means	P-value
Perc. Non-German	14.386	24.039	-9.653	0.000
Perc. Migration Background	28.987	44.440	-15.453	0.000
Total Population / 100	105.918	115.823	-9.904	0.529
10-year change in Perc. Foreign	5.373	5.841	-0.467	0.493
Unemployment Rate	5.885	5.843	0.042	0.934
Unemployment Gap	3.738	2.783	0.955	0.128
Minority Fractionalization	0.884	0.901	-0.017	0.190
Perc. Longterm Residents	65.352	64.147	1.205	0.388

Note:  $N = 53$  neighborhoods

**Table A2.** Balance on neighborhood characteristics around the 30% threshold.

	<30% non-Germans	>30% non-Germans	Diff in Means	P-value
Perc. Non-German	25.450	33.057	-7.606	0.000
Perc. Migration Background	44.277	53.974	-9.697	0.000
Total Population / 100	102.436	113.026	-10.590	0.503
10-year change in Perc. Foreign	7.643	8.333	-0.690	0.589
Unemployment Rate	6.973	7.229	-0.256	0.727
Unemployment Gap	2.834	2.596	0.238	0.778
Minority Fractionalization	0.892	0.908	-0.016	0.188
Perc. Longterm Residents	63.093	61.896	1.198	0.404

Note:  $N = 27$  neighborhoods