Childhood exposure to ethnic outgroups predicts interethnic marriage

Leonard Wendering[1,*], Nan Zhang[1,2], Kerstin Ostermann[3,4]

[1]Mannheim Centre for European Social Research (MZES), University of Mannheim, 68159 Mannheim, Germany.

[2]School of Social Sciences, University of Mannheim, 68159 Mannheim, Germany.

[3]Institute for Employment Research (IAB), 90478 Nuremberg, Germany.

[4] Leibniz ScienceCampus RegioHub, University Bielefeld, 33615 Bielefeld, Germany.

[*]Corresponding author. Email: leonard.wendering@uni-mannheim.de

Scholars posit that childhood exposure to ethnic outgroups may durably improve intergroup relations. However, few studies to date track the consequences of childhood experiences across multiple decades. Moreover, existing research focuses on majority populations and largely overlooks the impact of exposure for minorities. Using linked geocoded US census records from 1880 to 1900/1910, we examine childhood outgroup exposure and subsequent marriage patterns for over 400,000 American, German and Irish men. To account for residential self-selection, we use machine-learning to identify "ethnic" neighborhoods and compare individuals living within the same residential contexts but having different next-door neighbors. In majority-minority dyads, outgroup exposure is consistently associated with a higher likelihood of marrying a member of that group, as well as a lower likelihood of ingroup marriage. The evidence is less robust for minority-minority dyads but nonetheless supports an overall pattern wherein childhood experiences can serve to erode ethnic group boundaries.

Recent waves of immigration are reshaping the demographics of Western societies while simultaneously fueling public anxieties about the social and political consequences of increasing ethnoracial diversity. Indeed, demographic changes have been linked to prejudicial attitudes and discriminatory behavior (*1–4*), diminished social trust (*5*), and support for far-right parties (*6, 7*). At the same time, scholars and policymakers hold out hope that societies will adapt to diversity over the long-run (*8, 9*). According to the contact hypothesis, positive inter-ethnic contact reduces group conflict and prejudice (*10–12*). Further, contact's effect should be most pronounced for children and adolescents (*13–15*) as racial attitudes are formed during these early years (*16, 17*).

To date however, most studies examine short-term outcomes measured immediately (*18–20*), weeks (*4, 21–23*), or a few months (*24–26*) after contact takes place (*10, 27*). In contrast, only a handful of studies are able to track the consequences of childhood outgroup exposure across decades (*13, 28–30*) and thereby examine whether these experiences durably affect intergroup relations. Further, these studies focus on the experiences of majority populations and thereby overlook how contact shapes relations among ethnic and racial minorities. In particular, minority reactions to other minority groups may be conditioned by heightened sensitivity to group status (*31, 32*), as well as different baseline levels of prejudice (in comparison to majority-group members) (*33*).

To address these limitations, we study the long-term consequences of outgroup exposure during the "age of mass migration" to the United States at the turn of the 20th century. Then, as now, foreign-born individuals made up approximately 15% of the American population (*34*). Many immigrants settled in industrializing cities, giving rise to opportunities for intergroup contact within urban neighborhoods (*35*). Using over 400,000 individually-linked census records between 1880 and 1900/1910, we investigate how childhood exposure to outgroup neighbors influences subsequent marriage decisions for "native-born" Americans and German and Irish immigrants (the two largest immigrant groups at the time). In this way, we study how outgroup exposure shapes (i) majority relations with minority groups, as well as minorities' relations towards both (ii) the dominant majority and (iii) other minority outgroups.

Marital outcomes figure prominently in the literature on group relations (*36*). On the one hand, marriage patterns display a high degree of homophily, reflecting both ingroup preferences and norms as well as social distance between groups (*37*). On the other hand, ethnic intermarriage indicates a breakdown of parochial identities and group boundaries (*32*). In our context, childhood exposure may affect marriage patterns by increasing *opportunities* to meet and partner with members of the outgroup, and/or altering *preferences* for outgroup vs. ingroup partners (*28–30*). This latter mechanism may have been particularly relevant for marriages involving Irish immigrants who, in comparison to Germans and other northern Europeans, were viewed more negatively during the period (*35, 38*).

Our empirical strategy exploits the ordering of households within census enumeration sheets to identify childhood neighbors. More specifically, since census enumerators visited households door-to-door, neighboring households appear as adjacent records on the census page (*13, 39*). Combining this feature with census information on place of birth, we construct measures of childhood exposure for males aged 0-18 in 1880. We then use linked records from 1880-1900 and 1880-1910 (*40, 41*) to measure subsequent marital outcomes (since women are likely to change their names upon marriage, we restrict our attention to males only). We code the spouse's ethnicity based on information on her place of birth as well as the birthplace of her parents (see METHODS AND MATERIALS for details). This combination of data allows us to examine the relationship between marital choices and early-life exposure to outgroups.

One potential issue with this analysis is that individuals who were (more) exposed to a particular outgroup during childhood may differ from the unexposed on observable and unobservable characteristics that also influence subsequent marriage decisions. For instance, parents who are more favorably disposed towards ethnic outgroups may be more likely to settle or remain in diverse neighborhoods (*42*). To the extent that parents' attitudes are then transmitted to their children (*43, 44*), they may influence the next generation's marriage decisions *independently* of intergroup exposure.

We address this possibility by accounting for 1880 neighborhood characteristics that plausibly reflect parental attitudes. Specifically, we leverage the georeferencing of all households in 40 cities by (*45*) (see Fig.1, left panel). Using these geolocations, we apply a machine-learning algorithm from (*46*) (see METHODS AND MATERIALS) that identifies neighborhoods based on the density and ethnicity of households (Fig.1, right panel). This allows us to compare marital outcomes among individuals living within the same types of neighborhoods.

Our analysis focuses on white native-born Americans and German and Irish immigrants. Together, these three groups make up approximately 85% of linked records in our 40 cities. For each group, we examine exposure to the other two outgroups, resulting in six dyads represented in

Fig. 2A. For each dyad, we use multinomial logistic regression to model the conditional probabilities of being (i) married to a member of the TARGET OUTGROUP, (ii) the INGROUP, (iii) a THIRD OUTGROUP, (iv) UNMARRIED, or (v) OTHERWISE CLASSIFIED as divorced / widowed / “spouse not present”. This last designation captures a small number of married individuals for whom the identity of the spouse could not be determined among listed household members. Finally, in dyads involving German and Irish men with, respectively, Irish and German neighbors, we also include (vi) marriage to an AMERICAN as an additional outcome.

We apply two complementary analytical strategies to operationalize outgroup exposure (see METHODS AND MATERIALS). First, we use continuous explanatory variables which COUNT the number of households belonging to a target ethnic group among the ten closest neighbors up and down the census enumeration sheet (see Fig. 2B). Our models include individual-level covariates (e.g. birthyear, nativity status, school attendance, and occupational information) and neighborhood fixed effects to account for residential selection. Second, we employ Coarsened Exact Matching (CEM) (*47*) on the aforementioned covariates, where TREATED units are defined as those having $\geq$ 1 (out of max. = 2) *immediate* next-door neighbors belonging to the target ethnic outgroup, and CONTROL units are defined as having two immediate ingroup neighbors (to approximate “close exposure” to the ingroup only)(see Fig. 2C). While CEM reduces model dependency, it comes at the cost of fewer observations.

Separate analyses are conducted for the 1880-1900 and 1880-1910 linked samples. Given the large number of analytical decisions and individual estimates, we adopt a multiverse approach which focuses on broadly consistent patterns across an extensive set of analyses. Accordingly, our results are presented as a series of specification curves (*48*). Full regression results are presented in tables S1 – S4 of the Supplementary Materials.

Outgroup Exposure and Marriage Patterns

Childhood exposure to outgroup neighbors is associated with a higher likelihood of marrying a member of the target group (Fig. 3). To illustrate, the CEM analysis shows that American boys growing up next to German households were 3.5pp (95% CI = [1.7pp – 5.3pp]) more likely to marry German wives by 1910 compared to similar American boys with only American neighbors.

This is a sizable difference, given that the baseline probability of American-German marriages in the control group is only around 5pp. Alternative analyses utilizing the NEIGHBOR COUNT variable yield a substantively consistent pattern. Since NEIGHBOR COUNT ranges from 0 to 20, the coefficients in the right panel of Fig. 3 represent the difference in marriage probabilities associated with a 5pp change in the share of neighbors belonging to the target outgroup. If we again consider American boys in the 1880-1910 sample, our estimate implies that, for each additional German neighbor, the likelihood of marrying a German wife is 0.6pp higher. This is equivalent to about 9% of the sample mean (of 6.3pp). The magnitude of this relationship is similar to findings from (*28, 30*) who study contemporary outcomes.

Overall, the estimates in Fig. 3 indicate that childhood exposure consistently and positively predicts marriage to the target outgroup. Additionally, the similarity in findings for German→American and Irish→American dyads suggests that the greater skepticism directed against the Irish at the time did not constitute an insurmountable barrier to partnership formation. Further, the patterns we uncover are not unique to immigrant→native relations, but also apply in native→immigrant encounters (i.e. German boys exposed to American neighbors). Finally, we note that the weakest evidence comes from our CEM estimates involving German→Irish and Irish→German dyads (where the 1880-1910 samples yield statistically insignificant results). Even here, however, the coefficients point in the direction of a positive (but imprecisely estimated) relationship between outgroup exposure and ethnic intermarriage.

Fig. 4 presents evidence that childhood exposure is also associated with a lower likelihood of ingroup marriages, at least for American and German boys. For these groups, the CEM analyses indicate that TREATED individuals are between 1.2pp to 7.1pp less likely to marry endogamously, reflecting a decrease of between 7.6% to 18.5% relative to the CONTROL baseline. For Irish boys, we find patterns of decreasing endogamous marriage using the NEIGHBOR COUNT variable; however, results from the CEM analysis are not statistically significant.

Finally, we consider whether exposure to the target outgroup predicts marriage to a *third* outgroup. If so, this would indicate a general erosion of ingroup boundaries (beyond improved relations towards a specific outgroup). Results are presented in Fig. S1 of the Supplementary Materials. Here, our evidence is inconclusive: while we find positive coefficients in half of the models employing the NEIGHBOR COUNT variable, the majority of the CEM estimates are not significant. We also find no consistent evidence that childhood exposure is associated with the likelihood of remaining unmarried (Fig. S2) or of having an OTHERWISE CLASSIFIED marital status (Fig. S3).

To assess the robustness of our results to unobserved confounders, we conduct sensitivity analyses using the E-value framework (*49*) (see METHODS AND MATERIALS). For American boys exposed to immigrant neighbors, an unmeasured confounder would need to increase the likelihood of both exposure and intermarriage by a factor of 2.3 to 3.0 to nullify our results (tables S5 – S52). To put these values in context, covariates such as BIRTH YEAR, SCHOOL ATTENDANCE, and NEIGHBORHOOD DENSITY are rarely associated with either outgroup exposure or the outcome at these levels. Associations between OCCUPATION OF THE HOUSEHOLD HEAD (comparing professionals to laborers) and the outcome occasionally exceeds the E-value. However, OCCUPATION tends to be more weakly associated with outgroup exposure. Overall, our analyses suggest that an unobserved confounder would need to be relatively powerful to explain away our findings.

Summary and Implications

In summary, we analyzed linked census records from the "age of mass migration" to study the association between childhood exposure to other ethnic groups and the subsequent likelihood of marriage across ethnic boundaries. We find that, for dyads involving immigrants and "native-born" Americans, exposure to an outgroup in 1880 is consistently associated with a higher likelihood of marriage to a member of that group in later life, as well as a lower likelihood of ingroup marriage. The evidence is less robust for German-Irish encounters but nonetheless supportive of an overall pattern wherein childhood experiences can serve to erode ethnic group boundaries.

We acknowledge that our analysis is not without limitations. Although we attempted to mitigate endogeneity concerns by comparing boys living in similar neighborhoods and who are identical on a range of individual-level sociodemographic characteristics, we cannot rule out the influence of unobserved omitted variables entirely. Additionally, our study only examines marriages between white "ethnic" groups, but not between whites and non-whites (since inter-racial marriages were exceedingly rare during this period). Finally, our reliance on names-based data linkage restricts our attention to boys, while side-stepping an examination of childhood exposure among girls.

That said, we note the convergence in findings between our study and other research (employing alternative identification strategies) on the long-term consequences of childhood exposure in contemporary Sweden (*29*), Finland (*28*) and the United States (*30*). Moreover, while prior studies focus on the consequences of exposure for majority group members, our study additionally documents similar patterns among minority groups. We thus extend the evidence base beyond the typical majority-centric paradigm (*31, 50*).

Our results have implications for the broader literature on intergroup contact. Scholars in this field have typically focused on outcomes such as prejudice reduction (*10, 21, 23, 25, 51*) and social integration (*8, 52, 53*). We argue that many of these outcomes are causally prior to interethnic marriage – which some even refer to as the "final stage of assimilation" (*32*). Hence, our findings imply that early childhood exposure may have positive effects on these antecedent outcomes as well (*54*).

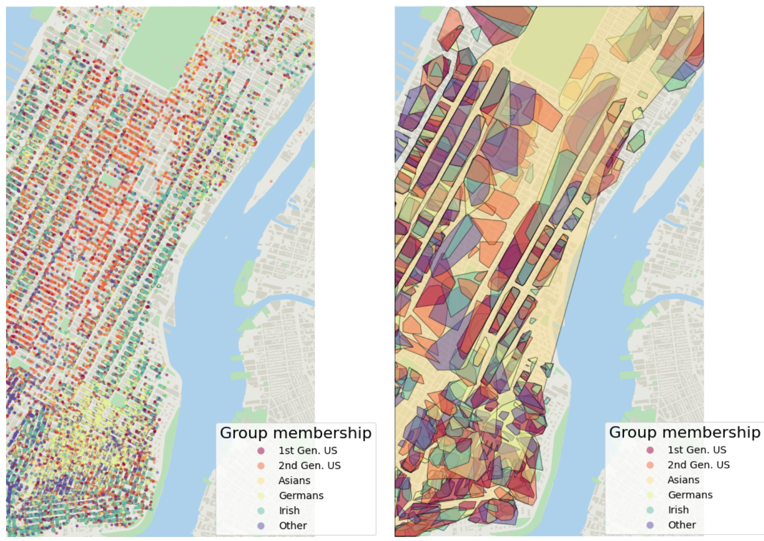

Figure 1: **Neighborhoods in Manhattan in 1880.** The left panel displays the geolocations of 1880 households, tagged by the ethnicity of the household head. Data from the US Census 1880. The right panel displays the results of our neighborhood identification HDBSCAN algorithm with $min\_cluster\_size = 80$. Each colored polygon represents one neighborhood. Many neighborhoods are confined to single city blocks but some, especially in the center of Manhattan and on the Lower East Side, expand over larger spatial areas. Our analyses account for the *combination* of polygons in which an individual resides. The background layers (e.g. buildings, water) depict present-day New York City (https://opendata.cityofnewyork.us/).

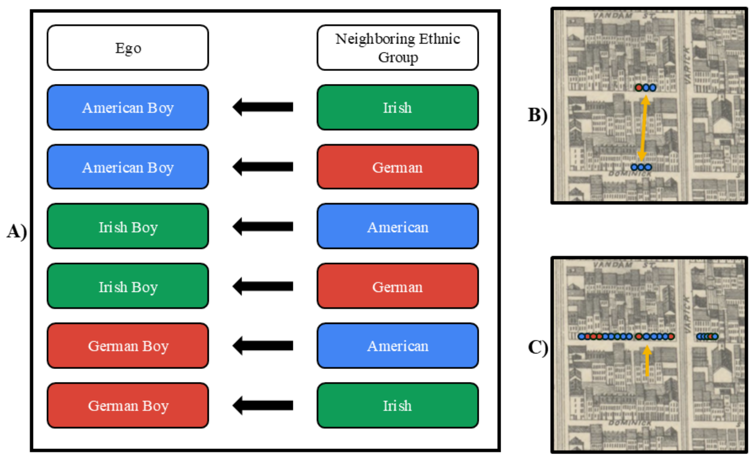

Figure 2: **Visualization of Analytical Strategy.** A) Dyads Analyzed: For each American, German, or Irish boy ("Ego") in the 1880 census, we examine exposure to the other two outgroups. B) CEM Analysis: We compare marriage outcomes of boys with at least one ethnic outgroup neighbor (denoted by the red dot) to matched boys (indicated by the double-headed yellow arrow) with two ethnic ingroup neighbors. C) Neighbor Count Analysis: For each Ego (denoted by the single-headed yellow arrow), our explanatory variable counts the number of households belonging to a target ethnic group (red dots) among the ten closest neighbors up and down the census enumeration sheet. The background layers of B and C depict a block of "The City of New York" (1879) by Will L. Taylor (https://www.loc.gov/item/75694818/)

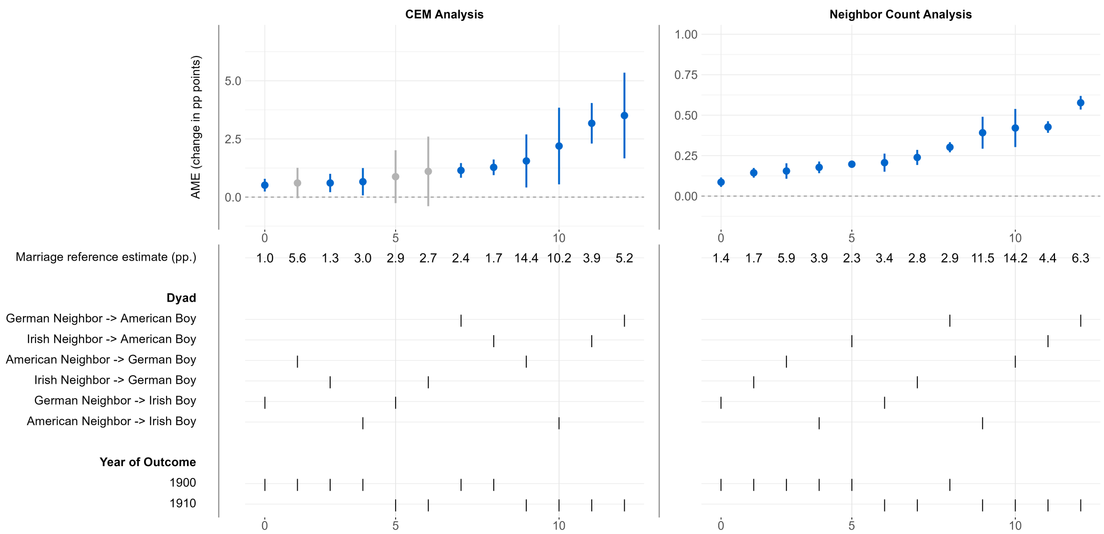

Figure 3: **Specification Curve of Predicted Probabilities of Marrying the Target Group.** The upper panels display point estimates and 95% confidence intervals from separate models (indicated in the bottom panel). The left panel presents results from the CEM analysis, while the right panel utilizes the NEIGHBOR COUNT variable. Reference estimates indicate the predicted probability of the outcome among CONTROL observations in the left panel, and the unconditional sample mean in the right panel. AMEs (Average Marginal Effects) in the left panel refer to the difference in predicted probabilities between TREATED VS. CONTROL observations. AMEs in the right panel denote the change in predicted probabilities resulting from every additional outgroup household among the 20 closest neighbors.

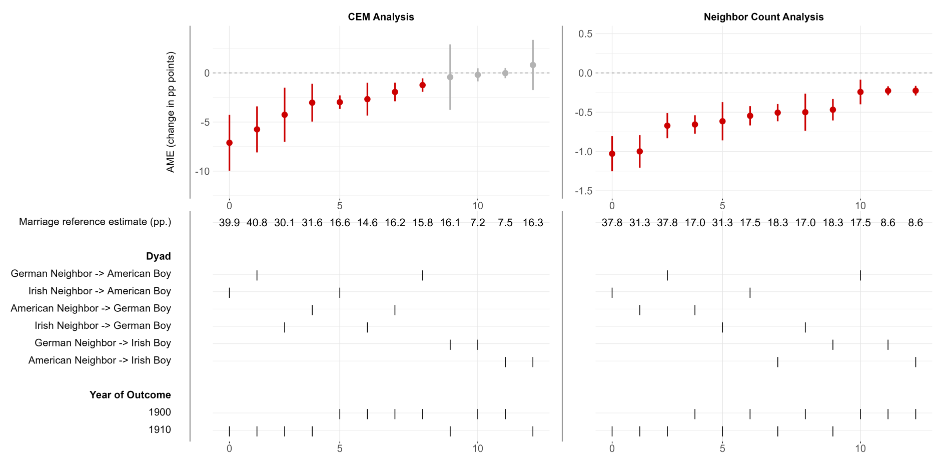

Figure 4: **Specification Curve of Predicted Probabilities of Ingroup Marriage.** The upper panels display point estimates and 95% confidence intervals from separate models (indicated in the bottom panel). The left panel presents results from the CEM analysis, while the right panel utilizes the NEIGHBOR COUNT variable. Reference estimates indicate the predicted probability of the outcome among CONTROL observations in the left panel, and the unconditional sample mean in the right panel. AMEs (Average Marginal Effects) in the left panel refer to the difference in predicted probabilities between TREATED VS. CONTROL observations. AMEs in the right panel denote the change in predicted probabilities resulting from every additional outgroup household among the 20 closest neighbors.

References and Notes

1. M. Abrajano, Z. L. Hajnal, *White Backlash: Immigration, Race, and American Politics* (Princeton University Press) (2015).

2. M. A. Craig, J. M. Rucker, J. A. Richeson, 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 (2018), doi:10.1177/0002716218766269, http://dx.doi.org/10.1177/0002716218766269.

3. D. Hangartner, E. Dinas, M. Marbach, K. Matakos, D. Xefteris, Does exposure to the refugee crisis make natives more hostile? *American Political Science Review* **113** (2), 442–455 (2019).

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

5. P. T. Dinesen, M. Schaeffer, K. M. Sønderskov, Ethnic Diversity and Social Trust: A Narrative and Meta-Analytical Review. *Annual Review of Political Science* **23** (1), 441–465 (2020), doi:10.1146/annurev-polisci-052918-020708, http://dx.doi.org/10.1146/annurev-polisci-052918-020708.

6. A. Abramowitz, J. McCoy, United States: Racial Resentment, Negative Partisanship, and Polarization in Trump's America. *The ANNALS of the American Academy of Political and Social Science* **681** (1), 137–156 (2018), doi:10.1177/0002716218811309, http://dx.doi.org/10.1177/0002716218811309.

7. E. Dinas, K. Matakos, D. Xefteris, D. Hangartner, Waking up the golden dawn: does exposure to the refugee crisis increase support for extreme-right parties? *Political Analysis* **27** (2), 244–254 (2019).

8. R. D. Putnam, E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies* **30** (2), 137–174 (2007), doi:10.1111/j.1467-9477.2007.00176.x, http://dx.doi.org/10.1111/j.1467-9477.2007.00176.x.

9. R. D. Enos, *The Space between Us: Social Geography and Politics* (Cambridge University Press) (2017), doi:10.1017/9781108354943, http://dx.doi.org/10.1017/9781108354943.

10. E. L. Paluck, S. A. Green, D. P. Green, The contact hypothesis re-evaluated. *Behavioural Public Policy* **3** (02), 129–158 (2018), doi:10.1017/bpp.2018.25, http://dx.doi.org/10.1017/bpp.2018.25.

11. T. F. Pettigrew, L. R. Tropp, A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology* **90** (5), 751–783 (2006), doi:10.1037/0022-3514.90.5.751, http://dx.doi.org/10.1037/0022-3514.90.5.751.

12. G. Allport, *The Nature of Prejudice* (Addison-Wesley) (1954).

13. J. R. Brown, R. D. Enos, J. Feigenbaum, S. Mazumder, Childhood cross-ethnic exposure predicts political behavior seven decades later: Evidence from linked administrative data. *Science Advances* **7** (24) (2021), doi:10.1126/sciadv.abe8432, http://dx.doi.org/10.1126/sciadv.abe8432.

14. D. Stolle, A. Harell, Social Capital and Ethno-racial Diversity: Learning to Trust in an Immigrant Society. *Political Studies* **61** (1), 42–66 (2012), doi:10.1111/j.1467-9248.2012.00969.x, http://dx.doi.org/10.1111/j.1467-9248.2012.00969.x.

15. S. K. Goldman, D. J. Hopkins, Past Place, Present Prejudice: The Impact of Adolescent Racial Context on White Racial Attitudes. *The Journal of Politics* **82** (2), 529–542 (2020), doi:10.1086/706461, http://dx.doi.org/10.1086/706461.

16. S. D. Miller, D. O. Sears, Stability and Change in Social Tolerance: A Test of the Persistence Hypothesis. *American Journal of Political Science* **30** (1), 214 (1986), doi:10.2307/2111302, http://dx.doi.org/10.2307/2111302.

17. D. O. Sears, C. Brown, *Childhood and Adult Political Development* (Oxford University Press), p. 65–108 (2023), doi:10.1093/oxfordhb/9780197541302.013.30, http://dx.doi.org/10.1093/oxfordhb/9780197541302.013.30.

18. H. Finseraas, A. Kotsadam, Does personal contact with ethnic minorities affect anti-immigrant sentiments? Evidence from a field experiment. *European Journal of Political Research* **56** (3), 703–722 (2017), doi:10.1111/1475-6765.12199, http://dx.doi.org/10.1111/1475-6765.12199.

19. L. Corno, E. La Ferrara, J. Burns, Interaction, Stereotypes, and Performance: Evidence from South Africa. *American Economic Review* **112** (12), 3848–3875 (2022), doi:10.1257/aer.20181805, http://dx.doi.org/10.1257/aer.20181805.

20. C. Grady, R. Wolfe, D. Dawop, L. Inks, How contact can promote societal change amid conflict: An intergroup contact field experiment in Nigeria. *Proceedings of the National Academy of Sciences* **120** (43) (2023), doi:10.1073/pnas.2304882120, http://dx.doi.org/10.1073/pnas.2304882120.

21. A. Scacco, S. S. Warren, Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria. *American Political Science Review* **112** (3), 654–677 (2018), doi:10.1017/s0003055418000151, http://dx.doi.org/10.1017/S0003055418000151.

22. M. Lowe, Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration. *American Economic Review* **111** (6), 1807–1844 (2021), doi:10.1257/aer.20191780, http://dx.doi.org/10.1257/aer.20191780.

23. F. Elwert, T. Keller, A. Kotsadam, 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 (2023), doi:10.1086/724865, http://dx.doi.org/10.1086/724865.

24. S. E. Carrell, M. Hoekstra, J. E. West, The Impact of College Diversity on Behavior toward Minorities. *American Economic Journal: Economic Policy* **11** (4), 159–182 (2019), doi:10.1257/pol.20170069, http://dx.doi.org/10.1257/pol.20170069.

25. S. Mousa, Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science* **369** (6505), 866–870 (2020), doi:10.1126/science.abb3153, http://dx.doi.org/10.1126/science.abb3153.

26. G. B. Dahl, A. Kotsadam, D.-O. Rooth, Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams. *The Quarterly Journal of Economics* **136** (2), 987–1030 (2020), doi:10.1093/qje/qjaa047, http://dx.doi.org/10.1093/qje/qjaa047.

27. M. Lowe, Has Intergroup Contact Delivered? *Annual Review of Economics* **17** (1), 321–344 (2025), doi:10.1146/annurev-economics-081324-091109, http://dx.doi.org/10.1146/annurev-economics-081324-091109.

28. A. Kalmbach, M. Sarvimäki, L. Wren-Lewis, The Impact of Childhood Immigrant Exposure on Adult Intermarriage (2022), working paper.

29. H. Holmlund, E. Lindahl, S. Roman, Immigrant peers in the class: Effects on natives' long-run revealed preferences. *Labour Economics* **82**, 102360 (2023).

30. L. P. Merlino, M. F. Steinhardt, L. Wren-Lewis, More than just friends? School peers and adult interracial relationships. *Journal of Labor Economics* **37** (3), 663–713 (2019).

31. M. Abascal, Us and Them: Black-White Relations in the Wake of Hispanic Population Growth. *American Sociological Review* **80** (4), 789–813 (2015), doi:10.1177/0003122415587313, http://dx.doi.org/10.1177/0003122415587313.

32. V. Fouka, S. Mazumder, M. Tabellini, From immigrants to Americans: Race and assimilation during the Great Migration. *The Review of Economic Studies* **89** (2), 811–842 (2022).

33. L. R. Tropp, T. F. Pettigrew, Relationships Between Intergroup Contact and Prejudice Among Minority and Majority Status Groups. *Psychological Science* **16** (12), 951–957 (2005), doi:10.1111/j.1467-9280.2005.01643.x, http://dx.doi.org/10.1111/j.1467-9280.2005.01643.x.

34. C. Gibson, K. Jung, *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*, Population Division Working Paper 81, U.S. Census Bureau, Washington, DC (2006).

35. R. Kazal, *Becoming Old Stock: The Paradox of German-American Identity* (Princeton University Press) (2021).

36. M. M. Gordon, *Assimilation in American Life: The Role of Race, Religion, and National Origins* (Oxford University Press) (1964), OCLC: 133168206.

37. M. Kalmijn, Intermarriage and homogamy: Causes, patterns, trends. *Annual Review of Sociology* **24** (1), 395–421 (1998).

38. J. Higham, *Strangers in the Land: Patterns of American Nativism, 1860-1925* (Rutgers University Press) (2002).

39. T. D. Logan, J. M. Parman, The National Rise in Residential Segregation. *The Journal of Economic History* **77** (1), 127–170 (2017), doi:10.1017/s0022050717000079, `http://dx.doi.org/10.1017/S0022050717000079`.

40. S. Ruggles, *et al.*, IPUMS Ancestry Full Count Data: Version 4.0 (2024), doi:10.18128/D014.V4.0, `https://usa.ipums.org/usa/full_count.shtml`.

41. S. Ruggles, *et al.*, IPUMS USA: Version 15.0 (2024), doi:10.18128/D010.V15.0, `https://www.ipums.org/projects/ipums-usa/d010.V15.0`.

42. R. Maxwell, Cosmopolitan immigration attitudes in large European cities: Contextual or compositional effects? *American Political Science Review* **113** (2), 456–474 (2019).

43. A. Acharya, M. Blackwell, M. Sen, The political legacy of American slavery. *The Journal of Politics* **78** (3), 621–641 (2016).

44. D. Schindler, M. Westcott, Shocking racial attitudes: Black GIs in Europe. *The Review of Economic Studies* **88** (1), 489–520 (2021).

45. J. R. Logan, J. Jindrich, H. Shin, W. Zhang, Mapping America in 1880: The Urban Transition Historical GIS Project. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* **44** (1), 49–60 (2011), doi:10.1080/01615440.2010.517509, `http://dx.doi.org/10.1080/01615440.2010.517509`.

46. K. Ostermann, Beyond Proximity: Constructing Organic Neighborhoods Using a Two-Stage Unsupervised Learning Approach. *SocArXiv* **https://doi.org/10.31235/osf.io/xpmkr** (2024).

47. S. M. Iacus, G. King, G. Porro, Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* **20** (1), 1–24 (2012), doi:10.1093/pan/mpr013, `http://dx.doi.org/10.1093/pan/mpr013`.

48. U. Simonsohn, J. P. Simmons, L. D. Nelson, Specification curve analysis. *Nature Human Behaviour* **4** (11), 1208–1214 (2020).

49. T. J. VanderWeele, P. Ding, Sensitivity Analysis in Observational Research: Introducing the E-Value. *Annals of Internal Medicine* **167** (4), 268–274 (2017), doi:10.7326/m16-2607, `http://dx.doi.org/10.7326/M16-2607`.

50. S. T. Fiske, C. H. Dupree, G. Nicolas, J. K. Swencionis, Status, power, and intergroup relations: the personal is the societal. *Current Opinion in Psychology* **11**, 44–48 (2016), doi:10.1016/j.copsyc.2016.05.012, `http://dx.doi.org/10.1016/j.copsyc.2016.05.012`.

51. S. Bubritzki, F. van Tubergen, J. Weesie, S. Smith, Ethnic composition of the school class and interethnic attitudes: a multi-group perspective. *Journal of Ethnic and Migration Studies* **44** (3), 482–502 (2018).

52. S. Smith, D. A. McFarland, F. Van Tubergen, I. Maas, Ethnic Composition and Friendship Segregation: Differential Effects for Adolescent Natives and Immigrants. *American Journal of Sociology* **121** (4), 1223–1272 (2016), doi:10.1086/684032, `http://dx.doi.org/10.1086/684032`.

53. D. Kretschmer, J. Gereke, F. Winter, N. Zhang, No Differential Effects of Classroom Ethnic Composition on Native and Immigrant Friendship Segregation: Comment on Smith et al.,

2016. *American Journal of Sociology* **129** (2), 570–585 (2023), doi:10.1086/727823, http://dx.doi.org/10.1086/727823.

54. N. Egami, E. Hartman, Elements of External Validity: Framework, Design, and Analysis. *American Political Science Review* **117** (3), 1070–1088 (2022), doi:10.1017/s0003055422000880, http://dx.doi.org/10.1017/S0003055422000880.

55. J. Helgertz, *et al.*, IPUMS Multigenerational Longitudinal Panel: Version 1.2 (2023), doi: 10.18128/D016.V1.2, https://usa.ipums.org/usa/mlp/mlp.shtml.

56. J. Helgertz, *et al.*, A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* **55** (1), 12–29 (2021), doi:10.1080/01615440.2021.1985027, http://dx.doi.org/10.1080/01615440.2021.1985027.

57. A. Grigoryeva, M. Ruef, The Historical Demography of Racial Segregation. *American Sociological Review* **80** (4), 814–842 (2015), doi:10.1177/0003122415589170, http://dx.doi.org/10.1177/0003122415589170.

58. U.S. Bureau of the Census, Alphabetical Index of Occupations and Industries: 1950 (1950), prepared by: David L. Kaplan, Claire Casey, and Walter B. Pimm, with assistance from Gladys M. Dodd, Inez B. Lange, and Hettie F. Owens.

59. D. Arthur, S. Vassilvitskii, k-means++: the advantages of careful seeding, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07 (Society for Industrial and Applied Mathematics, USA) (2007), p. 1027–1035.

60. A. J. Onumanyi, D. N. Molokomme, S. J. Isaac, A. M. Abu-Mahfouz, AutoElbow: An automatic elbow detection method for estimating the number of clusters in a dataset. *Applied Sciences* **12** (15), 7515 (2022).

61. L. McInnes, J. Healy, S. Astels, hdbscan: Hierarchical density based clustering. *Journal of Open Source Software* **2** (11), 205 (2017).

62. J. L. Abu-Lughod, *From Urban Village to East Village: The Battle for New York's Lower East Side* (Blackwell) (1994).

63. E. A. Stuart, Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science* **25** (1) (2010), doi:10.1214/09-sts313, http://dx.doi.org/10.1214/09-STS313.

64. P. Ding, T. J. VanderWeele, Sensitivity Analysis Without Assumptions. *Epidemiology* **27** (3), 368–377 (2016), doi:10.1097/ede.0000000000000457, http://dx.doi.org/10.1097/EDE.0000000000000457.

Acknowledgments

We also gratefully acknowledge helpful feedback from Ozan Aksoy, Ala Alrababah, Julius Diener, Per Engzell, Mary Waters, Joscha Legewie, seminar participants at the 2024 GESS research day, the Mannheim CDSS workshop, and conference participants at the 8th International Meeting on Experimental and Behavioral Social Sciences (IMEBESS) in Riga, the 5th International Sociological Association (ISA) Forum of Sociology in Rabat, the International Conference of Computational Social Science (IC2S2) in Norrköpping, and the 2025 European Consortium for Sociological Research (ECSR) Conference in Cologne. The authors acknowledge support by the state of Baden-Württemberg through bwHPC.

Funding: This work was supported by the German Research Foundation (DFG)'s Emmy Noether Programme (Projektnummer 437007381), and the joint PhD Scholarship Program of the Institute for Employment Research (IAB) and the Friedrich-Alexander University Erlangen-Nuremberg.

Author contributions: Designed research: LW, NZ; Contributed methods and materials: LW, KO; Analyzed data: LW; Wrote the paper: LW, NZ, KO.

Competing interests: There are no competing interests to declare.

Data and materials availability: Replication code is available from OSF after publication. Historical US census microdata were obtained from IPUMS (*40, 55*). Access to these source datasets is unrestricted but requires registration with IPUMS. For analyses we use PYTHON 3.9, R 4.4.1 and STATA 19.

Supplementary materials

Materials and Methods

Figs. S1 to S20

Tables S1 to S71

References *(7-64)*

Supplementary Materials for
Childhood exposure to ethnic outgroups predicts interethnic marriage

Leonard Wendering , Nan Zhang , Kerstin Ostermann

*Corresponding author. Email: leonard.wendering@uni-mannheim.de

This PDF file includes:

Materials and Methods

Data Sources

The starting point for our analysis is the publicly available 1880 US Census with information on 50,140,482 individuals (*40, 41*). Our analysis excludes individuals living outside of private households.

To apply the machine learning algorithm that identifies neighborhoods (*46*), we need georeferenced household data. We utilize data from 40 cities for which households have been geolocated by (*45*). This results in precise locations for 6,726,553 individuals in 1,520,078 households, covering 96% of all census entries for the 40 cities.

We further restrict our sample to boys aged 0-18 in 1880 who are linked to their respective entries in the 1900/1910 US Census (*41, 55*). Since women are likely to change their names upon marriage, we restrict our attention to men only.

Racial information is incorporated in our analysis in the following way. Non-whites are treated as separate groups in the construction of neighborhoods (see below) and when coding the ethnicity of next-door neighbors. This is because (i) we believe that exposure to non-whites is substantively different than exposure to white Americans (even of differing ethnicities), and (ii) the rates of inter-racial marriage in this period were negligible. For the same reasons, we restrict our attention to the marriage decisions of white Americans (whom we label "American" for convenience) and white immigrants.

We arrive at a linked sample of 428,202 individuals in the 1880-1900 sample (N_{American} = 150,512; N_{German} = 119,438; N_{Irish} = 93,226; N_{Other} = 65,026; overall linkage rate = 31.4%), and 207,438 individuals in the 1880-1910 sample (N_{American} = 75,447; N_{German} = 63,553; N_{Irish} = 38,131; N_{Other} = 30,307; overall linkage rate = 15.2%).

The linking algorithm produces a reported accuracy of 98% (*56*). To compare whether linked individuals vary systematically from the full population before linkage, we calculate the standardized mean differences for an extensive set of demographic variables (Fig. S4 and S5). The linked samples resemble the full population with three exceptions: (i) American-born are overrepresented in both linked samples, and the 1880-1910 linked sample overrepresents (ii) individuals from New York and (iii) laborers.

Variables

The ethnicity of a boy in the analysis is determined by the birthplace of his household head as recorded in 1880. This implies that second-generation immigrants are given the ethnicity of their household head (e.g. a child born in the United States to German-born parents is classified as German). We also use the birthplace of the household head to code the ethnicity of all neighbors in 1880. Neighbors are determined based on the ordering of entries in the census enumeration sheets, reflecting enumerators' practice of collecting data sequentially, household by household and street by street (*13, 39, 57*).

Explanatory Variables. We take two different approaches to operationalize childhood exposure. Firstly, we create COUNTS of the number of American, German, and Irish neighbors among the 20 nearest census entries (10 up and 10 down the enumeration sheet). Summary statistics can be found in Table S53.

Secondly, we employ a Coarsened Exact Matching (CEM) approach where we define TREATMENT and CONTROL observations based on the ethnicity of the two immediate neighbors (1 up and 1 down the enumeration sheet). For instance, $\text{TREAT}_{\text{GERMAN}} = 1$ if one or both immediate neighbors are German, and $\text{TREAT}_{\text{GERMAN}} = 0$ only if both immediate neighbors are from the ingroup. Importantly, we define the CONTROL group to represent cases without "close exposure" to any ethnic outgroups. Thus, in this example, if one neighbor were Italian and the other neighbor were American, $\text{TREAT}_{\text{GERMAN}}$ is set to missing.

Additional Covariates. We utilize additional information from the 1880 census as controls/matching variables. SCHOOL ATTENDANCE is a dichotomous variable indicating whether an individual attended school. OCCUPATION categories of the household head are defined based on the following 1950 Census Bureau classifications (*58*): professional/technical workers, farmers, managers/officials/proprietors, clerics, sales workers, craftsmen, operatives, service workers, farm laborers, laborers, and non-occupational. BIRTHYEAR, NATIVITY, and CITY of residence are taken from the census without change. Finally, we control for / match on 1880 NEIGHBORHOODS (see below).

Outcomes. Marriage outcomes are derived from the boys' spousal data recorded in the 1900 or 1910 censuses. Wives are coded based on their own nativity and that of their parents. Specifically, first-generation immigrant women are coded based on their country of birth. Native-born women are assigned the ethnicity of their parents' birth countries. This allows us separate second-generation immigrants from the native-born of "American stock." We code as second-generation Irish / Germans only native-born women for whom *both parents* were born in Ireland / Germany. Second-generation immigrants with other combinations of parentage are coded as "Other".

Using this information, we create categorical dependent variables denoting (i) marriage to a member of the TARGET OUTGROUP, (ii) the INGROUP, (iii) a THIRD OUTGROUP, (iv) being UNMARRIED (i.e. "single"), or (v) having an OTHER STATUS as divorced or widowed, as well as cases where the identity of the spouse could not be determined among listed household members. Finally, in dyads involving German and Irish men with, respectively, Irish and German neighbors, we also include (vi) marriage to an AMERICAN as an additional outcome. Tables S54 and S55 present the distributions of the dependent variables.

Neighborhoods

We employ a machine-learning algorithm (*46*) to delineate neighborhoods based on geolocation data and demographic characteristics of households. A primary advantage of this approach over non-spatial network methods is its explicit acknowledgment of space by identifying boundaries where the spatial density of social groups shifts (e.g. due to physical barriers like large roads or parks). These units are typically smaller than census tracts (averaging 258 residents) and vary in shape, size, and density. As shown in Fig.1, neighborhoods defined this manner are internally homogeneous but can overlap with one another.

To identify neighborhoods, the algorithm proceeds in two steps. In the first step, we cluster individuals into relevant social groups using the k-means++ algorithm (*59*). Using this algorithm, we create groups of individuals who are most similar on the variables RACE (white, Black or Asian) and BIRTHPLACE (first-generation Americans, second-generation Americans, other North Americans, Central American & Caribbean, South Americans, other Europeans, Irish, Germans, Asians, Africans, Oceanians, Other/Unknown. See Table S56).

We map these variables into an n-dimensional space where centroids represent distinct social groups. K-means++ uses a probabilistic seeding technique to choose initial centroids, ensuring stability and reproducibility of the results. The k-means iterates as follows: each individual is assigned to the nearest centroid based on their characteristics, and centroids are updated as the mean of their assigned individuals. These assignment and update steps repeat until the centroids remain stable, representing our assignment of individuals to social groups. Since the number and type of relevant social groups might vary by city, we run the algorithm separately for each city and use the AutoElbow-algorithm (*60*) to determine the optimal number of social groups. Fig. S7 shows the distribution of number of social groups by city.

Using Manhattan as an illustrative case, Fig. S8 displays the mean values for race, generational status, and parental birthplace across the six optimal groups identified in Manhattan. These profiles reveal distinct social strata: Groups 0, 1, and 4 consist mainly of white foreigners from Germany, Ireland, or other states, respectively, while Group 5 contains mainly Asian foreign-born individuals. Among the US-born population, Group 2 comprises individuals with foreign-born parents, whereas Group 3 consists of those with US-born parents. The spatial distribution of these households is visualized in the left panel of Figure 1.

In the second step, we create polygons for each of the social groups in areas where these groups are prevalent (*46*). Employing a Hierarchical Density-Based Clustering Algorithm with an Application to Noise (HDBSCAN (*61*)), HDBSCAN clusters group members if their location shows a similar local density value. The algorithm relies on two hyperparameters: MINIMUM CLUSTER SIZE defines the smallest number of households required to constitute a valid neighborhood, and MINIMUM SAMPLES determines the number of neighbors used to estimate local density. Higher values of minimum samples lead to smaller outcome groups as observations at the boundaries are not considered. To be conservative, we set minimum samples to 1, ensuring that only substantial density fluctuations form new cluster boundaries. We vary MINIMUM CLUSTER SIZE from 15 to 300 and run the main regression analysis on the matched sample across the full range of cluster sizes, selecting a minimum cluster size of 80 as it minimizes the Akaike Information Criterion (AIC). Results are robust to different hyperparameter settings (see Fig. S6 and Tables S57 to S70).

The neighborhoods in the right panel of Figure 1 illustrate how ethnic groups of different sizes and spatial expansion share or monopolize certain areas of Manhattan. In some areas, neighborhoods follow the block structure, whereas other areas show a more complex pattern. The Lower East Side (at the bottom left of Map B) in particular exhibits larger neighborhoods and more overlaps mirroring earlier ethnographic descriptions of this part of the city (*62*). Germans (yellow), seem to distribute quite evenly across Manhattan as one large yellow polygon covers large parts of the island. In our sample of 40 cities, the algorithm generates roughly 81,700 neighborhoods (polygons).

Finally, we create a variable measuring neighborhood DENSITY. This is defined as the number of households in each polygon divided by its area in square kilometers. Individuals in 1880 are assigned a density value that equals the average of all of the overlapping polygons in which they reside.

Analysis

Our first empirical strategy is to model the conditional probability of marriage outcomes as a function of the COUNT of outgroup neighbors among the 20 closest households in 1880. Specifically, we perform a multinomial logistic regression to estimate the odds of a given marriage outcome k (with $k = 1, \dots, K-1$) relative to a baseline marriage category (indexed as K):

$$\log\left(\frac{\Pr(Y_{i,e} = k \mid \mathbf{X}_i)}{\Pr(Y_{i,e} = K \mid \mathbf{X}_i)}\right) = \beta_{0,k} + \beta_{1,k}\,\text{Count}_{i,e} + X\Gamma + \varepsilon_{i,e}$$

where:

- i indexes individuals belong to ethnic group e
- $\text{Count}_{i,e}$ denotes the number of neighbors belonging to the target outgroup
- X denotes a vector of controls for SCHOOL ATTENDANCE, OCCUPATION of the household head, NATIVITY, BIRTH YEAR, CITY of residence, NEIGHBORHOOD DENSITY, and the COMPOSITION OF SOCIAL GROUPS in the residential surroundings (i.e. the combination of overlapping polygons).
- $\beta_{0,k}$ is the intercept for outcome k, representing the baseline log-odds of being in category k (relative to the baseline category) when $\text{Count}_{i,e} = 0$.
- $\beta_{1,k}$ represents the effect of $\text{Count}_{i,e}$ on the log-odds of being in outcome k relative to the baseline category.

Robust standard errors are clustered at the city level. To prevent quasi-separation between social-group dummies and city fixed effects, we recode any social group that appears in only one city and contains fewer than 100 observations into a pooled ”single-city groups” category before estimation. We prefer this approach to either excluding these observations or omitting social-group or city fixed effects entirely.

After estimating the model, we calculate the predicted probabilities for all outcome categories including the baseline category. The Average Marginal Effect (AME) is the average change in the predicted probability of each marriage outcome for one additional outgroup neighbor. Separate analyses are conducted for each of the six dyads depicted in Fig.2, and for each of the linked samples (1900 and 1910). Results are presented in Tables S3 and S4.

Additionally, we apply Coarsened Exact Matching (CEM) (*47*) to compare marriage outcomes between TREATMENT and CONTROL observations. More specifically, each TREATED boy with at least one suitable counterpart in the CONTROL group becomes part of a matched set. Matching is conducted with replacement, and matched boys in the CONTROL group receive weights based on the number and size of the matched sets to which they are assigned. We use the above-mentioned covariates as matching variables (with neighborhood density coarsened using cutpoints at 50, 100, 250, 750, 1500, and 10,000 individuals per km^2). As before, we conduct the matching analysis separately for each dyad, and for each of the linked samples (1900 and 1910). Details on the number of observations in each matched sample can be found in Table S71.

Figs. S9–S20 display covariate balance before and after CEM using standardized mean differences (*63*). We do not observe imbalance (std. diff. > 0.1) in any of the matched samples.

We then estimate a multinomial logistic regression for each of the twelve matched samples, but now with TREATMENT (instead of COUNT) as our main explanatory variable. After estimating the model, we calculate the predicted probabilities for all outcome categories, including the baseline. The AMEs represent how the predicted probabilities for each outcome change as we move from the TREATMENT to the CONTROL group. Full estimation results are presented in Tables S1 and S2.

Sensitivity Analysis

We employ the E-value framework to quantify the minimum strength of association that an unmeasured confounder would need to have with both the treatment (exposure to outgroup neighbors) and the outcome (marriage) to explain away the observed treatment effect (*49*). The E-value starts from 1, which indicates that even a trivial confounder could overturn the estimated effect, while larger values imply that a confounder must induce severe bias to explain away the observed results.

To arrive at an E-value, we first calculate risk ratios (RR) using the predicted probabilities of the outcome for the control group and the treatment group:

$$RR = \frac{Pr(Y \mid T = 1)}{Pr(Y \mid T = 0)}$$

where $Pr(Y)$ represents the probability for a specific marriage category in our multinomial regression.

Under the null hypothesis of no association, there is no "true" treatment effect, implying a risk ratio of 1 ($RR_{\text{true}} = 1$). Thus, any observed non-null relationship between treatment and outcome (i.e. $RR_{\text{observed}} \neq 1$) must be due entirely to bias. In other words, $RR_{\text{true}} = RR_{\text{observed}}/\text{Bias}$ under the null.

Further, consider an unmeasured confounder C. We can define two sensitivity parameters:

$$RR_{CY} = \frac{Pr(Y \mid C = 1)}{Pr(Y \mid C = 0)}$$

where RR_{CY} measures how much the confounder increases the risk of the outcome; and

$$RR_{TC} = \frac{Pr(C \mid T = 1)}{Pr(C \mid T = 0)}$$

where RR_{TC} measures how much more common the confounder is in the treatment vs. control group.

Ding and VanderWeele (*64*) demonstrate that the maximum bias a confounder can possibly induce is:

$$\text{Bias}_{\text{max}} = \frac{RR_{TC} \times RR_{CY}}{RR_{TC} + RR_{CY} - 1}$$

Under the null hypothesis, we can then substitute RR_{observed} for Bias_{max}.

Finally, the E-value is defined for the "worst-case" scenario where the confounder is equally associated with both the treatment and the outcome ($RR_{\text{CY}} = RR_{\text{TC}} = \text{E-value}$). This allows us to solve for the strength of the confounder necessary to produce a bias equal to the observed RR:

$$\text{E-value} = RR_{\text{observed}} + \sqrt{RR_{\text{observed}}(RR_{\text{observed}} - 1)}$$

We also calculate the E-value for the limit of the 95% confidence interval closest to zero. This provides a conservative estimate of the minimum confounding strength required to turn the result statistically insignificant.

To benchmark these calculations, we compare the E-values to the Risk Ratios from the associations between measured covariates X and our independent and dependent variables. If all observed covariates have RRs below the E-value, an unobserved confounder would need to be stronger than any measured variable to nullify the effect. These benchmarks are reported in tables S5 to S52. RR_{XY} captures the association between a covariate and the outcome: it shows how much more likely the outcome (marrying a neighbor-group or own-group member) is when the boy falls into one category of a control variable versus another. For example, a RR_{XY} value of 2.97 in Table S5 indicates that American boys whose household head worked as a laborer in 1880 were almost three times more likely to marry an outgroup member compared to boys whose household head was a professional. Similarly, RR_{TX} refers to the association between a covariate and the treatment.

AI-Usage

AI models (OpenAI GPT-4o/o1/o3/5/5.1, Anthropic Claude 3.5/4/4.5 Sonnet and 4.5 Opus, Google Gemini 2.5/3 Pro and Cursor IDE) were used to assist with writing code in the R and Stata analysis scripts. Prompts were used for code generation, debugging, and optimizing script efficiency. All AI-assisted code was reviewed.

Supplementary Figures

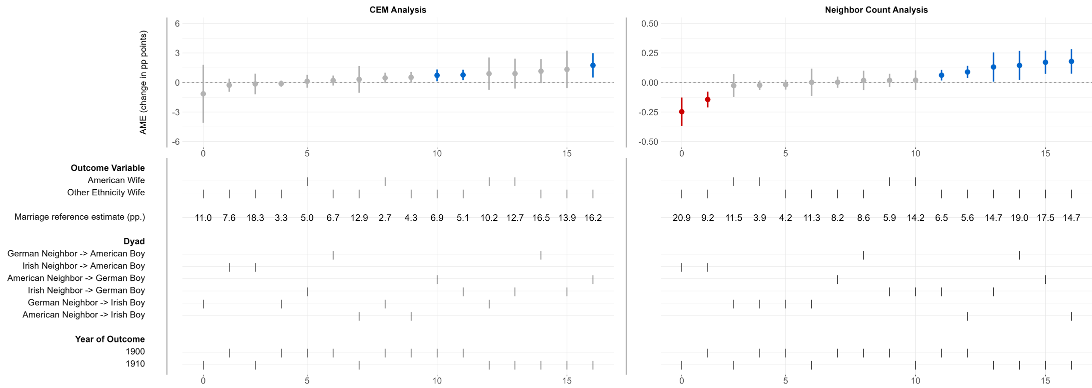

Figure S1: **Specification Curve of Predicted Probabilities of Marrying a Third Ethnic Group.** The upper panels display point estimates and 95% confidence intervals from separate models (indicated in the bottom panel). The left panel presents results from the CEM analysis, while the right panel utilizes the NEIGHBOR COUNT variable. Reference estimates indicate the predicted probability of the outcome among CONTROL observations in the left panel, and the unconditional sample mean in the right panel. AMEs (Average Marginal Effects) in the left panel refer to the difference in predicted probabilities between TREATED VS. CONTROL observations. AMEs in the right panel denote the change in predicted probabilities resulting from every additional outgroup household among the 20 closest neighbors.

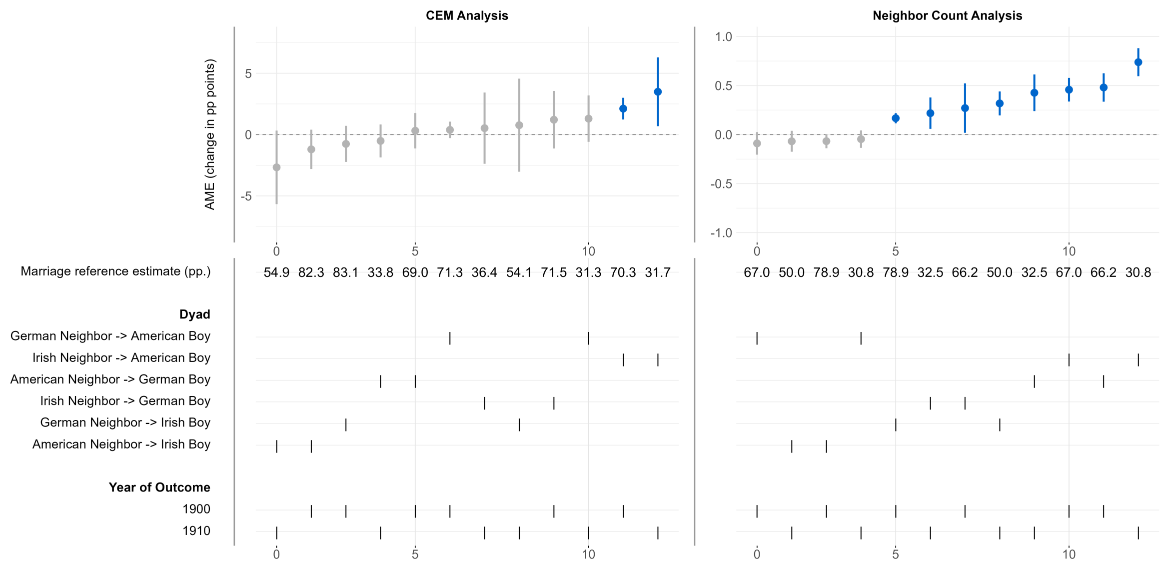

Figure S2: **Specification Curve of Predicted Probabilities of Not Being Married.** The upper panels display point estimates and 95% confidence intervals from separate models (indicated in the bottom panel). The left panel presents results from the CEM analysis, while the right panel utilizes the Neighbor Count variable. Reference estimates indicate the predicted probability of the outcome among control observations in the left panel, and the unconditional sample mean in the right panel. AMEs (Average Marginal Effects) in the left panel refer to the difference in predicted probabilities between treated vs. control observations. AMEs in the right panel denote the change in predicted probabilities resulting from every additional outgroup household among the 20 closest neighbors.

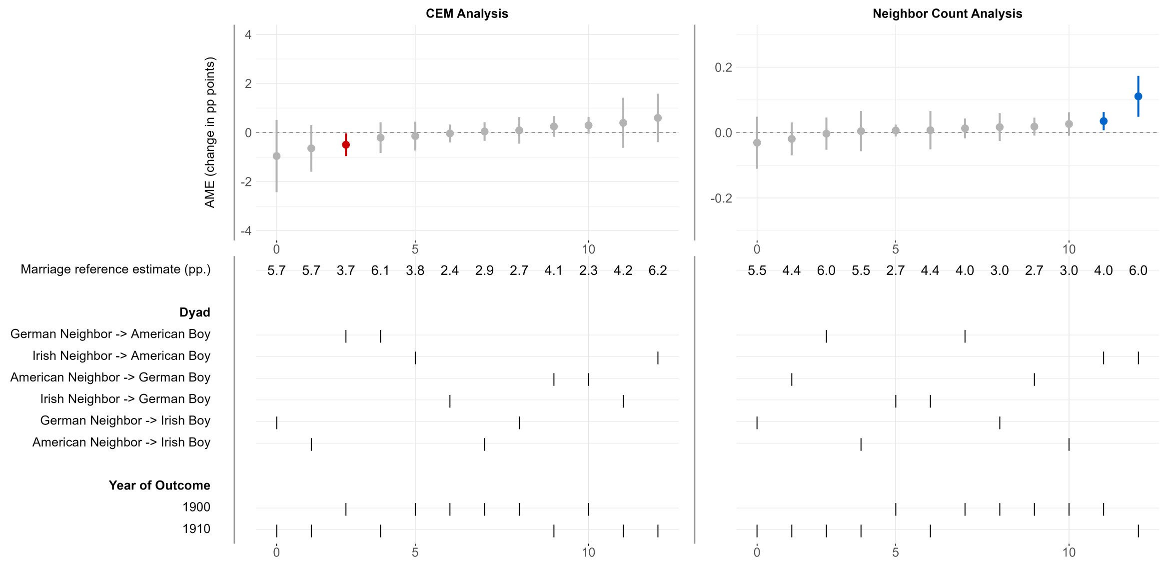

Figure S3: **Specification Curve of Predicted Probabilities of having an Other Marriage Outcome (Spouse Absent, Divorced, Widowed).** The upper panels display point estimates and 95% confidence intervals from separate models (indicated in the bottom panel). The left panel presents results from the CEM analysis, while the right panel utilizes the NEIGHBOR COUNT variable. Reference estimates indicate the predicted probability of the outcome among CONTROL observations in the left panel, and the unconditional sample mean in the right panel. AMEs (Average Marginal Effects) in the left panel refer to the difference in predicted probabilities between TREATED VS. CONTROL observations. AMEs in the right panel denote the change in predicted probabilities resulting from every additional outgroup household among the 20 closest neighbors.

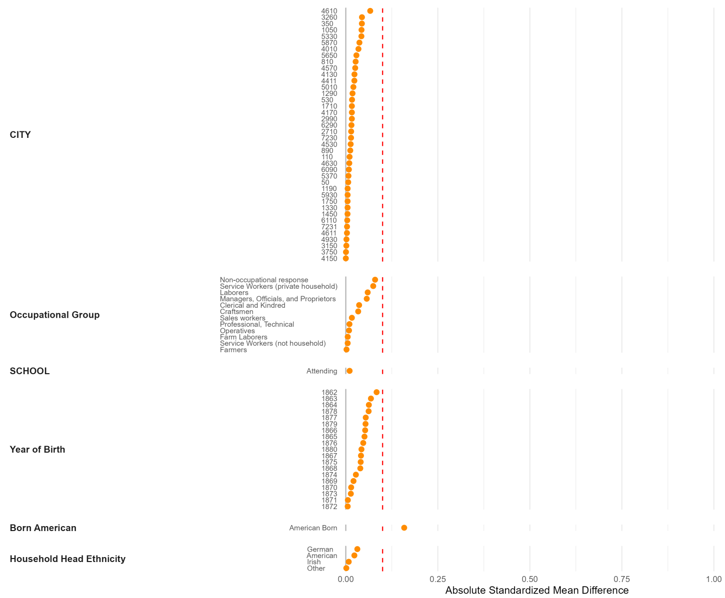

Figure S4: **Covariate balance of the samples before and after linkage (1880-1900).** Each point represents the absolute standardized difference between the linked sample and the full population before linkage. The dashed red line indicates the 0.1 threshold for imbalance.

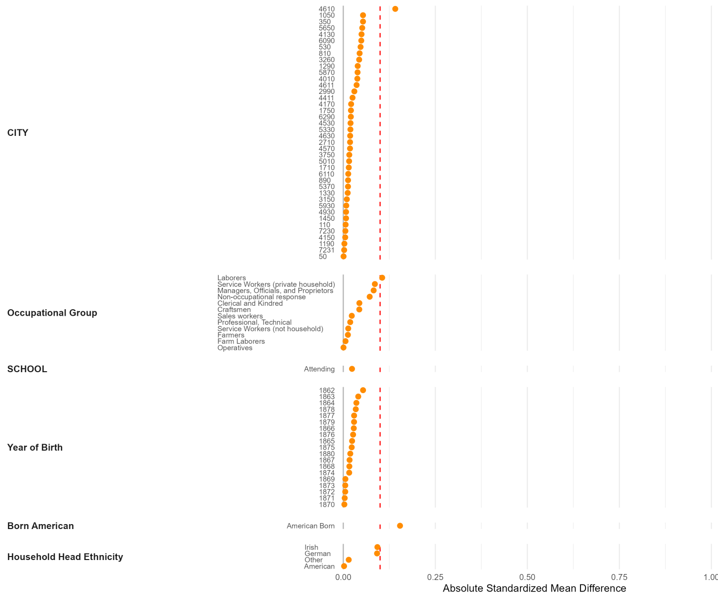

Figure S5: **Covariate balance of the samples before and after linkage (1880-1910).** Each point represents the absolute standardized difference between the linked sample and the full population before linkage. The dashed red line indicates the 0.1 threshold for imbalance.

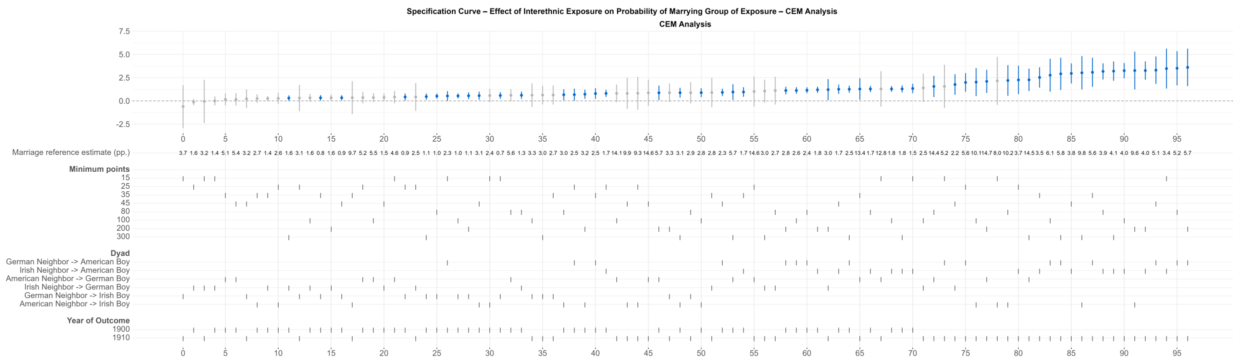

Figure S6: **Specification Curve of Predicted Probabilities of Marrying the Neighboring Ethnic Group by Minimum Cluster Size.** Each point in the upper panels refers to an estimate (and 95% confidence intervals) from a model using the CEM-matched sample. Reference estimates indicate the predicted probability of the outcome among CONTROL observations. AMEs (Average Marginal Effects) refer to the difference in predicted probabilities between TREATED VS. CONTROL observations. The marks in the bottom panel indicate which modeling choices are basis for the estimate. MINIMUM CLUSTER SIZE refers to the minimum number of households used by the neighborhood algorithm to identify a neighborhood.

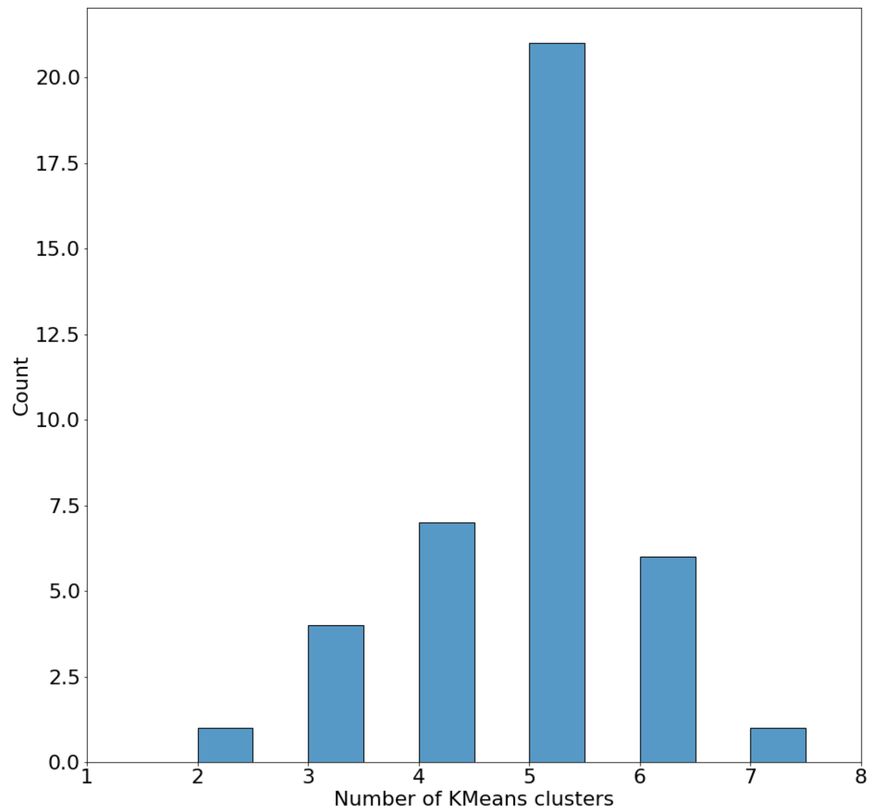

Figure S7: **Distribution of KMeans groups across the all cities.** The figure displays the distribution of the optimal cluster number for the algorithm's first stage. The five cluster solution is most prevalent.

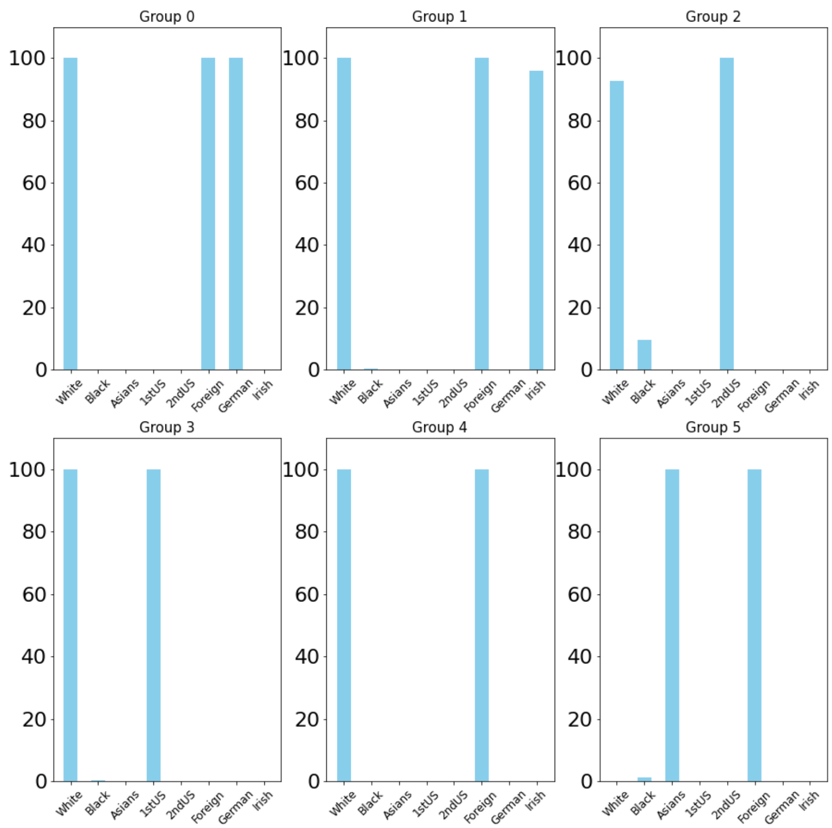

Figure S8: **Group descriptions for organic neighborhoods in Manhattan.** The figure displays six graphs each describing the algorithm's first stage group compositions. The y-axis depicts the share of persons exhibiting the respective characteristics in the group.

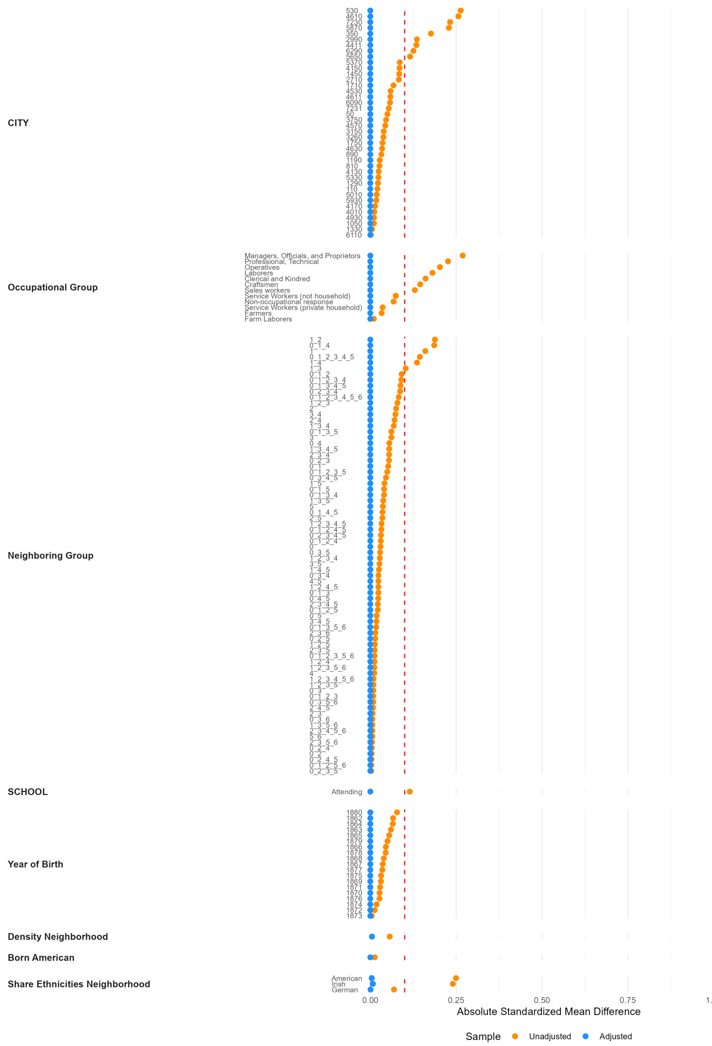

Figure S9: **Covariate balance before and after CEM for American Boys with Irish Neighbors (1880-1900).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

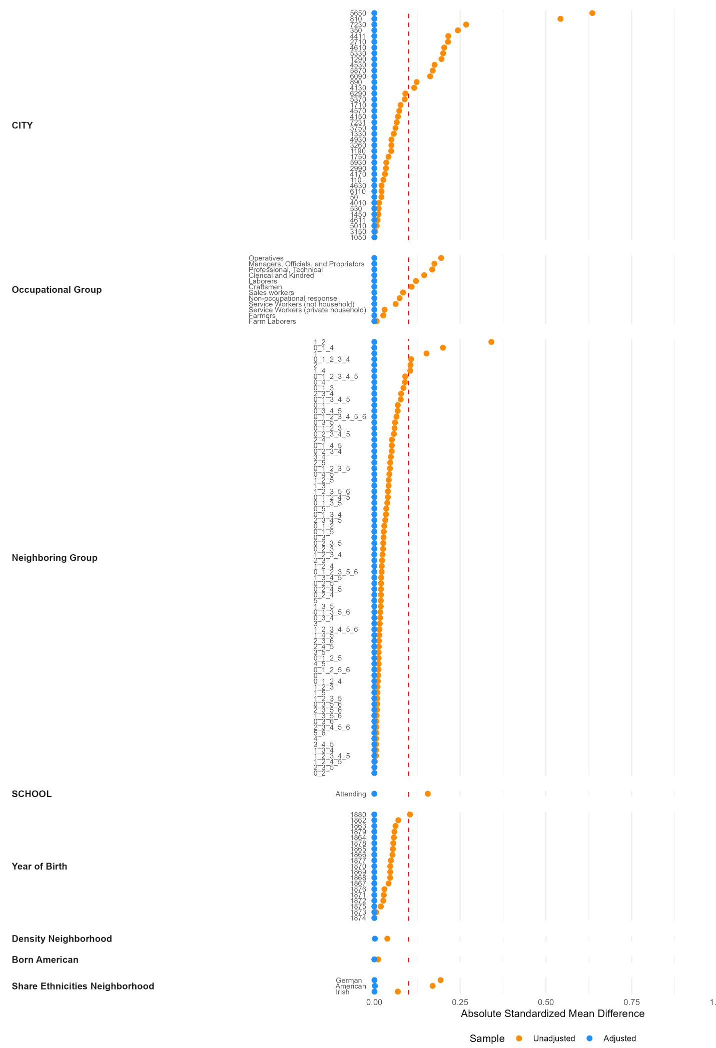

Figure S10: **Covariate balance before and after CEM for American Boys with German Neighbors (1880-1900).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

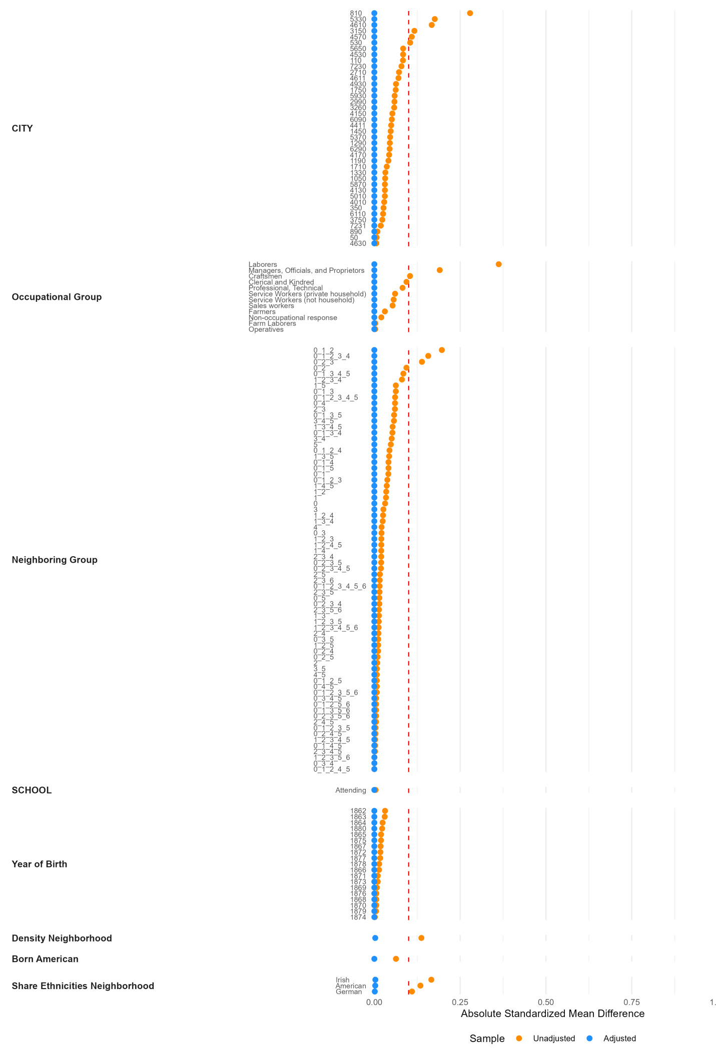

Figure S11: **Covariate balance before and after CEM for Irish Boys with American Neighbors (1880-1900).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

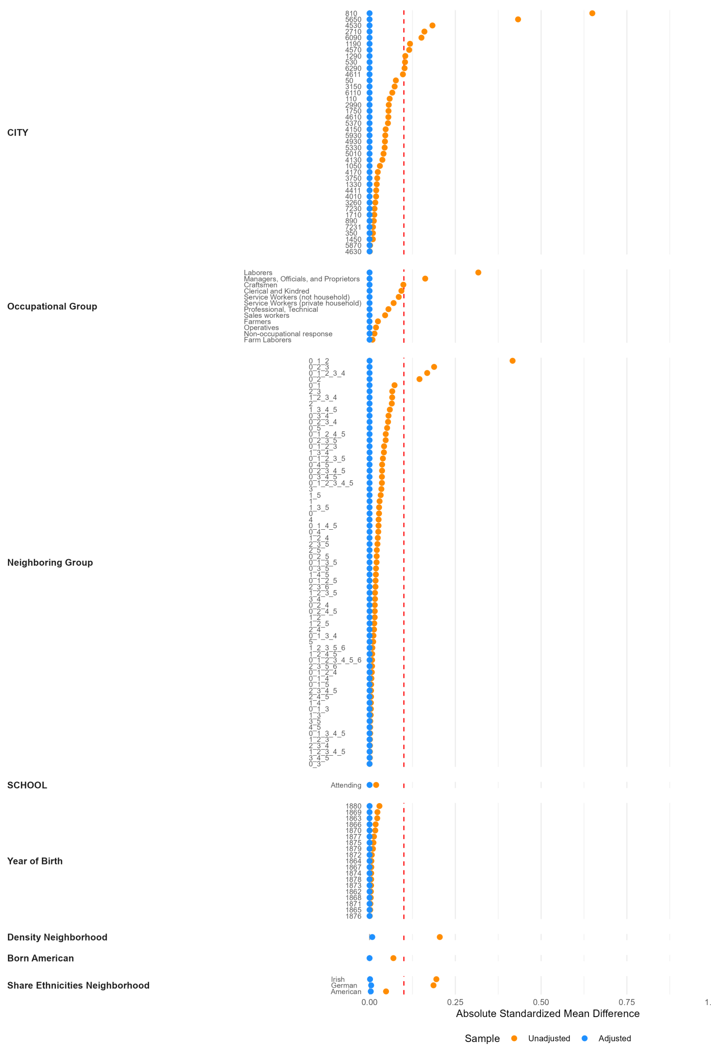

Figure S12: **Covariate balance before and after CEM for Irish Boys with German Neighbors (1880-1900).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

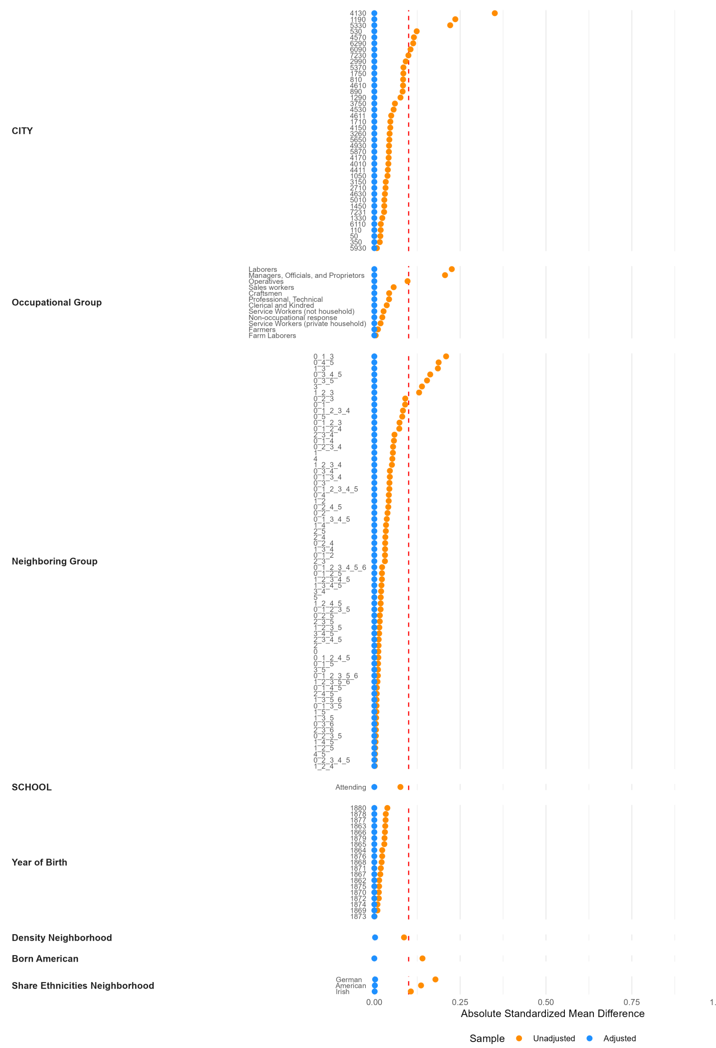

Figure S13: **Covariate balance before and after CEM for German Boys with American Neighbors (1880-1900).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

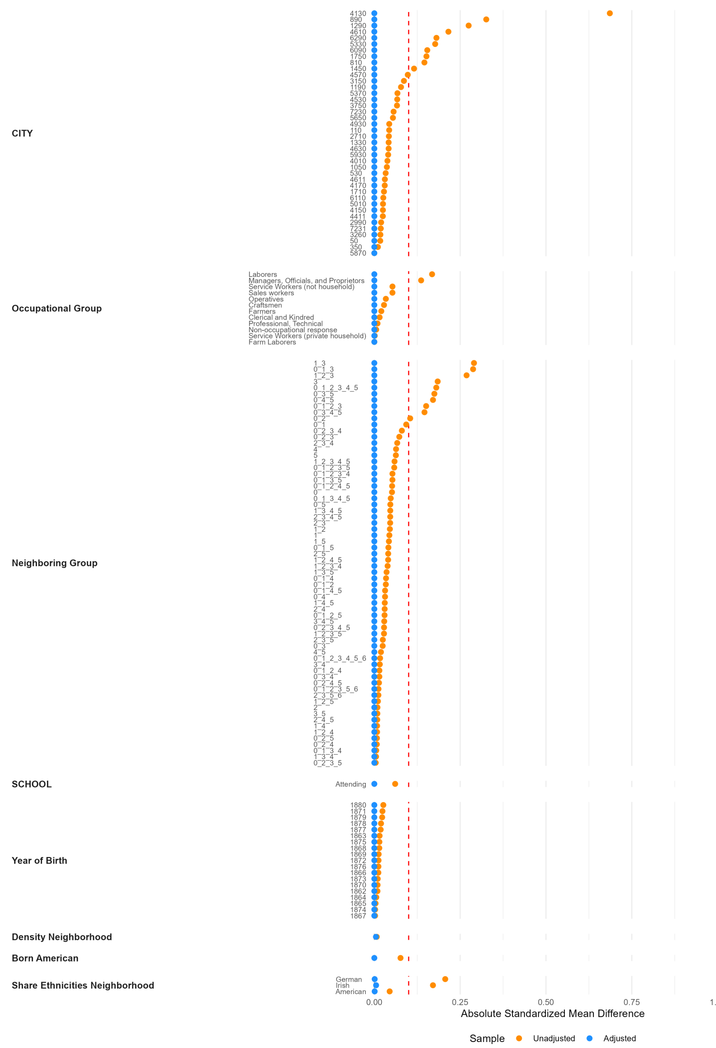

Figure S14: **Covariate balance before and after CEM for German Boys with Irish Neighbors (1880-1900).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

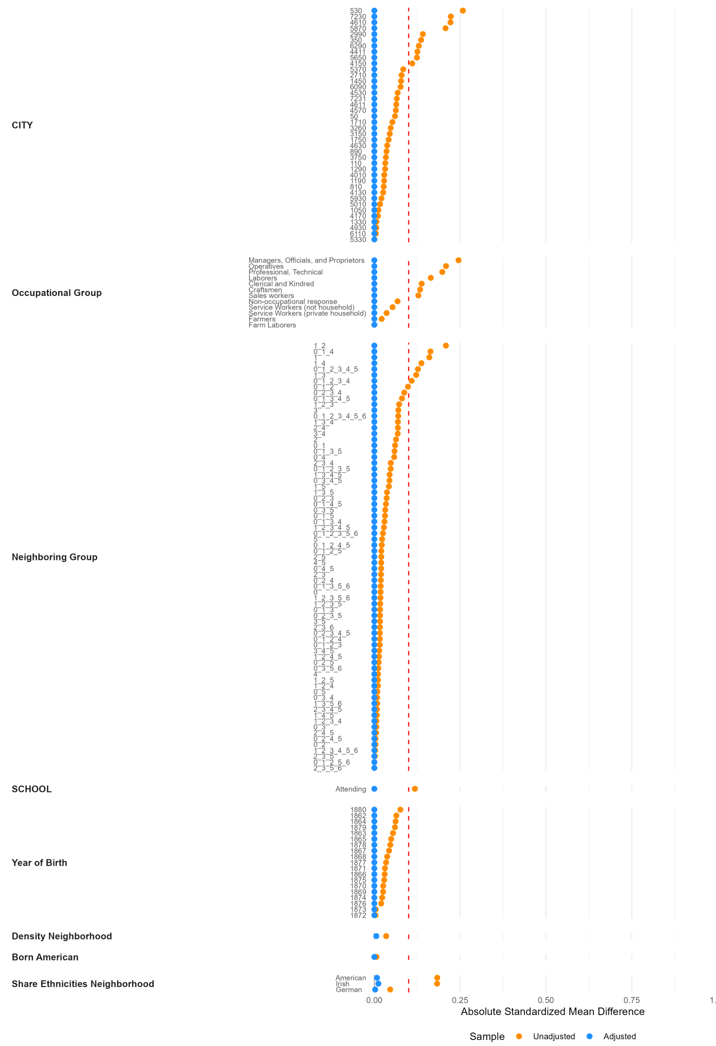

Figure S15: **Covariate balance before and after CEM for American Boys with Irish Neighbors (1880-1910).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

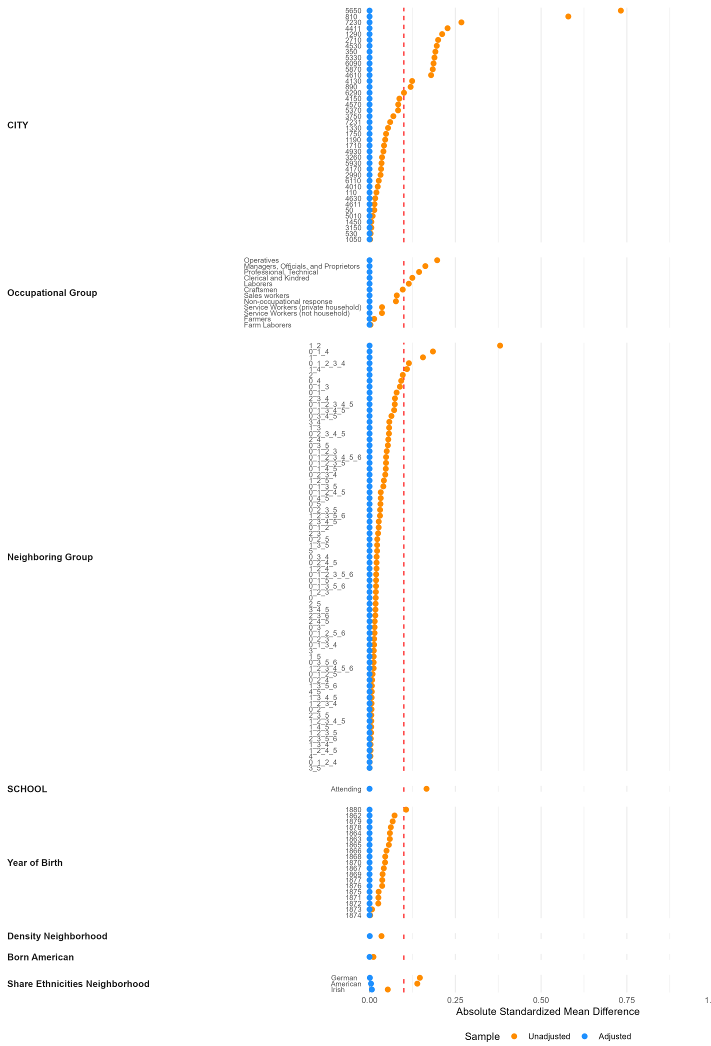

Figure S16: **Covariate balance before and after CEM for American Boys with German Neighbors (1880-1910).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

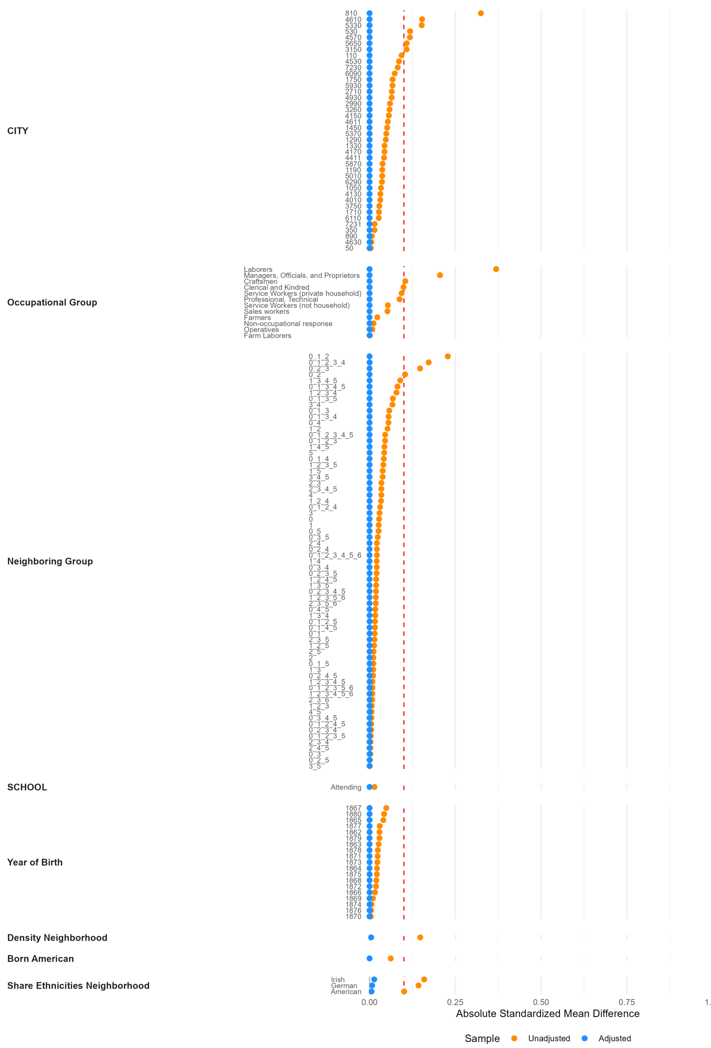

Figure S17: **Covariate balance before and after CEM for Irish Boys with American Neighbors (1880-1910).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

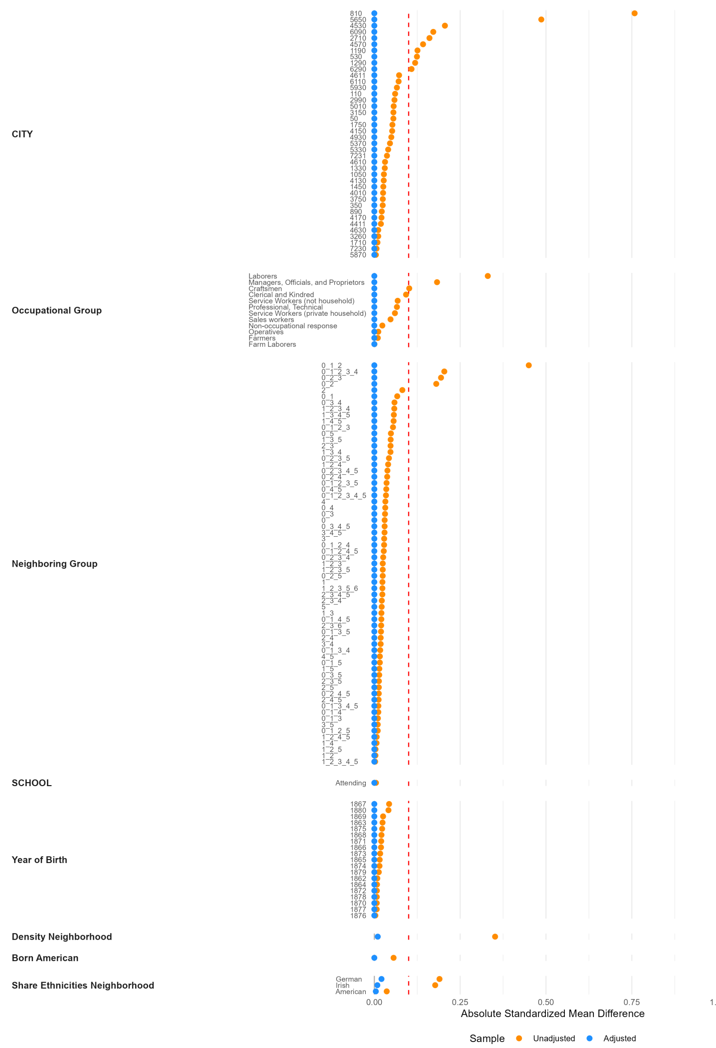

Figure S18: **Covariate balance before and after CEM for Irish Boys with German Neighbors (1880-1910).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

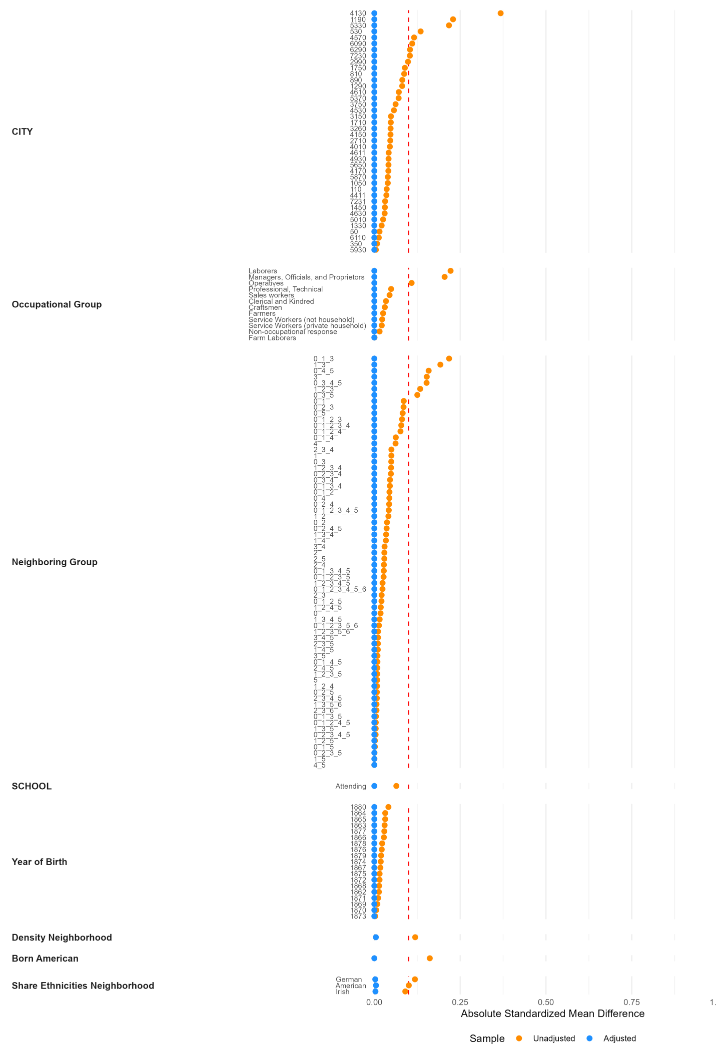

Figure S19: **Covariate balance before and after CEM for German Boys with American Neighbors (1880-1910).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

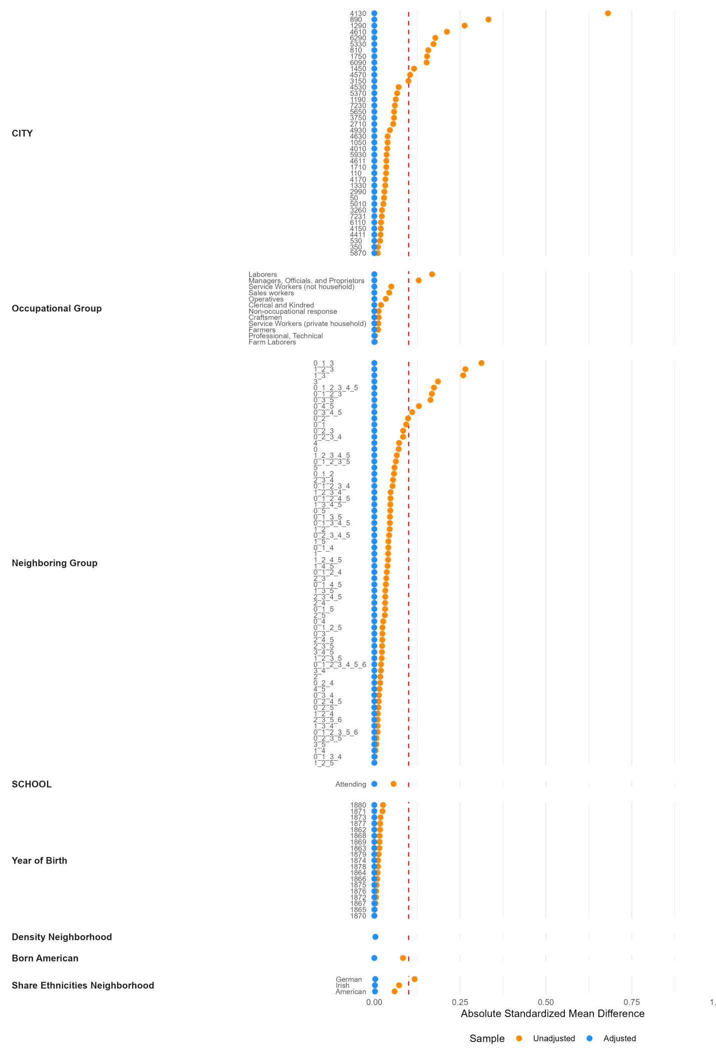

Figure S20: **Covariate balance before and after CEM for German Boys with Irish Neighbors (1880-1910).** Each point represents the absolute standardized mean difference for a covariate comparing treatment and control groups. Orange points show balance in the unmatched sample, blue points show balance after matching. The dashed red line indicates the 0.1 threshold for imbalance.

Supplementary Tables

Table S1: **Predicted Probabilities CEM Analysis (Year of Outcome: 1900).** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least one target out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	1.74	2.44	2.96	0.99	5.62	1.33
ethnic group	AME	+1.28***	+1.15***	+0.66*	+0.51***	+0.61	+0.61**
	(SE)	(0.17)	(0.16)	(0.30)	(0.14)	(0.33)	(0.20)
	p-value	<.001	<.001	0.028	<.001	0.069	0.003
Wife own	Ref.	16.55	15.84	7.48	7.16	16.16	14.59
ethnic group	AME	-2.98***	-1.24***	-0.02	-0.19	-1.94***	-2.67**
	(SE)	(0.35)	(0.35)	(0.26)	(0.34)	(0.48)	(0.85)
	p-value	<.001	<.001	0.938	0.574	<.001	0.002
Wife other group	Ref.	7.59	6.74	4.34	3.33	6.86	5.13
	AME	-0.28	+0.20	+0.52	-0.12	+0.72*	+0.77**
	(SE)	(0.34)	(0.26)	(0.28)	(0.16)	(0.31)	(0.27)
	p-value	0.418	0.433	0.061	0.464	0.018	0.005
American wife	Ref.	-	-	-	2.70	-	5.03
	AME	-	-	-	+0.46	-	+0.12
	(SE)	-	-	-	(0.27)	-	(0.33)
	p-value	-	-	-	0.090	-	0.711
Other outcomes	Ref.	3.85	3.71	2.95	2.74	2.34	2.44
	AME	-0.14	-0.50*	+0.04	+0.09	+0.29	-0.04
	(SE)	(0.30)	(0.24)	(0.19)	(0.28)	(0.17)	(0.19)
	p-value	0.630	0.037	0.821	0.741	0.087	0.850
Not married	Ref.	70.26	71.26	82.28	83.08	69.02	71.48
	AME	+2.12***	+0.38	-1.21	-0.76	+0.31	+1.21
	(SE)	(0.45)	(0.34)	(0.82)	(0.75)	(0.74)	(1.20)
	p-value	<.001	0.263	0.141	0.311	0.669	0.311
n		44,698	43,702	33,265	20,233	45,434	24,608

Table S2: **Predicted Probabilities CEM Analysis (Year of Outcome: 1910).** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least one target out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	3.90	5.24	10.18	2.86	14.40	2.65
ethnic group	AME	+3.17***	+3.51***	+2.20**	+0.88	+1.55**	+1.10
	(SE)	(0.44)	(0.94)	(0.84)	(0.58)	(0.58)	(0.76)
	p-value	<.001	<.001	0.009	0.130	0.008	0.148
Wife own	Ref.	39.86	40.84	16.33	16.12	31.55	30.12
ethnic group	AME	-7.11***	-5.75***	+0.81	-0.43	-3.03**	-4.26**
	(SE)	(1.45)	(1.20)	(1.30)	(1.70)	(0.98)	(1.41)
	p-value	<.001	<.001	0.536	0.801	0.002	0.002
Wife other group	Ref.	18.29	16.46	12.86	11.01	16.15	13.93
	AME	-0.15	+1.15	+0.31	-1.15	+1.74**	+1.33
	(SE)	(0.54)	(0.62)	(0.69)	(1.50)	(0.63)	(0.97)
	p-value	0.780	0.065	0.649	0.444	0.006	0.172
American wife	Ref.	-	-	-	10.22	-	12.70
	AME	-	-	-	+0.89	-	+0.91
	(SE)	-	-	-	(0.84)	-	(0.77)
	p-value	-	-	-	0.288	-	0.242
Other outcomes	Ref.	6.23	6.13	5.74	5.74	4.09	4.16
	AME	+0.60	-0.21	-0.64	-0.96	+0.25	+0.40
	(SE)	(0.50)	(0.32)	(0.49)	(0.75)	(0.22)	(0.52)
	p-value	0.235	0.519	0.187	0.203	0.246	0.444
Not married	Ref.	31.73	31.34	54.89	54.06	33.80	36.44
	AME	+3.49*	+1.30	-2.67	+0.77	-0.52	+0.52
	(SE)	(1.43)	(0.97)	(1.53)	(1.94)	(0.69)	(1.48)
	p-value	0.015	0.178	0.081	0.693	0.452	0.724
n		16,625	16,910	9,127	5,149	19,485	9,649

Table S3: **Predicted Probabilities Count Analysis (Year of Outcome: 1900).** Each column displays the results of a separate multinomial regression for each dyad. "Ref." notes the unconditional sample mean of the outcome. AME is the percentage-point change in that probability for every additional target household among the 20 closest neighbors. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of boys resulting from the linked sample.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	2.25	2.88	3.88	1.43	5.93	1.69
ethnic group	AME	+0.20***	+0.30***	+0.18***	+0.09***	+0.15***	+0.14***
	(SE)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
	p-value	<.001	<.001	<.001	<.001	<.001	<.001
Wife own	Ref.	17.51	17.51	8.56	8.56	17.01	17.01
ethnic group	AME	-0.55***	-0.24**	-0.23***	-0.23***	-0.66***	-0.50***
	(SE)	(0.06)	(0.08)	(0.03)	(0.03)	(0.06)	(0.12)
	p-value	<.001	0.002	<.001	<.001	<.001	<.001
Wife other group	Ref.	9.20	8.57	5.61	4.18	8.17	6.49
	AME	-0.14***	+0.02	+0.09***	-0.02	+0.00	+0.06**
	(SE)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)
	p-value	<.001	0.677	<.001	0.379	0.912	0.007
American wife	Ref.	-	-	-	3.88	-	5.93
	AME	-	-	-	-0.02	-	+0.02
	(SE)	-	-	-	(0.02)	-	(0.03)
	p-value	-	-	-	0.249	-	0.535
Other outcomes	Ref.	3.99	3.99	3.05	3.05	2.71	2.71
	AME	+0.04*	+0.01	+0.03	+0.02	+0.02	+0.01
	(SE)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
	p-value	0.013	0.407	0.152	0.444	0.188	0.495
Not married	Ref.	67.04	67.04	78.91	78.91	66.18	66.18
	AME	+0.46***	-0.09	-0.07	+0.17***	+0.48***	+0.27*
	(SE)	(0.06)	(0.06)	(0.04)	(0.03)	(0.07)	(0.13)
	p-value	<.001	0.126	0.067	<.001	<.001	0.036
n		144,721	144,721	90,442	90,442	116,177	116,177

Table S4: **Predicted Probabilities Count Analysis (Year of Outcome: 1910).** Each column displays the results of a separate multinomial regression for each dyad. "Ref." notes the unconditional sample mean of the outcome. AME is the percentage-point change in that probability for every additional target household among the 20 closest neighbors. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of boys resulting from the linked sample.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	4.40	6.30	11.47	3.42	14.20	2.84
ethnic group	AME	+0.43***	+0.58***	+0.39***	+0.21***	+0.42***	+0.24***
	(SE)	(0.02)	(0.02)	(0.05)	(0.03)	(0.06)	(0.02)
	p-value	<.001	<.001	<.001	<.001	<.001	<.001
Wife own	Ref.	37.81	37.81	18.26	18.26	31.34	31.34
ethnic group	AME	-1.03***	-0.67***	-0.51***	-0.47***	-1.00***	-0.62***
	(SE)	(0.11)	(0.08)	(0.06)	(0.07)	(0.11)	(0.12)
	p-value	<.001	<.001	<.001	<.001	<.001	<.001
Wife other group	Ref.	20.92	19.02	14.73	11.31	17.53	14.69
	AME	-0.25***	+0.14*	+0.18***	+0.00	+0.17***	+0.13*
	(SE)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)
	p-value	<.001	0.021	<.001	0.984	<.001	0.037
American wife	Ref.	-	-	-	11.47	-	14.20
	AME	-	-	-	-0.03	-	+0.02
	(SE)	-	-	-	(0.05)	-	(0.04)
	p-value	-	-	-	0.593	-	0.644
Other outcomes	Ref.	6.05	6.05	5.50	5.50	4.39	4.39
	AME	+0.11***	-0.00	+0.00	-0.03	-0.02	+0.01
	(SE)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
	p-value	<.001	0.898	0.889	0.447	0.454	0.807
Not married	Ref.	30.82	30.82	50.05	50.05	32.54	32.54
	AME	+0.74***	-0.05	-0.07	+0.32***	+0.43***	+0.22**
	(SE)	(0.07)	(0.05)	(0.05)	(0.06)	(0.10)	(0.08)
	p-value	<.001	0.308	0.208	<.001	<.001	0.008
n		72,417	72,417	36,928	36,928	61,724	61,724

Table S5: **Sensitivity Analysis: American Boy with Irish Neighbor predicting Neighbor Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.86 (2.46)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**2.97**	2.31
Birth Year	1862 vs 1867	1.24	1.08
School Attendance	Attending vs Not attending	1.82	1.12
Density	Density 75 vs 1125	1.09	1.27

Table S6: **Sensitivity Analysis: American Boy with Irish Neighbor predicting Own Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and both TREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.74 (1.59)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.40	**2.31**
Birth Year	1862 vs 1867	1.17	1.08
School Attendance	Attending vs Not attending	**2.04**	1.12
Density	Density 75 vs 1125	1.18	1.27

Table S7: **Sensitivity Analysis: American Boy with Irish Neighbor predicting Neighbor Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 3.03 (2.56)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**3.08**	2.30
Birth Year	1862 vs 1867	1.26	1.05
School Attendance	Attending vs Not attending	1.16	1.12
Density	Density 75 vs 1125	1.12	1.20

Table S8: **Sensitivity Analysis: American Boy with Irish Neighbor predicting Own Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.73 (1.49)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.40	**2.30**
Birth Year	1862 vs 1867	1.04	1.05
School Attendance	Attending vs Not attending	1.18	1.12
Density	Density 75 vs 1125	1.10	1.20

Table S9: **Sensitivity Analysis: American Boy with Irish Neighbor predicting Neighbor Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.51 (2.40)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**2.97**	2.22
Birth Year	1862 vs 1867	1.24	1.05
School Attendance	Attending vs Not attending	1.82	1.10
Density	Density 75 vs 1125	1.09	1.28

Table S10: **Sensitivity Analysis: American Boy with Irish Neighbor predicting Own Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.61 (1.50)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.40	**2.22**
Birth Year	1862 vs 1867	1.17	1.05
School Attendance	Attending vs Not attending	**2.04**	1.10
Density	Density 75 vs 1125	1.18	1.28

Table S11: **Sensitivity Analysis: American Boy with Irish Neighbor predicting Neighbor Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.36 (2.29)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**3.08**	2.12
Birth Year	1862 vs 1867	1.26	1.03
School Attendance	Attending vs Not attending	1.16	1.10
Density	Density 75 vs 1125	1.12	1.20

Table S12: **Sensitivity Analysis: American Boy with Irish Neighbor predicting Own Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.47 (1.39)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.40	**2.12**
Birth Year	1862 vs 1867	1.04	1.03
School Attendance	Attending vs Not attending	1.18	1.10
Density	Density 75 vs 1125	1.10	1.20

Table S13: **Sensitivity Analysis: American Boy with German Neighbor predicting Neighbor Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.30 (2.02)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	2.00	1.75
Birth Year	1862 vs 1867	1.37	1.07
School Attendance	Attending vs Not attending	1.76	1.12
Density	Density 75 vs 1125	1.15	1.11

Table S14: **Sensitivity Analysis: American Boy with German Neighbor predicting Own Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.39 (1.23)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**1.40**	**1.75**
Birth Year	1862 vs 1867	1.16	1.07
School Attendance	Attending vs Not attending	**2.04**	1.12
Density	Density 75 vs 1125	1.18	1.11

Table S15: **Sensitivity Analysis: American Boy with German Neighbor predicting Neighbor Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.73 (1.96)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	2.08	1.75
Birth Year	1862 vs 1867	1.11	1.09
School Attendance	Attending vs Not attending	1.08	1.12
Density	Density 75 vs 1125	1.14	1.09

Table S16: **Sensitivity Analysis: American Boy with German Neighbor predicting Own Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.60 (1.41)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.40	**1.75**
Birth Year	1862 vs 1867	1.04	1.09
School Attendance	Attending vs Not attending	1.18	1.12
Density	Density 75 vs 1125	1.11	1.09

Table S17: **Sensitivity Analysis: American Boy with German Neighbor predicting Neighbor Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.82 (2.70)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	2.00	1.40
Birth Year	1862 vs 1867	1.37	1.04
School Attendance	Attending vs Not attending	1.76	1.11
Density	Density 75 vs 1125	1.15	1.02

Table S18: **Sensitivity Analysis: American Boy with German Neighbor predicting Own Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.34 (1.17)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**1.40**	**1.40**
Birth Year	1862 vs 1867	1.16	1.04
School Attendance	Attending vs Not attending	**2.04**	1.11
Density	Density 75 vs 1125	1.18	1.02

Table S19: **Sensitivity Analysis: American Boy with German Neighbor predicting Neighbor Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.61 (2.53)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	2.08	1.48
Birth Year	1862 vs 1867	1.11	1.06
School Attendance	Attending vs Not attending	1.08	1.10
Density	Density 75 vs 1125	1.14	1.01

Table S20: **Sensitivity Analysis: American Boy with German Neighbor predicting Own Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.40 (1.33)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.40	**1.48**
Birth Year	1862 vs 1867	1.04	1.06
School Attendance	Attending vs Not attending	1.18	1.10
Density	Density 75 vs 1125	1.11	1.01

Table S21: **Sensitivity Analysis: Irish Boy with American Neighbor predicting Neighbor Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.75 (1.18)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**2.19**	1.69
Birth Year	1862 vs 1867	1.20	1.03
School Attendance	Attending vs Not attending	1.71	1.01
Density	Density 75 vs 1125	1.04	1.06

Table S22: **Sensitivity Analysis: Irish Boy with American Neighbor predicting Own Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.06 (1.00)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**1.42**	**1.69**
Birth Year	1862 vs 1867	**1.48**	1.03
School Attendance	Attending vs Not attending	**1.51**	1.01
Density	Density 75 vs 1125	1.02	**1.06**

Table S23: **Sensitivity Analysis: Irish Boy with American Neighbor predicting Neighbor Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.73 (1.29)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**2.01**	1.70
Birth Year	1862 vs 1867	1.02	1.03
School Attendance	Attending vs Not attending	1.11	1.02
Density	Density 75 vs 1125	1.06	1.06

Table S24: **Sensitivity Analysis: Irish Boy with American Neighbor predicting Own Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.28 (1.00)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**2.06**	**1.70**
Birth Year	1862 vs 1867	1.26	1.03
School Attendance	Attending vs Not attending	**1.34**	1.02
Density	Density 75 vs 1125	1.03	1.06

Table S25: **Sensitivity Analysis: Irish Boy with American Neighbor predicting Neighbor Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.27 (2.09)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	2.19	1.78
Birth Year	1862 vs 1867	1.20	1.04
School Attendance	Attending vs Not attending	1.71	1.02
Density	Density 75 vs 1125	1.04	1.04

Table S26: **Sensitivity Analysis: Irish Boy with American Neighbor predicting Own Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.77 (1.58)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.42	**1.78**
Birth Year	1862 vs 1867	1.48	1.04
School Attendance	Attending vs Not attending	1.51	1.02
Density	Density 75 vs 1125	1.02	1.04

Table S27: **Sensitivity Analysis: Irish Boy with American Neighbor predicting Neighbor Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.98 (1.81)			
Covariate	Contrast	RR$_{XY}$	RR$_{TX}$
Occupation	Laborers vs Professional	**2.01**	1.80
Birth Year	1862 vs 1867	1.02	1.04
School Attendance	Attending vs Not attending	1.11	1.01
Density	Density 75 vs 1125	1.06	1.04

Table S28: **Sensitivity Analysis: Irish Boy with American Neighbor predicting Own Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.80 (1.64)			
Covariate	Contrast	RR$_{XY}$	RR$_{TX}$
Occupation	Laborers vs Professional	**2.06**	**1.80**
Birth Year	1862 vs 1867	1.26	1.04
School Attendance	Attending vs Not attending	1.34	1.01
Density	Density 75 vs 1125	1.03	1.04

Table S29: **Sensitivity Analysis: Irish Boy with German Neighbor predicting Neighbor Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.40 (1.79)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.55	1.83
Birth Year	1862 vs 1867	1.12	1.01
School Attendance	Attending vs Not attending	1.62	1.00
Density	Density 75 vs 1125	1.02	1.12

Table S30: **Sensitivity Analysis: Irish Boy with German Neighbor predicting Own Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.20 (1.00)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**1.42**	**1.83**
Birth Year	1862 vs 1867	**1.48**	1.01
School Attendance	Attending vs Not attending	**1.51**	1.00
Density	Density 75 vs 1125	1.02	1.12

Table S31: **Sensitivity Analysis: Irish Boy with German Neighbor predicting Neighbor Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.94 (1.00)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**3.34**	**1.95**
Birth Year	1862 vs 1867	1.11	1.13
School Attendance	Attending vs Not attending	1.17	1.03
Density	Density 75 vs 1125	1.03	1.15

Table S32: **Sensitivity Analysis: Irish Boy with German Neighbor predicting Own Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.19 (1.00)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**2.06**	**1.95**
Birth Year	1862 vs 1867	**1.26**	1.13
School Attendance	Attending vs Not attending	**1.34**	1.03
Density	Density 75 vs 1125	1.03	1.15

Table S33: **Sensitivity Analysis: Irish Boy with German Neighbor predicting Neighbor Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.87 (1.67)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.55	1.08
Birth Year	1862 vs 1867	1.12	1.01
School Attendance	Attending vs Not attending	1.62	1.00
Density	Density 75 vs 1125	1.02	1.02

Table S34: **Sensitivity Analysis: Irish Boy with German Neighbor predicting Own Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.46 (1.38)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.42	1.08
Birth Year	1862 vs 1867	**1.48**	1.01
School Attendance	Attending vs Not attending	**1.51**	1.00
Density	Density 75 vs 1125	1.02	1.02

Table S35: **Sensitivity Analysis: Irish Boy with German Neighbor predicting Neighbor Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.87 (1.71)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**3.34**	1.06
Birth Year	1862 vs 1867	1.11	1.04
School Attendance	Attending vs Not attending	1.17	1.01
Density	Density 75 vs 1125	1.03	1.02

Table S36: **Sensitivity Analysis: Irish Boy with German Neighbor predicting Own Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.45 (1.35)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**2.06**	1.06
Birth Year	1862 vs 1867	1.26	1.04
School Attendance	Attending vs Not attending	1.34	1.01
Density	Density 75 vs 1125	1.03	1.02

Table S37: **Sensitivity Analysis: German Boy with American Neighbor predicting Neighbor Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.45 (1.00)			
Covariate	Contrast	RR$_{XY}$	RR$_{TX}$
Occupation	Laborers vs Professional	1.42	1.32
Birth Year	1862 vs 1867	1.05	1.01
School Attendance	Attending vs Not attending	**1.73**	1.05
Density	Density 75 vs 1125	1.19	1.06

Table S38: **Sensitivity Analysis: German Boy with American Neighbor predicting Own Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.53 (1.33)			
Covariate	Contrast	RR$_{XY}$	RR$_{TX}$
Occupation	Laborers vs Professional	1.39	1.32
Birth Year	1862 vs 1867	1.39	1.01
School Attendance	Attending vs Not attending	1.45	1.05
Density	Density 75 vs 1125	1.01	1.06

Table S39: **Sensitivity Analysis: German Boy with American Neighbor predicting Neighbor Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.45 (1.20)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**1.49**	1.33
Birth Year	1862 vs 1867	1.22	1.01
School Attendance	Attending vs Not attending	1.01	1.05
Density	Density 75 vs 1125	1.12	1.06

Table S40: **Sensitivity Analysis: German Boy with American Neighbor predicting Own Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.45 (1.23)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.41	1.33
Birth Year	1862 vs 1867	1.25	1.01
School Attendance	Attending vs Not attending	1.13	1.05
Density	Density 75 vs 1125	1.00	1.06

Table S41: **Sensitivity Analysis: German Boy with American Neighbor predicting Neighbor Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.80 (1.63)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.42	1.54
Birth Year	1862 vs 1867	1.05	1.00
School Attendance	Attending vs Not attending	1.73	1.05
Density	Density 75 vs 1125	1.19	1.04

Table S42: **Sensitivity Analysis: German Boy with American Neighbor predicting Own Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.06 (1.87)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.39	1.54
Birth Year	1862 vs 1867	1.39	1.00
School Attendance	Attending vs Not attending	1.45	1.05
Density	Density 75 vs 1125	1.01	1.04

Table S43: **Sensitivity Analysis: German Boy with American Neighbor predicting Neighbor Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.88 (1.71)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.49	1.53
Birth Year	1862 vs 1867	1.22	1.00
School Attendance	Attending vs Not attending	1.01	1.04
Density	Density 75 vs 1125	1.12	1.04

Table S44: **Sensitivity Analysis: German Boy with American Neighbor predicting Own Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.90 (1.72)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.41	1.53
Birth Year	1862 vs 1867	1.25	1.00
School Attendance	Attending vs Not attending	1.13	1.04
Density	Density 75 vs 1125	1.00	1.04

Table S45: **Sensitivity Analysis: German Boy with Irish Neighbor predicting Neighbor Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.27 (1.59)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**2.31**	1.00
Birth Year	1862 vs 1867	1.02	1.03
School Attendance	Attending vs Not attending	1.58	1.05
Density	Density 75 vs 1125	1.19	1.01

Table S46: **Sensitivity Analysis: German Boy with Irish Neighbor predicting Own Wife (1900, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.75 (1.36)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.39	1.00
Birth Year	1862 vs 1867	1.39	1.03
School Attendance	Attending vs Not attending	1.45	1.05
Density	Density 75 vs 1125	1.01	1.01

Table S47: **Sensitivity Analysis: German Boy with Irish Neighbor predicting Neighbor Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.18 (1.00)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	2.08	1.03
Birth Year	1862 vs 1867	1.05	1.06
School Attendance	Attending vs Not attending	1.19	1.05
Density	Density 75 vs 1125	1.07	1.00

Table S48: **Sensitivity Analysis: German Boy with Irish Neighbor predicting Own Wife (1910, CEM).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the TREATMENT and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothTREATMENT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.60 (1.29)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.41	1.03
Birth Year	1862 vs 1867	1.25	1.06
School Attendance	Attending vs Not attending	1.13	1.05
Density	Density 75 vs 1125	1.00	1.00

Table S49: **Sensitivity Analysis: German Boy with Irish Neighbor predicting Neighbor Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 2.18 (2.01)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**2.31**	1.44
Birth Year	1862 vs 1867	1.02	1.00
School Attendance	Attending vs Not attending	1.58	1.02
Density	Density 75 vs 1125	1.19	1.01

Table S50: **Sensitivity Analysis: German Boy with Irish Neighbor predicting Own Wife (1900, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.50 (1.32)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	1.39	1.44
Birth Year	1862 vs 1867	1.39	1.00
School Attendance	Attending vs Not attending	1.45	1.02
Density	Density 75 vs 1125	1.01	1.01

Table S51: **Sensitivity Analysis: German Boy with Irish Neighbor predicting Neighbor Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.91 (1.79)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**2.08**	1.41
Birth Year	1862 vs 1867	1.05	1.01
School Attendance	Attending vs Not attending	1.19	1.02
Density	Density 75 vs 1125	1.07	1.01

Table S52: **Sensitivity Analysis: German Boy with Irish Neighbor predicting Own Wife (1910, Count).** The table compares the model's E-value (top row) against observed covariate strengths. To bring our coefficients to 0, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the first E-value. To render our estimates statistically insignificant, an unmeasured confounder would need to be associated with both the COUNT variable and the outcome by at least the second E-value (CI). To benchmark these values, the remainder of the table shows the associations between observed covariates and bothCOUNT and the outcome. RR_{XY} denotes the Covariate-Outcome association; RR_{TX} denotes the Treatment-Covariate association. Values in **bold** exceed the first E-value.

E-value estimate (CI): 1.31 (1.22)			
Covariate	Contrast	RR_{XY}	RR_{TX}
Occupation	Laborers vs Professional	**1.41**	**1.41**
Birth Year	1862 vs 1867	1.25	1.01
School Attendance	Attending vs Not attending	1.13	1.02
Density	Density 75 vs 1125	1.00	1.01

Table S53: **Summary statistics for the continuous treatment variable.** Mean and standard deviation of the number of neighboring households of a given ethnic group among the 20 neighboring households for the 1900 and 1910 linked samples.

Own Ethnicity	Neighboring Group	1880-1900	1880-1910
American	German	3.31 (3.71)	3.28 (3.72)
American	Irish	3.19 (3.5)	2.99 (3.35)
German	American	5.43 (3.92)	5.47 (3.94)
German	Irish	2.47 (3.11)	2.35 (3)
Irish	American	5.74 (3.91)	5.93 (3.97)
Irish	German	2.8 (3.07)	2.78 (3.07)

Table S54: **Spouse's Outcome by Men's Marital status in the Linked 1880–1900 Sample.** Counts show the distribution of spousal ethnicity categories (American, Irish, German, Other, Not married, NA) by men's reported marital status in 1900. "NA" indicates missing or unclassifiable spousal ethnicity.

Wife Outcome	Married, spouse present	Married, spouse absent	Divorced	Widowed	Never married	Total
American	43,908	0	0	1	0	43,909
Irish	14,958	0	0	0	0	14,958
German	28,492	0	0	0	0	28,492
Other	27,079	0	0	0	0	27,079
Not married	0	0	0	0	299,202	299,202
Other marriage status	0	9,136	812	4,614	0	14,562
Total	114,437	9,136	812	4,615	299,202	428,202

Table S55: **Wife Outcome by Marital status for Boys in the Linked 1880–1910 Sample.** Counts show the distribution of spousal ethnicity categories (American, Irish, German, Other, Not married, NA) by men's reported marital status in 1910. "NA" indicates missing or unclassifiable spousal ethnicity.

Wife Outcome	Married, spouse present	Married, spouse absent	Divorced	Widowed	Never married	Total
American	49,175	0	1	0	0	49,176
Irish	13,511	0	0	1	0	13,512
German	28,839	0	0	0	0	28,839
Other	31,515	0	0	2	0	31,517
Not married	0	0	0	0	73,233	73,233
Other marriage status	0	5,243	1,157	4,761	0	11,161
Total	123,040	5,243	1,158	4,764	73,233	207,438

Table S56: **Mapping of Census Birthplace Codes to Categories.** Definition of the twelve birthplace categories used as input features for the K-Means++ algorithm (Stage 1). Raw values refer to the numeric codes in the 1880 US Census (*40*).

Aggregated Category	Census Birthplace Code Range
First-generation Americans	< 120 (both parents born outside the US)
Second-generation Americans	< 120 (at least one parent born in the US)
Other North Americans	150 – 199
Central American & Caribbean	200 – 299
South Americans	300
Other Europeans	400 – 499 (excluding 414, 453)
Irish	414
Germans	453
Asians	500 – 599
Africans	600 – 699
Oceanians	700 – 799
Other/Unknown	All other values

Table S57: **Replication of Table S1, Minimum cluster size = 15.** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	1.51	2.34	2.35	0.70	4.59	1.37
ethnic group	AME	+1.34***	+0.53*	+0.57	+0.58**	+0.41	-0.00
	(SE)	(0.26)	(0.27)	(0.36)	(0.18)	(0.34)	(0.25)
	p-value	<.001	0.047	0.114	0.001	0.229	0.991
Wife own	Ref.	16.11	16.66	7.04	6.83	15.20	13.12
ethnic group	AME	-3.05***	-1.29**	-0.09	-0.03	-1.86**	-1.23
	(SE)	(0.49)	(0.49)	(0.26)	(0.61)	(0.68)	(0.81)
	p-value	<.001	0.009	0.718	0.963	0.006	0.128
Wife other group	Ref.	7.33	5.90	3.50	2.49	5.78	3.75
	AME	-0.48	+0.30	+0.76*	+0.19	+0.91**	+1.36***
	(SE)	(0.35)	(0.32)	(0.33)	(0.35)	(0.34)	(0.32)
	p-value	0.165	0.342	0.020	0.585	0.007	<.001
American wife	Ref.	-	-	-	2.23	-	3.60
	AME	-	-	-	+0.40	-	+0.63
	(SE)	-	-	-	(0.59)	-	(0.36)
	p-value	-	-	-	0.502	-	0.081
Other outcomes	Ref.	3.58	3.68	2.59	2.55	2.06	2.02
	AME	+0.14	-0.49	+0.06	-0.03	+0.30	+0.41
	(SE)	(0.23)	(0.32)	(0.18)	(0.30)	(0.23)	(0.29)
	p-value	0.546	0.126	0.726	0.910	0.199	0.151
Not married	Ref.	71.47	71.42	84.52	85.19	72.37	76.14
	AME	+2.06***	+0.95	-1.31	-1.11	+0.25	-1.17
	(SE)	(0.45)	(0.56)	(0.78)	(0.81)	(0.82)	(0.93)
	p-value	<.001	0.092	0.093	0.171	0.763	0.208
n		19,168	17,354	15,016	8,052	19,322	8,999

Table S58: **Replication of Table S2, Minimum cluster size = 15.** Each column displays the results of a separate multinomial regression for each dyad. “Ref.” denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring ethnic group	Ref.	3.37	5.16	8.00	3.69	12.79	3.18
	AME	+3.48**	+1.56	+2.15	-0.61	+1.28	-0.06
	(SE)	(1.09)	(1.18)	(1.32)	(1.18)	(0.97)	(1.19)
	p-value	0.001	0.188	0.104	0.606	0.188	0.959
Wife own ethnic group	Ref.	39.11	41.42	16.87	14.99	31.42	30.03
	AME	-5.79***	-1.97	-0.43	+2.40	-1.68	-1.42
	(SE)	(1.07)	(1.63)	(0.76)	(1.74)	(1.53)	(1.78)
	p-value	<.001	0.227	0.571	0.167	0.272	0.425
Wife other group	Ref.	16.57	15.22	10.94	9.55	15.27	12.78
	AME	+1.47	+0.67	+1.10	-0.29	+1.89*	+1.15
	(SE)	(0.87)	(0.99)	(0.94)	(1.55)	(0.89)	(1.61)
	p-value	0.092	0.498	0.244	0.849	0.033	0.474
American wife	Ref.	-	-	-	8.46	-	9.63
	AME	-	-	-	+1.64	-	+2.42***
	(SE)	-	-	-	(2.28)	-	(0.67)
	p-value	-	-	-	0.473	-	<.001
Other outcomes	Ref.	8.00	6.64	6.18	5.42	4.03	3.56
	AME	-0.97	-1.11**	-0.51	+0.05	+0.07	+0.54
	(SE)	(0.52)	(0.36)	(0.64)	(1.18)	(0.44)	(0.63)
	p-value	0.061	0.002	0.424	0.964	0.875	0.397
Not married	Ref.	32.95	31.56	58.01	57.89	36.49	40.82
	AME	+1.82	+0.86	-2.31	-3.19	-1.57	-2.62
	(SE)	(1.02)	(1.79)	(1.89)	(2.51)	(1.24)	(1.46)
	p-value	0.074	0.633	0.222	0.204	0.206	0.072
n		5,887	5,554	3,282	1,562	6,425	2,630

Table S59: **Replication of Table S1, Minimum cluster size = 25.** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	1.73	2.55	2.60	0.90	5.15	1.58
ethnic group	AME	+0.80***	+0.67*	+0.27	+0.42*	+0.36	-0.12
	(SE)	(0.19)	(0.33)	(0.30)	(0.19)	(0.32)	(0.17)
	p-value	<.001	0.042	0.370	0.027	0.265	0.471
Wife own	Ref.	16.47	16.77	7.31	6.56	15.41	13.62
ethnic group	AME	-2.71***	-0.97*	-0.50	-0.36	-1.58***	-2.18***
	(SE)	(0.38)	(0.39)	(0.50)	(0.26)	(0.39)	(0.57)
	p-value	<.001	0.014	0.323	0.169	<.001	<.001
Wife other group	Ref.	6.96	6.17	3.69	2.89	6.27	4.02
	AME	+0.11	-0.13	+0.76	+0.01	+0.38	+1.12*
	(SE)	(0.31)	(0.28)	(0.43)	(0.45)	(0.33)	(0.49)
	p-value	0.707	0.648	0.078	0.975	0.253	0.021
American wife	Ref.	-	-	-	1.96	-	3.99
	AME	-	-	-	+0.97*	-	+0.45
	(SE)	-	-	-	(0.38)	-	(0.50)
	p-value	-	-	-	0.011	-	0.372
Other outcomes	Ref.	3.56	3.52	2.81	2.26	2.39	2.02
	AME	+0.15	-0.12	-0.17	+0.22	+0.23	+0.34
	(SE)	(0.19)	(0.26)	(0.25)	(0.20)	(0.24)	(0.22)
	p-value	0.420	0.646	0.488	0.276	0.333	0.125
Not married	Ref.	71.29	70.99	83.59	85.43	70.77	74.76
	AME	+1.64**	+0.54	-0.37	-1.26	+0.61	+0.39
	(SE)	(0.52)	(0.53)	(0.91)	(0.79)	(0.81)	(1.03)
	p-value	0.002	0.300	0.687	0.110	0.453	0.704
n		24,768	24,115	18,000	9,870	24,733	10,911

Table S60: **Replication of Table S2, Minimum cluster size = 25.** Each column displays the results of a separate multinomial regression for each dyad. “Ref.” denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring ethnic group	Ref.	3.67	5.58	9.27	3.06	14.57	2.51
	AME	+2.26**	+1.99***	+0.81	+0.30	+1.00	+0.42
	(SE)	(0.78)	(0.51)	(0.90)	(0.72)	(0.82)	(0.76)
	p-value	0.004	<.001	0.367	0.674	0.223	0.579
Wife own ethnic group	Ref.	41.13	41.69	17.98	16.84	31.97	31.24
	AME	-6.11***	-2.32	-1.76	-0.42	-2.03*	-4.37**
	(SE)	(1.12)	(1.93)	(1.06)	(1.55)	(0.82)	(1.55)
	p-value	<.001	0.229	0.096	0.787	0.014	0.005
Wife other group	Ref.	18.02	15.58	10.83	7.91	15.10	14.26
	AME	+0.19	-0.30	+1.83**	+1.35	+1.01	+0.58
	(SE)	(0.72)	(0.69)	(0.66)	(1.23)	(0.95)	(1.31)
	p-value	0.794	0.662	0.005	0.272	0.290	0.659
American wife	Ref.	-	-	-	9.07	-	11.18
	AME	-	-	-	+0.51	-	+1.54
	(SE)	-	-	-	(1.63)	-	(1.16)
	p-value	-	-	-	0.755	-	0.184
Other outcomes	Ref.	5.87	5.99	5.51	5.37	3.94	3.79
	AME	+1.01*	+0.30	+0.09	-0.95	+0.04	-0.18
	(SE)	(0.50)	(0.28)	(0.53)	(0.86)	(0.28)	(0.81)
	p-value	0.044	0.281	0.864	0.271	0.874	0.826
Not married	Ref.	31.31	31.16	56.41	57.75	34.43	37.03
	AME	+2.66*	+0.33	-0.98	-0.80	-0.02	+2.00
	(SE)	(1.20)	(1.59)	(1.74)	(1.91)	(0.90)	(2.39)
	p-value	0.027	0.833	0.574	0.676	0.978	0.401
n		8,237	8,604	4,113	2,035	8,865	3,429

Table S61: **Replication of Table S1, Minimum cluster size = 35.** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	1.67	2.48	2.72	0.82	5.12	1.41
ethnic group	AME	+0.96***	+0.79**	+0.25	+0.33*	+0.16	+0.25
	(SE)	(0.26)	(0.29)	(0.23)	(0.15)	(0.33)	(0.13)
	p-value	<.001	0.007	0.283	0.026	0.628	0.056
Wife own	Ref.	15.99	15.76	7.48	6.96	15.95	14.16
ethnic group	AME	-2.48***	-0.83*	-0.48	+0.05	-1.45***	-1.96**
	(SE)	(0.43)	(0.34)	(0.37)	(0.41)	(0.34)	(0.68)
	p-value	<.001	0.014	0.190	0.904	<.001	0.004
Wife other group	Ref.	7.00	6.38	3.95	2.77	6.28	4.88
	AME	-0.27	+0.25	+0.29	+0.17	+0.69	+0.61*
	(SE)	(0.31)	(0.44)	(0.24)	(0.29)	(0.39)	(0.29)
	p-value	0.388	0.569	0.219	0.555	0.078	0.034
American wife	Ref.	-	-	-	2.59	-	4.34
	AME	-	-	-	+0.02	-	-0.09
	(SE)	-	-	-	(0.34)	-	(0.30)
	p-value	-	-	-	0.945	-	0.770
Other outcomes	Ref.	3.47	3.55	2.67	2.32	2.16	2.24
	AME	+0.25	-0.55	+0.21	+0.18	+0.38*	-0.11
	(SE)	(0.33)	(0.42)	(0.17)	(0.17)	(0.19)	(0.14)
	p-value	0.462	0.190	0.213	0.295	0.043	0.413
Not married	Ref.	71.87	71.83	83.17	84.55	70.49	72.97
	AME	+1.54*	+0.34	-0.27	-0.75	+0.22	+1.29
	(SE)	(0.72)	(0.61)	(0.44)	(0.65)	(0.65)	(0.94)
	p-value	0.032	0.576	0.539	0.254	0.739	0.171
n		26,975	25,845	21,230	12,292	30,897	15,586

Table S62: **Replication of Table S2, Minimum cluster size = 35.** Each column displays the results of a separate multinomial regression for each dyad. “Ref.” denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	3.49	5.64	9.75	2.71	13.42	2.84
ethnic group	AME	+2.52***	+3.07***	+0.34	+0.63	+1.28*	+0.89
	(SE)	(0.56)	(0.80)	(0.90)	(0.52)	(0.59)	(0.81)
	p-value	<.001	<.001	0.704	0.225	0.030	0.267
Wife own	Ref.	41.07	40.24	18.69	17.45	32.85	31.82
ethnic group	AME	-7.22***	-3.38**	-0.94	-0.74	-2.62*	-4.51***
	(SE)	(1.03)	(1.19)	(1.10)	(1.50)	(1.03)	(0.70)
	p-value	<.001	0.005	0.390	0.622	0.011	<.001
Wife other group	Ref.	17.49	16.67	11.11	10.10	15.98	13.37
	AME	+0.99	-0.03	+0.56	-0.74	+1.10**	+1.40
	(SE)	(1.18)	(0.75)	(0.88)	(1.42)	(0.43)	(0.90)
	p-value	0.401	0.964	0.527	0.599	0.010	0.121
American wife	Ref.	-	-	-	9.18	-	11.55
	AME	-	-	-	-0.32	-	+0.82
	(SE)	-	-	-	(0.80)	-	(0.88)
	p-value	-	-	-	0.685	-	0.350
Other outcomes	Ref.	5.85	6.14	5.37	5.58	3.89	3.18
	AME	+0.98*	+0.07	-0.02	-0.81	+0.20	+0.81
	(SE)	(0.46)	(0.53)	(0.47)	(1.10)	(0.32)	(0.78)
	p-value	0.034	0.898	0.972	0.460	0.520	0.302
Not married	Ref.	32.10	31.31	55.08	54.98	33.87	37.25
	AME	+2.73	+0.27	+0.06	+2.00	+0.04	+0.59
	(SE)	(1.47)	(0.80)	(1.06)	(1.47)	(0.98)	(1.51)
	p-value	0.063	0.736	0.954	0.176	0.969	0.697
n		8,723	8,938	4,973	2,591	12,054	5,385

Table S63: **Replication of Table S1, Minimum cluster size = 45.** Each column displays the results of a separate multinomial regression for each dyad. “Ref.” denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	1.65	2.27	3.06	0.95	5.41	1.50
ethnic group	AME	+1.26***	+0.91***	+0.56*	+0.33*	+0.17	+0.37
	(SE)	(0.26)	(0.20)	(0.22)	(0.14)	(0.34)	(0.21)
	p-value	<.001	<.001	0.012	0.017	0.626	0.069
Wife own	Ref.	16.02	15.89	7.39	7.63	15.65	14.05
ethnic group	AME	-2.55***	-0.77*	-0.00	-0.54	-1.59**	-1.80*
	(SE)	(0.40)	(0.34)	(0.41)	(0.37)	(0.53)	(0.77)
	p-value	<.001	0.026	0.992	0.145	0.003	0.020
Wife other group	Ref.	7.29	6.72	3.98	2.90	6.35	4.59
	AME	-0.40	-0.07	+0.54*	+0.07	+0.72*	+0.71**
	(SE)	(0.26)	(0.30)	(0.26)	(0.21)	(0.31)	(0.27)
	p-value	0.130	0.827	0.035	0.748	0.019	0.008
American wife	Ref.	-	-	-	2.57	-	4.91
	AME	-	-	-	+0.39	-	-0.19
	(SE)	-	-	-	(0.20)	-	(0.37)
	p-value	-	-	-	0.053	-	0.610
Other outcomes	Ref.	3.83	3.60	2.86	2.26	2.21	2.27
	AME	-0.04	-0.32	+0.07	+0.45***	+0.33*	+0.18
	(SE)	(0.35)	(0.27)	(0.17)	(0.11)	(0.13)	(0.14)
	p-value	0.909	0.236	0.691	<.001	0.014	0.204
Not married	Ref.	71.21	71.53	82.71	83.70	70.39	72.69
	AME	+1.73*	+0.24	-1.16*	-0.70	+0.37	+0.72
	(SE)	(0.70)	(0.60)	(0.46)	(0.56)	(0.75)	(1.18)
	p-value	0.013	0.689	0.011	0.204	0.617	0.540
n		34,100	32,264	26,286	15,484	34,436	17,546

Table S64: **Replication of Table S2, Minimum cluster size = 45.** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring ethnic group	Ref.	3.81	5.12	9.94	3.16	14.57	2.20
	AME	+2.95***	+3.30***	+0.81	+0.23	+0.87	+1.76**
	(SE)	(0.55)	(0.75)	(0.85)	(0.51)	(0.72)	(0.56)
	p-value	<.001	<.001	0.344	0.655	0.228	0.002
Wife own ethnic group	Ref.	40.26	40.16	17.54	17.03	31.25	30.40
	AME	-7.32***	-3.42***	+0.04	+0.66	-1.54	-3.87*
	(SE)	(1.88)	(0.94)	(0.93)	(1.51)	(0.98)	(1.80)
	p-value	<.001	<.001	0.967	0.661	0.116	0.032
Wife other group	Ref.	17.95	16.38	11.52	9.29	15.66	12.68
	AME	+0.71	+0.44	+0.56	-0.18	+1.54**	+1.65
	(SE)	(0.66)	(0.57)	(0.65)	(1.01)	(0.49)	(1.69)
	p-value	0.285	0.444	0.384	0.862	0.002	0.327
American wife	Ref.	-	-	-	9.54	-	12.66
	AME	-	-	-	-0.16	-	+0.20
	(SE)	-	-	-	(0.94)	-	(0.83)
	p-value	-	-	-	0.865	-	0.806
Other outcomes	Ref.	5.55	5.89	5.06	4.71	4.32	3.91
	AME	+1.40*	+0.01	+0.68	-0.59	-0.35	+0.43
	(SE)	(0.62)	(0.44)	(0.44)	(1.09)	(0.34)	(0.40)
	p-value	0.023	0.989	0.128	0.590	0.308	0.280
Not married	Ref.	32.43	32.45	55.93	56.25	34.20	38.14
	AME	+2.25	-0.33	-2.09	+0.03	-0.52	-0.18
	(SE)	(1.38)	(0.86)	(1.36)	(1.85)	(0.99)	(1.59)
	p-value	0.103	0.700	0.124	0.986	0.602	0.908
n		11,918	11,771	6,598	3,427	13,428	6,181

Table S65: **Replication of Table S1, Minimum cluster size = 100.** Each column displays the results of a separate multinomial regression for each dyad. “Ref.” denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring ethnic group	Ref.	1.78	2.62	2.84	0.99	5.55	1.60
	AME	+1.28***	+1.12***	+0.89***	+0.54***	+0.36	+0.32
	(SE)	(0.16)	(0.18)	(0.24)	(0.14)	(0.24)	(0.21)
	p-value	<.001	<.001	<.001	<.001	0.127	0.120
Wife own ethnic group	Ref.	16.36	15.65	7.68	7.77	16.92	15.26
	AME	-3.13***	-1.50***	-0.22	-0.69	-2.03***	-2.66***
	(SE)	(0.35)	(0.39)	(0.37)	(0.36)	(0.58)	(0.61)
	p-value	<.001	<.001	0.551	0.059	<.001	<.001
Wife other group	Ref.	7.54	6.74	4.65	3.29	6.88	5.05
	AME	-0.19	+0.36	+0.31	+0.20	+0.87***	+0.87***
	(SE)	(0.23)	(0.28)	(0.30)	(0.11)	(0.26)	(0.26)
	p-value	0.414	0.193	0.311	0.075	<.001	<.001
American wife	Ref.	-	-	-	2.65	-	4.95
	AME	-	-	-	+0.75***	-	+0.09
	(SE)	-	-	-	(0.21)	-	(0.27)
	p-value	-	-	-	<.001	-	0.752
Other outcomes	Ref.	3.67	3.49	2.91	2.54	2.36	2.31
	AME	+0.06	-0.32	+0.02	+0.33	+0.35*	+0.19
	(SE)	(0.30)	(0.24)	(0.22)	(0.24)	(0.16)	(0.20)
	p-value	0.837	0.182	0.942	0.164	0.027	0.348
Not married	Ref.	70.65	71.50	81.91	82.77	68.29	70.84
	AME	+1.97***	+0.34	-0.99	-1.13	+0.45	+1.19
	(SE)	(0.43)	(0.52)	(0.67)	(0.73)	(0.70)	(1.04)
	p-value	<.001	0.506	0.140	0.125	0.518	0.254
n		46,589	45,862	35,489	21,925	48,274	27,617

Table S66: **Replication of Table S2, Minimum cluster size = 100.** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	3.96	6.08	10.12	3.29	14.15	2.47
ethnic group	AME	+3.25***	+2.76**	+2.02**	+0.61	+0.81	+1.40
	(SE)	(0.42)	(0.90)	(0.77)	(0.65)	(0.49)	(0.78)
	p-value	<.001	0.002	0.008	0.342	0.099	0.072
Wife own	Ref.	39.23	39.83	16.65	16.01	32.50	30.74
ethnic group	AME	-6.60***	-5.30***	+0.09	+0.05	-3.07**	-3.63*
	(SE)	(1.31)	(1.25)	(1.58)	(1.70)	(1.00)	(1.42)
	p-value	<.001	<.001	0.952	0.979	0.002	0.010
Wife other group	Ref.	19.03	16.57	13.22	10.54	16.02	14.10
	AME	-0.65	+1.44	+0.50	-0.35	+1.68**	+0.71
	(SE)	(0.65)	(0.78)	(0.83)	(1.37)	(0.53)	(0.65)
	p-value	0.319	0.064	0.546	0.799	0.002	0.274
American wife	Ref.	-	-	-	9.97	-	12.74
	AME	-	-	-	+1.10	-	+0.14
	(SE)	-	-	-	(1.05)	-	(0.85)
	p-value	-	-	-	0.295	-	0.868
Other outcomes	Ref.	5.93	5.77	5.54	5.78	4.15	4.21
	AME	+0.69	+0.05	-0.35	-1.07	+0.10	+0.20
	(SE)	(0.45)	(0.31)	(0.42)	(0.57)	(0.19)	(0.42)
	p-value	0.125	0.872	0.408	0.061	0.599	0.626
Not married	Ref.	31.85	31.75	54.46	54.41	33.18	35.73
	AME	+3.31*	+1.05	-2.27	-0.35	+0.48	+1.17
	(SE)	(1.36)	(0.90)	(1.24)	(1.90)	(0.93)	(1.48)
	p-value	0.015	0.243	0.068	0.855	0.604	0.428
n		17,660	17,827	10,033	5,766	21,210	11,006

Table S67: **Replication of Table S1, Minimum cluster size = 200.** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	1.83	2.79	3.23	1.10	5.72	1.63
ethnic group	AME	+1.20***	+1.12***	+0.71**	+0.55**	+0.87*	+0.33
	(SE)	(0.17)	(0.21)	(0.25)	(0.19)	(0.40)	(0.18)
	p-value	<.001	<.001	0.004	0.003	0.029	0.074
Wife own	Ref.	16.61	16.60	8.00	7.62	16.84	14.65
ethnic group	AME	-2.83***	-1.56***	-0.43*	-0.64*	-1.96***	-2.13*
	(SE)	(0.40)	(0.42)	(0.21)	(0.29)	(0.45)	(0.90)
	p-value	<.001	<.001	0.039	0.029	<.001	0.019
Wife other group	Ref.	7.58	6.36	4.60	3.23	7.06	5.23
	AME	-0.05	+0.51	+0.48	+0.33	+0.76**	+0.66**
	(SE)	(0.17)	(0.30)	(0.30)	(0.23)	(0.25)	(0.22)
	p-value	0.772	0.083	0.100	0.145	0.002	0.003
American wife	Ref.	-	-	-	2.91	-	5.05
	AME	-	-	-	+0.58***	-	+0.39
	(SE)	-	-	-	(0.17)	-	(0.34)
	p-value	-	-	-	<.001	-	0.261
Other outcomes	Ref.	3.95	3.64	3.00	2.73	2.42	2.46
	AME	-0.34	-0.29	+0.16	+0.18	+0.28	-0.02
	(SE)	(0.20)	(0.18)	(0.12)	(0.17)	(0.18)	(0.17)
	p-value	0.087	0.104	0.178	0.273	0.115	0.922
Not married	Ref.	70.04	70.61	81.17	82.41	67.96	70.98
	AME	+2.02***	+0.22	-0.93	-1.01**	+0.04	+0.77
	(SE)	(0.51)	(0.53)	(0.52)	(0.38)	(0.74)	(0.98)
	p-value	<.001	0.672	0.074	0.009	0.952	0.430
n		51,792	51,676	37,254	22,537	52,082	29,138

Table S68: **Replication of Table S2, Minimum cluster size = 200.** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring ethnic group	Ref.	4.04	5.65	9.59	3.25	14.70	2.99
	AME	+3.26***	+3.60***	+3.26**	+0.88	+2.10***	+1.21*
	(SE)	(0.50)	(1.03)	(1.04)	(0.51)	(0.64)	(0.58)
	p-value	<.001	<.001	0.002	0.086	<.001	0.037
Wife own ethnic group	Ref.	40.79	41.52	16.89	16.09	31.15	30.19
	AME	-7.60***	-6.08***	+0.07	-0.49	-2.84***	-4.33***
	(SE)	(1.21)	(1.29)	(1.22)	(1.52)	(0.82)	(1.04)
	p-value	<.001	<.001	0.955	0.747	<.001	<.001
Wife other group	Ref.	18.49	15.37	13.67	11.00	15.81	13.81
	AME	-0.26	+1.79*	-0.17	-0.10	+1.42**	+1.22
	(SE)	(0.45)	(0.74)	(0.77)	(1.25)	(0.53)	(0.84)
	p-value	0.559	0.016	0.822	0.935	0.007	0.145
American wife	Ref.	-	-	-	9.63	-	13.08
	AME	-	-	-	+2.58**	-	+1.27
	(SE)	-	-	-	(0.90)	-	(0.72)
	p-value	-	-	-	0.004	-	0.077
Other outcomes	Ref.	5.88	6.10	6.04	5.97	4.10	4.14
	AME	+0.93*	-0.12	-0.44	-1.10*	+0.15	+0.34
	(SE)	(0.39)	(0.43)	(0.28)	(0.56)	(0.15)	(0.35)
	p-value	0.016	0.787	0.117	0.050	0.317	0.332
Not married	Ref.	30.80	31.36	53.80	54.05	34.24	35.79
	AME	+3.67***	+0.80	-2.72	-1.76	-0.83	+0.29
	(SE)	(0.93)	(0.73)	(1.52)	(1.99)	(0.90)	(1.43)
	p-value	<.001	0.274	0.074	0.378	0.358	0.839
n		20,650	21,672	11,097	6,224	23,449	11,994

Table S69: **Replication of Table S1, Minimum cluster size = 300.** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	1.84	2.53	3.11	1.06	5.65	1.65
ethnic group	AME	+1.29***	+1.26***	+0.88**	+0.46**	+0.95*	+0.30*
	(SE)	(0.19)	(0.19)	(0.27)	(0.17)	(0.43)	(0.15)
	p-value	<.001	<.001	0.001	0.007	0.026	0.044
Wife own	Ref.	16.31	16.22	8.23	7.77	16.67	15.22
ethnic group	AME	-2.68***	-1.37***	-0.56*	-0.59	-1.76***	-2.43***
	(SE)	(0.36)	(0.33)	(0.28)	(0.32)	(0.42)	(0.59)
	p-value	<.001	<.001	0.044	0.065	<.001	<.001
Wife other group	Ref.	7.60	6.36	4.54	3.43	7.17	5.46
	AME	-0.18	+0.43	+0.57	-0.02	+0.60**	+0.37
	(SE)	(0.17)	(0.33)	(0.33)	(0.25)	(0.23)	(0.21)
	p-value	0.307	0.189	0.085	0.932	0.009	0.077
American wife	Ref.	-	-	-	2.71	-	4.84
	AME	-	-	-	+0.77***	-	+0.36
	(SE)	-	-	-	(0.16)	-	(0.31)
	p-value	-	-	-	<.001	-	0.246
Other outcomes	Ref.	3.84	3.52	2.86	2.66	2.58	2.47
	AME	-0.24	-0.17	+0.34**	+0.53*	+0.21	+0.14
	(SE)	(0.21)	(0.22)	(0.13)	(0.22)	(0.19)	(0.14)
	p-value	0.250	0.445	0.008	0.018	0.253	0.314
Not married	Ref.	70.42	71.36	81.27	82.38	67.92	70.37
	AME	+1.81***	-0.14	-1.22	-1.15*	-0.01	+1.25
	(SE)	(0.54)	(0.42)	(0.63)	(0.57)	(0.78)	(0.81)
	p-value	<.001	0.731	0.051	0.045	0.994	0.122
n		53,722	52,880	39,148	24,108	54,155	30,123

Table S70: **Replication of Table S2, Minimum cluster size = 300.** Each column displays the results of a separate multinomial regression for each dyad. "Ref." denotes the predicted probability in percentage points of each outcome for boys with two immediate ingroup neighbors. AME is the percentage-point change in that probability for boys living next to at least target one out-group neighbor. Robust city-clustered standard errors in parentheses. Stars display statistical significance: * p<0.05, ** p<0.01, *** p<0.001. n is the number of matched boys resulting from the CEM procedure.

Outcome	Statistic	Am. Boy Irish Neighbor	Am. Boy German Neighbor	Ir. Boy American Neighbor	Ir. Boy German Neighbor	Ger. Boy American Neighbor	Ger. Boy Irish Neighbor
Wife neighboring	Ref.	4.09	5.80	9.79	3.04	14.48	2.99
ethnic group	AME	+3.21***	+2.91**	+3.02***	+0.63	+2.26***	+1.05
	(SE)	(0.53)	(0.89)	(0.92)	(0.51)	(0.62)	(0.63)
	p-value	<.001	0.001	<.001	0.221	<.001	0.096
Wife own	Ref.	40.34	40.94	16.98	17.11	31.34	30.17
ethnic group	AME	-7.55***	-5.50***	+0.21	-1.55	-2.70***	-3.30**
	(SE)	(1.14)	(1.03)	(1.21)	(1.66)	(0.77)	(1.06)
	p-value	<.001	<.001	0.861	0.350	<.001	0.002
Wife other group	Ref.	18.49	15.28	13.13	9.70	15.73	13.76
	AME	-0.22	+1.61*	+0.41	+1.01	+1.43*	+0.54
	(SE)	(0.64)	(0.67)	(0.51)	(0.68)	(0.64)	(0.65)
	p-value	0.733	0.016	0.414	0.140	0.024	0.406
American wife	Ref.	-	-	-	9.34	-	12.92
	AME	-	-	-	+2.65***	-	+0.80
	(SE)	-	-	-	(0.69)	-	(0.64)
	p-value	-	-	-	<.001	-	0.207
Other outcomes	Ref.	5.95	6.03	6.05	6.14	4.10	4.09
	AME	+0.85*	+0.05	-0.41	-0.92	+0.00	+0.46
	(SE)	(0.42)	(0.41)	(0.28)	(0.61)	(0.23)	(0.35)
	p-value	0.041	0.909	0.142	0.131	0.989	0.188
Not married	Ref.	31.12	31.94	54.06	54.67	34.35	36.06
	AME	+3.71**	+0.94	-3.24	-1.82	-1.00	+0.45
	(SE)	(1.16)	(0.66)	(1.72)	(1.97)	(0.77)	(1.33)
	p-value	0.001	0.157	0.060	0.355	0.194	0.736
n		21,518	21,733	11,911	6,664	24,351	12,368

Table S71: **Distribution of treatment and control units after CEM.** Counts by dyads for the 1900 and 1910 linked samples. "Treatment" = at least one immediate target outgroup neighbor, "Control" = both immediate next-door neighbors belong to the ingroup.

		1900			1910		
Own Ethnicity	Neighbor Ethnicity	Treatment	Control	Total	Treatment	Control	Total
American	German	20,484	23,218	43,702	7,991	8,919	16,910
American	Irish	21,033	23,665	44,698	7,708	8,917	16,625
German	American	24,482	20,952	45,434	10,322	9,163	19,485
German	Irish	10,638	13,970	24,608	4,123	5,526	9,649
Irish	American	18,130	15,135	33,265	4,920	4,207	9,127
Irish	German	8,901	11,332	20,233	2,384	2,765	5,149
Total		103,668	108,272	211,940	37,448	39,497	76,945