Like a Good Neighbor: Childhood Neighbors Influence Occupation Choice*

Michael Andrews† Ryan Hill‡
Joseph Price§ Riley Wilson¶

July 25, 2026

Abstract

Occupation choice is a key determinant of economic mobility. This paper studies how childhood neighbors influence occupational choices and later-life earnings. Using linked historical U.S. census records for over six million boys and four million girls, we reconstruct neighborhood microgeography to identify childhood neighbors and their occupations. We find that living next door to someone in a given occupation increases the probability of working in that same occupation 30 years later by about 10 percent relative to other children on the same street. The transmission effect is higher with more intense exposure, in more connected neighborhoods, and among more similar families. Exposure to high-income or highly educated neighbors has lasting economic effects, leading to significant gains in adult income and education even compared to other children from the same street. These findings suggest that neighborhood networks play an important role in shaping intergenerational economic mobility.

Keywords: Neighborhood networks, peer effects, occupational transmission
JEL Codes: J24, J62, N32

---

*Thanks to Christian Ames, Thomas Barden, Carver Coleman, Maxwell Dunbar, Zach Flynn, Jimena Kiser, Josh Nicholls, Ethan Pixton, Abbie Sanders, Rebekah Todd, Margaret Truman, and Conner Wilkinson for outstanding research assistance. We would like to thank Patrick Bayer, Barbara Biasi, Sandra Black, Nicholas Carollo, Raj Chetty, John Friedman, Nathan Hendren, Song Ma, Stephen Ross, Carolyn Stein, Zach Ward, Melanie Wasserman, and Seth Zimmerman for helpful comments as well as audiences at Clemson; Northwestern; NYU; UC Merced; Towson; Yale; Mannheim; University of Southern Denmark; Rutgers; Oklahoma; Queens; San Diego State; University of New South Wales; Vanderbilt; Economic History Association Meetings; the NBER Mobility Conference; Opportunity Insights Conference; CEPR Applied Microeconomic History Workshop; Association for the Study of Economics, Religion, and Culture Conference; and the Cliometric Conference for helpful comments.

†University of Maryland Baltimore County, Email: mandrews@umbc.edu
‡Northwestern University, NBER, Email: ryan.hill@kellogg.northwestern.edu
§Brigham Young University, NBER, Email: joseph_price@byu.edu
¶Brigham Young University, CESifo, IZA, Email: riley_wilson@byu.edu.

1 Introduction

Growing evidence suggests that the neighborhood in which a child grows up influences their future earnings and economic mobility.[1] But why do neighborhoods matter? Prior work emphasizes neighborhood amenities—such as schools, infrastructure, safety, and health—that shape investments in children (Billings et al., 2014; Duncan et al., eds, 2018; Laliberté, 2021; Sharkey and Torrats-Espinosa, 2017). In this paper, we focus on another potentially important mechanism: the role of neighbors. In addition to amenities, neighborhoods foster person-to-person interactions that potentially generate human capital spillovers from peers, mentors, and role models. This mechanism is especially relevant for occupational choice, a key channel through which early-life environment can affect adult earnings and intergenerational mobility.

We identify the impact of adult neighbors on occupation choice by comparing the future occupations of children raised immediately next door to an adult in a given occupation to other children who live nearby but farther away on the same street. Our empirical strategy can be illustrated with a simple thought experiment: Suppose that a boy named Max lives next door to Dr. Johnson. His classmate Carl lives five doors away from Dr. Johnson on the same street. Is Max more likely to become a doctor than Carl? We scale this thought experiment to estimate exposure effects for all children who lived next door to doctors (and many other occupations) in the 1910 U.S. population census. Using machine-learning record linkage, we follow more than 10 million children, including 4 million girls, from childhood neighborhoods into adult careers. Because historical U.S. censuses were collected by enumerators going door-to-door, we are able to reconstruct local neighborhoods and identify each child's immediate neighbors. This microgeographic detail allows us to separate neighborhood selection from variation in next-door exposure.

[1] See Bergman et al. (2024); Chetty and Hendren (2018a,b); Chetty et al. (2026, 2016, 2014); Chyn (2018); Chyn et al. (2025); Haltiwanger et al. (2024); Katz et al. (2001); Kawano et al. (2024); Kling et al. (2007); Rosenbaum (1995).

Using narrow geographic fixed effects, we compare children whose families select into the same census manuscript sheet (typically a subset of a single street, about 8 households on average) but are exposed to next-door neighbors with different occupations. This strategy relies on the identification assumption that, while selection into neighborhoods may not be random, selection of immediate next-door neighbors is as-good-as-randomly assigned. We argue this assumption is plausible given the unpredictability of housing markets and the difficulty of observing occupations of specific neighbors before moving into a residence. We show that the degree of within-census sheet sorting on observable dimensions is not systematic across occupations and is small—too small to account for the occupation transmission effects.

We find that childhood neighbors matter for occupation choice. For example, boys who live next door to doctors in 1910 are 41% more likely to be doctors as adults in 1940 than are other boys residing on the same census manuscript sheet but farther away from the doctor. This magnitude is large relative to other transmission effects we present in the paper, but noticeably smaller than the effect of having a doctor in the child's own household. Next-door neighbor occupation transmission is not unique to doctors. Across the 50 largest non-farm occupations for men in the 1910 census, we find substantial heterogeneity in transmission effects, with positive point estimates from next-door neighbors for all but four occupations. The transmission effect is statistically significantly greater than zero in 26 out of the 50 occupations, and is never statistically significantly negative. Among the 25 largest occupations for women, we similarly find positive (albeit smaller in magnitude) effects for girls, despite lower statistical power due to lower female labor force participation in the early 1900s.[2]

Combining information on all of these occupations into a single "stacked regression," we estimate that, on average, a boy is about 10% more likely to enter an occupation when they live next door to an individual in that occupation than are other children on the same census sheet. A girl is about 6% more likely to enter their neighbor's occupation compared to other

[2]Because women's employment was highly concentrated in both 1910 and 1940, we restrict our analysis to fewer occupations.

girls on the same sheet. We show that these findings are robust to alternative specifications and sample restrictions to further address concerns of unobservable within-sheet sorting.

We next turn to mechanisms, and argue that a likely determinant of occupation spillovers is social connectedness. First, we show that occupation transmission monotonically decays for children in households farther away on the street. Boys are more likely to adopt their next-door neighbors' occupation in rural areas relative to urban areas, and in places where a smaller share of residents were born in other places. We also find strong patterns of homophily between neighbors. Boys are more likely to adopt their adult neighbor's occupation when they share the neighbor's birthplace, race, or gender, or when the families have children of similar ages. Although there are few measures of social connectedness available for the time period, we show that a proxy for social connectedness—the probability that children in nearby households are married in a future census—similarly decays for households farther away on the street, varies across these same dimensions (e.g., urbanicity, out-of-state share, and neighbor homophily), and is positively correlated with patterns of occupation transmission. Overall, these results suggest that children's career choices are influenced by their interactions with people in their local social networks.

Importantly, the occupation composition of childhood neighbors has real economic implications for both boys and girls. Boys are more likely to adopt the occupation of their next-door neighbor when it is a high-income occupation. Additionally, boys growing up next door to a neighbor in a high-income, high-education occupation, such as a doctor or lawyer, have significantly higher earnings as adults relative to other boys on the same census sheet. Meanwhile, growing up next door to a lower income occupation such as a porter, truck driver, or laborer results in lower income for children 30 years later. We find evidence that, in addition to making children more likely to enter their own occupation, neighbors also convey more general human capital to children. For example, growing up next to a teacher increases the probability of working in other high-education occupations like doctor and manager which have higher average incomes than teachers.

Our paper builds on the work of economic and social historians that exploits the structure of the historical censuses to understand social phenomena. The use of census sheet order to measure residential racial segregation was pioneered by Agresti (1980) and Logan and Parman (2017a),[3] and detailed census geography, such as census block and page, have been used to study neighborhood-level hiring networks (Bayer et al., 2008), industry choice (Tan, 2022), and income shocks on homeownership (Quincy, 2022). Like some of the existing work, we focus on extremely local geographic differences in exposure, comparing next-door neighbors to those living on the same census sheet, minimizing concerns about unobserved sorting within neighborhoods. Perhaps the most similar research design to ours is Bayer et al. (2025), which uses modern mortgage and housing transaction data to compare move rates for homeowners who receive a next-door neighbor of a different race to others farther away on the same block. The richness and comprehensiveness of the historical census data allow us to investigate outcome variables, dimensions of heterogeneity, and alternative identification strategies not possible with most contemporary datasets.[4] We can also link highly detailed microgeographic data on children to their adult outcomes 30 years later.

This work adds to our understanding of neighborhood effects, intergenerational mobility, and occupation choice. Although connectedness is a strong predictor of intergenerational mobility (Chetty et al., 2022), there is limited causal evidence on this mechanism. We are able to provide causal estimates of how adult neighbors affect children's long-run trajectories. Most studies of adults' influence typically investigate the correlation between parents and children's later career outcomes (Bell et al., 2019; Corak and Piraino, 2011; Fairlie and Robb, 2007; Hvide and Oyer, 2018; Staiger, 2023). Studies of parents' influences, however, are often unable to disentangle the effects of environment and informational spillovers from genetic

[3]These measures have subsequently been used to study the relationship between residential racial segregation and lynching (Cook et al., 2018), homeownership (Logan and Parman, 2017b), mortality (Logan and Parman, 2018), and present-day neighborhood-level economic mobility (Andrews et al., 2017). Similar techniques have been used to measure residential segregation of immigrants (Eriksson and Ward, 2019).

[4]The mortgage data include only limited demographic markers, and cannot address broader questions of occupation transmission.

endowments and intra-family transfers such as inheriting a family business.[5] By estimating the effect of next-door neighbors, we remove these last two channels and investigate the importance of information and exposure spillovers.[6] This paper also joins a recent literature focused on the role of information, exposure, and mentoring in children's (especially girls') decisions to enter high-income and high-education careers (Andrews and Zhao, 2025; Bell and Petkova, 2023; Breda et al., 2023; Chung, 2020; Fruehwirth and Gagete-Miranda, 2019; Laitila et al., 2024; Mertz et al., 2025; Olivetti et al., 2020; Porter and Serra, 2020). Finally, we contribute to a growing literature on the importance of microgeography in facilitating face-to-face social interactions (Andrews, 2019; Andrews and Lensing, 2024; Arzaghi and Henderson, 2008; Atkin et al., 2022; Catalini, 2018); this literature shows how these interactions can influence the flow of ideas to produce new innovations, but has not established how variations in the microgeography can influence long-term labor market decisions.

The paper proceeds as follows. Section 2 discusses the census data, in particular how we construct links from children in 1910 to adults in 1940 and how we construct measures of microgeographic proximity. Section 3 describes our regression framework and discusses our identification strategy. Section 4 presents our baseline results for doctors, results for the 50 largest male occupations and the 25 largest female occupations separately, and for the stacked regressions. Section 5 explores mechanisms of transmission, focusing on the intensity of exposure, homophily within neighborhoods and between neighbors, and the strength of social ties. Section 6 investigates how exposure to next-door neighbors affects economic outcomes of children when they are adults. Finally, Section 7 briefly concludes.

[5]Some exceptions are Dal Bó et al. (2009) and Greenberg et al. (2024), which use regression discontinuities in vote shares and AFQT scores to isolate plausibly exogenous variation in the parent's probability of being a politician or serving in the military, respectively.

[6]Other studies argue that neighbors affect a host of household decisions, including employment, criminal activity, mortgage lending, and consumer purchases (Aliprantis, 2017; Bayer et al., 2008; Billings et al., 2024; Grinblatt et al., 2008; Hellerstein et al., 2011; McCartney and Shah, 2022; Redding and Sturm, 2024; Shah and McCartney, 2023; Tan, 2022).

2 Data

Our analysis requires two features: identifying children's neighborhood microgeography, and linking children to adult occupational outcomes. Most existing studies lack at least one of these features. Traditional surveys lack detailed information about neighbors. Although recent advances in census record linkage using the Personal Identification Key (PIK) have expanded available data, they either restrict analysis to survey samples or rely on PIK-linked full-count censuses that do not span a sufficiently long period to observe children into adulthood. Anonymization protocols in contemporary administrative data also make it difficult to reconstruct the microgeography of neighborhoods.

For this reason, we exploit individual-linked full-count census data from 1910 to 1940. The US Census Bureau releases personally identifiable information from the decennial census after 72 years. Digitized versions of the original census sheets, filled out by hand by Census enumerators, are publicly available and include individuals' names, sex, birth year, birth place, and occupation. Each individual's information is recorded on a census line with a family identifier, thus allowing us to connect together members of a family at a given point in time.

We use three main data sources: the 1910 full-count census, the 1940 full-count census, and the Census Tree database of individual links.

2.1 Full-count Census Data

We obtain full-count census microdata for 1910 through 1940 from Ancestry. This includes all of the digitized information on a census sheet including state, city, enumeration district, street address (when recorded), household id, census sheet number, census line number, name, sex, relation to head-of-household, marital status, year of birth, place of birth, employment status, and occupation. We merge the de-identified full-count census data to IPUMS full-count census records using historical IDs to work with cleaned, categorized measures of occupations. Using the 1940 census, we can observe each child's own occupational choice as an adult, 30 years later.

2.2 Census Tree Links

The Census Tree is a dataset that provides over 700 million links for individuals across historical US census records. This dataset was originally created by Price et al. (2021) and further expanded and refined in Buckles et al. (2025). The Census Tree builds on family tree data from one of the world's largest internet genealogy platforms, FamilySearch.org. FamilySearch users add and attach historic records to the public profiles of their own relatives. These profiles are connected together through family relationships into a large interconnected network of profiles called the Family Tree. FamilySearch users often have private information that allows them to link records that even a trained research assistant or machine learning algorithm could not link. The Family Tree is used to identify links between census records by looking for pairs of census records that are attached to the same profile. The census-to-census links for men alone number more than 158 million, greater than the number of conservative links in the Census Linking Project (Abramitzky et al., 2022).

The Census Tree uses the hand-linked records from the Family Tree as training data to develop a machine learning algorithm to identify additional linkages. The Census Tree then combines these new machine learning links with links from the Census Linking Project (CLP) and the Multigenerational Longitudinal Panel (MLP), as well as the Family Tree links and a set of machine learning links created by FamilySearch. One key innovation of the Census Tree is to combine together links from multiple methods and then use a set of decision rules to adjudicate disagreements across the different linking methods. The final result is a dataset with 391 million links for men and 314 million links for women across the 1850 to 1940 censuses.

To construct our analysis sample we start with the universe of children between the ages of 5 and 18 in the 1910 full-count census. The children observed in 1910 are then linked to outcomes in the 1940 census, when they are between 35 and 48 years old. In the 1940 census we observe the individual's occupation, along with other outcomes, such as employment, educational attainment, and wage income. To link individuals from 1910 to 1940 we use all available Census Tree links, allowing us to link 3-4 times as many boys and girls as alternative linking methods (Buckles et al., 2023). This yields a sample of approximately 10.5 million, including 6,346,719 men and 4,182,461 women. As seen in Table 1, this represents 40.2 percent of all children observed in 1910. Table 1 shows that our linked sample comes from household settings that are broadly representative of the full population, although our linked sample has relatively more White individuals than the full population due to the difficulty linking Black individuals during this time period.

Because women historically change their last name when they get married, traditional linking methods perform poorly when linking women. For this reason researchers often exclude women or focus on subsamples of women that are easier to follow over time. Our ability to evaluate outcomes for girls is therefore an innovation relative to the prior literature. The patterns and opportunities for employment were quite different for boys and girls in 1910, however. In 1910, only 29.3 percent of women 18-54 were in the labor force, compared to 95.5 percent of men. These gendered patterns were largely unchanged by 1940, when the labor force participation rate of women 18-54 was 31.7 percent compared to 90.7 percent for men. Given these patterns, we view the occupational transmission effects for boys and girls as both interesting, but distinct, questions. As such, the set of occupations we consider and the estimated effects will be gender-specific to correctly interpret patterns in the historical context.

2.3 Measuring Neighbors in the Full-count Census

Our empirical approach relies on identifying geographically proximate households using the 1910 census records. Census enumerators were given explicit instructions on the order in which they were to collect and record household information: "It is your duty *personally* to visit every family and farm within your district... Canvassing of blocks should go in order around the block, not switching back and forth across the street" (U.S. Bureau of the Census, 1910, emphasis in original). As specified in the 1910 enumerator instruction guide, the intent was to record families on a census sheet *neighbor-by-neighbor*. Because the enumerator recorded households in the sequence they encountered them while canvassing, the order of households on the census sheet is a good proxy for physical distance between households. We use information on the enumeration district, sheet (i.e., page) number, and line number to identify likely neighbors with fine geographic precision.

Although we estimate occupational transmission for many occupations, for clarity we describe our data creation process focusing first on doctors only.[7] First, we flag all households across all 1910 census sheets that include a child age 5-18 and all households across all 1910 census sheets that include an individual with the target occupation: physicians and surgeons (1950 occupation code 075). We then collapse the data to include one observation per household, preserving the geographic measures (city, state, enumeration district, census sheet number) and the within-sheet ordering of households. If household members spread across multiple sheets we reassign them to the census sheet of the household head. For each household we then construct two binary indicators. The first equals one if there is a doctor in one's own household. The second equals one if there is a doctor on the same sheet and one household above or one household below one's own household ("next door"). For our baseline analysis we do not restrict the gender of the next-door neighbor in the target occupation; we explore gender homophily in Section 5. We repeat this process for each target occupation separately. We focus on the 50 largest occupations in terms of 1910 male workforce for boys and the 25 largest occupations in terms of 1910 female workforce for girls. In both cases we exclude farmers and farm laborers.[8]

[7] Several studies use doctors as a case study of intergenerational transmission of occupations from parents to children (Lentz and Laband, 1989; Polyakova et al., 2020; Ventura, 2025).

[8] We exclude farmers from this set of occupations for four reasons. First, farmers often live in communities where most households have the same occupation; it is likely that farmers who live on a street with no other farmers, for instance a lone farmer in an urban neighborhood, are very different from the average farmer. Second, the nature of farming changed substantially between 1910 and 1940. Third, farmers often lived on

Using the written ordering on census sheets could introduce measurement error when identifying next-door neighbors if, for example, enumerators did not follow instructions or families were not home when the enumerator stopped by. We describe how we use census sheet order to reconstruct neighborhood microgeography, potential sources of measurement error, and various ways to address potential measurement error, including using address-based measures when available, in detail in Appendix B.[9] Overall, error in our measurement of next-door neighbors is limited. For a random sample of the subset of sheets with transcribed street addresses, we validate census sheet order using historical Sanborn Fire Insurance maps (Sanborn Map Company, Various Years). For this hand-checked subsample, we find that the sheet-order based next-door neighbor measure of whether a doctor lived next door is correct 91 percent of the time (see Appendix B for more detail on this process).[10]

3 Empirical Approach

To fix ideas, we introduce our empirical approach with an illustrative example of a single occupation: doctors. We test whether a boy living next door to a doctor in 1910 is more likely to become a doctor by 1940 relative to other boys on the same census sheet. With this simple framework in hand, we then broaden the analysis to study the transmission of other occupations. We also introduce a stacked regression framework that allows us to test average transmission effects across occupations in a unified approach.

Our baseline estimating equation is as follows

$$\text{Doctor1940}_i = \alpha + \beta_0 \text{OwnDoc1910}_i + \beta_1 \text{NextDoorDoc1910}_i + \delta_{\text{a}(i)} + \gamma_{\text{s}(i)} + \epsilon_i, \quad (1)$$

large farms, and distance between households may have impeded social interaction. Finally, farmers often report no wage income making it difficult to study effects on household income in Section 6. We show that our baseline results are robust to including farmers in Appendix Table A1.

[9] About 35% of households in the 1910 census lack street address information; in contrast, census sheet order is available for all households.

[10] Using the street address-based measure of proximity to correct for non-classical measurement error in the census sheet-order analysis as suggested by Pischke (2007) suggests that, if anything, our next-door neighbor transmission results are somewhat attenuated (see Appendix B).

where OwnDoc1910_i is an indicator equal to one if there is an adult doctor living in the same household as i and NextDoorDoc1910_i is an indicator equal to one if one of the households next door to i has an adult who is a doctor. $\delta_{\text{a}(i)}$ is a fixed effect for the focal child's age in the 1910 census. Finally, $\gamma_{\text{s}(i)}$ is a fixed effect for i's census sheet, and ϵ_i is an idiosyncratic error term that includes all unobservable local amenities and individual- and household-level characteristics. We cluster standard errors at the household level to account for common treatment across siblings. The inclusion of census sheet fixed effects is key to our identification strategy, allowing us to compare households in the same narrow geographic location that have different next-door neighbor occupation exposure. We discuss our identification assumptions further in the next section.

To generalize this doctor regression to other occupations, we replace the doctor next door and own household indicators with an indicator for a target occupation living next door or in the same household in 1910. We then replace the outcome with an indicator for the focal child reporting that target occupation in 1940. For example, we can test whether a girl living next door to a teacher in 1910 is a teacher in 1940.

We also implement a stacked regression approach, which allows us to systematically explore average effects, patterns of heterogeneity, and causal mechanisms later in the paper. To do this, we take each occupation-specific regression sample described above and stack them, so that all 50 occupations for boys appear in one regression sample and all 25 occupations for girls appear in another regression sample. We can also restrict to specific subsets of occupations, such as all high-income occupations, to explore dimensions of heterogeneity. The unit of observation in the stacked regressions is a child-occupation pair. We estimate

$$\text{Occ1940}_{ij} = \alpha + \beta_0 \text{OwnOcc1910}_{ij} + \beta_1 \text{NextDoorOcc1910}_{ij} + \delta_{\text{a}(i)\times\text{occ}(j)} + \gamma_{\text{s}(i)\times\text{occ}(j)} + \epsilon_{ij}, \tag{2}$$

where j indexes each occupation. In this specification, each child i is included in the regression multiple times, one for each occupation that occurs on the child's sheet.[11] Since the same household appears multiple times for each occupation on their sheet, we continue to cluster standard errors at the household level to account for common treatment across siblings and dependence in occupation choices at the individual level. Our fixed effects for age and census sheet are interacted with occupation to make the same within-occupation comparisons that we made in the single occupation specification.

3.1 Identification

In order to identify a causal occupation transmission effect, we assume that while households might select into a particular neighborhood, their specific next-door neighbors are as-good-as-randomly assigned. Given the limited availability of housing at a point in time and imperfect information, it is difficult for individuals to residentially sort based on the occupations of potential next-door neighbors. In other words, conditional on living in the same small portion of a street, a household's proximity to a neighbor in a particular occupation is plausibly exogenous to children's future occupation outcomes.

To formalize this idea, consider a stylized, linear model of a child's occupational choice:

$$\text{Occ1940}_i = \alpha \text{Amenities}_{s(i)} + \beta \text{NextDoorOcc1910}_i + \nu_i, \tag{3}$$

where Occ1940_i is an indicator equal to one if a child in household i has a particular occupation in 1940. $\text{Amenities}_{s(i)}$ capture any local characteristics, such as the quality of the public school district, that may change the probability that child i enters a particular occupation as an adult. NextDoorOcc1910_i is equal to one if an adult living next door to child i in 1910 is in that occupation. ν_i captures any individual- or household-level idiosyncratic characteristics that affect the probability that child i enters a particular occupation.

[11] Because we include sheet-by-occupation fixed effects, the estimated effects are identified off of sheets where there is variation in OwnOcc1910_{ij} and $\text{NextDoorOcc1910}_{ij}$. As such, we can eliminate child observations on sheets without people in the target occupation without affecting the identifying variation. This reduces the overall size of the stacked sample from over 316 million possible individual-occupation observations for the 50 male occupations to 26.6 million observations.

We are interested in estimating β, the causal effect of exposure to adult neighbors on occupation choice. The challenge is that we do not observe ν_i nor the full set of local amenities, both of which are likely correlated with the occupations of nearby adults. For an extreme example, children growing up in coal-mining towns are more likely to live next door to a coal miner and also more likely to become coal miners themselves, even absent any true exposure effect, because proximity to coal mines and related local conditions jointly shape residential patterns and occupational choices. This illustrates the broader reflection problem articulated by Manski (1993).

We argue that the ability to compare individuals over tightly specified geographies resolves the reflection problem in this setting. Individuals who reside on the same census manuscript sheet (on average, only 7.6 households) are exposed to the same local amenities, and only have limited ability to choose their immediate next-door neighbors based on their occupations. Therefore, conditional on being on the same manuscript sheet, all households should be equally likely to be immediately next door to a neighbor in a given occupation. Our identification assumption is therefore that

$$\mathrm{Cov}(\mathrm{NextDoorOcc1910}_i, \mathrm{Amenities}_{s(i)}|s(i)) = \mathrm{Cov}(\mathrm{NextDoorOcc1910}_i, \nu_i|s(i)) = 0, \quad (4)$$

where we condition on i's census manuscript sheet $s(i)$. Similar identification has been used in housing transaction data in a modern context (Bayer et al., 2025).

3.2 Testing the Identification Assumption

Our identification assumption fails if households sort *within census manuscript sheets* in ways that are correlated with both their neighbor's occupation and their child's proclivity to work in that occupation as an adult. In the data, this would likely result in individuals in the same occupation living next to each other or individuals in similar occupations (along dimensions such as industry, skill-level, or income) living next to each other in a non-random fashion. We document the degree to which households sort along both of these dimensions.

We assess within-sheet occupational sorting directly and find it to be limited. In Appendix Figure A1, we show that for the 52 largest occupations in the 1910 census (including farm owners and farm laborers), among the sheets that contain at least one individual with each occupation, 65% of sheets contain only one occurrence of that occupation. With the exception of geographically constrained resource-based occupations like mining, most people are the only person in their occupation on their census sheet.[12] In addition, we show in Appendix Figure A2 that occupation-specific door-to-door segregation indices, constructed analogously to the racial segregation measures in Logan and Parman (2017a), are near zero for nearly all occupations and substantially smaller than racial or ethnic segregation, further supporting the view that immediate neighbors are close to as-good-as-randomly assigned.[13]

Even though people in the same occupation tend not to sort next-door to each other, a more nuanced form of within-sheet sorting may still occur on dimensions that are correlated with potential occupation choice, such as income and education. For example, suppose a neighborhood street adjacent to a city park has house values that are correlated with distance from the park within the same street. Since doctors and lawyers have high income, they might be more likely to cluster near the park, and thus be more likely to live next door to one another. In this case, lawyers' children might be more likely to become doctors, not due to a neighbor transmission effect, but because proximity to a doctor reflects underlying socioeconomic selection even within streets.

We test for this degree of selection in Table 2. We report means of household characteristics that could reasonably be correlated with potential occupation choice for children who live on the same census sheet as a doctor (Panel A), a teacher (Panel B), or a brickmason (Panel C). For this table, we put all children in mutually exclusive groups based on the distance to that particular occupation. Column 1 reports means for children with someone in

[12]Consistent with minimal same-occupation sorting at this microgeographic scale, our estimates are unchanged when we restrict attention to sheets with exactly one occurrence of each occupation, which mechanically rules out same-occupation adjacency sorting.

[13]The exception is once again for resource-based occupations like farmers, miners, and lumbermen.

that occupation in their own household, Column 2 reports the mean for children immediately next door to someone in the occupation, and Column 3 reports the mean for other children who live further away on the same census sheet. In Column 4, we report the difference between Columns 2 and 3 and provide the standard error for the difference in Column 5.[14]

Doctors have occupation scores and education scores that are significantly larger than other families on the same sheet. Since we control for own household occupation indicators in our regression specification, this does not affect our identification assumption. The key comparison is between next-door neighbors and farther neighbors reported in Column 4. We find evidence of some sorting along several dimensions, but the magnitudes are economically small. Next-door neighbors of doctors have about $100, or 5%, higher occupational income scores than farther neighbors. But this difference is very small compared to the $4,564, or 188%, higher occupational income scores for doctors than their farther neighbors. Along other dimensions, like education scores, the probability of being white, or the probability of being a US native, averages for next-door neighbors and neighbors further away are very similar. Other occupations are even more balanced; for example, next-door neighbors of teachers have very small differences in occupational income scores (1%) relative to farther neighbors, but have larger differences in education scores (7%); again, these differences are swamped by the difference between the teachers themselves and the farther neighbors (107%). Lastly, brickmasons show strong balance between next-door and farther neighbors across most characteristics, with no difference being more than a few percentage points. Differences in Column 4 are often statistically significant due to the large sample size, but the differences are always small in magnitude.

These correlations between household characteristics and neighbor proximity only bias our causal estimates if they are correlated with the outcome. To assess this directly, we

[14] In our main regression specification, we allow children to have own household and next door indicators turned on simultaneously in cases where two adjacent houses hold the same occupation. In this balance table, we put each child in this situation into the own household category only for ease of interpretation. In Appendix Table A2, we show a version of this table where children are allowed to appear in both Columns 1 and 2, and in Appendix Table A3, we restrict to sheets that have exactly one person in that occupation per sheet. We find very similar patterns in each of these alternative versions.

use observable characteristics from the census to predict occupation choice for each of our 50 occupations. Because of the large number of potential covariates, we implement this prediction using a lasso model. For each occupation, we regress an indicator for reporting that occupation in the 1940 census on household and individual covariates from 1910, including household head occupation score, education score, home ownership status, race and nationality indicators, marital status, family size and member ages, as well as indicators for household head's occupation and industry. We do not include any measures of occupation proximity, census sheet fixed effects, or geographic fixed effects. We run this lasso prediction separately for each occupation, allowing different features to be selected for each job. This allows us to determine the extent to which 1910 household and individual observables predict being in a particular occupation in 1940. In Table 2, we also provide the mean of the lasso predictions for each occupation. As expected, the predicted values are very high among children of doctors, teachers, and brickmasons. However, consistent with our identification assumption, the lasso predictions are nearly identical between next-door and farther neighbors on the sheet. Below in the robustness section (Section 4), we compare the coefficients from a regression where the predicted occupation is the outcome to the coefficients from the regression where the realized occupation is the outcome, and find that for all occupations, living next door to someone in a particular occupation has a much smaller effect on the predicted occupation than the realized occupation. We also control directly for the prediction and find that the main results are affected very little, suggesting within-sheet residential sorting is not a primary driver of our results.

4 Results: Occupation Transmission

We present our main results in three steps: First we estimate transmission for the doctors case study. Second, we estimate transmission for the largest occupations for boys and girls individually. Third, we estimate the average transmission effects across all occupations in our stacked regression framework. Following the main results, we explore the robustness of our estimation strategy.

4.1 Doctors

Table 3 shows results for boys in the doctor regression described in Equation (1). To highlight the role of narrow geographic fixed effects, we present different levels of fixed effects in each column. Column 1 does not include any location fixed effects and includes all households with a child in the 1910 census. In Column 2, we run the same regression in a sample restricted to sheets with at least one doctor. Column 3 adds fixed effects for the enumeration district, and Column 4 adds our preferred, more narrow sheet-level fixed effects. Across all four specifications, we find a stable and positive effect of within-household transmission. Boys who grow up with a doctor in the home are 9.54–9.83 percentage points more likely to become doctors than boys without a doctor in the home. Relative to the probability of becoming a doctor for the untreated group, boys with doctors in the household are at least ten times more likely to become doctors than their peers on the same sheet.

Our key coefficient of interest is the effect of living next door to a doctor in 1910. For this estimate, the level of fixed effects is relevant for our identification assumption. The estimate with no fixed effects includes not only the causal effect of exposure but also a selection effect where children who live in the same neighborhood as doctors may have been more likely to become doctors even absent exposure to their next-door neighbors. We see that the coefficient in Column 1 of 0.0072 falls to 0.0036 in Column 2 when considering only sheets with doctors, and falls slightly further to 0.0033 in Column 3 with enumeration district fixed effects. In Column 4, with sheet-level fixed effects, the estimate is 0.0032. This estimate removes the street and neighborhood-level sorting, and we interpret this coefficient as the causal transmission effect. Boys who live next door to a doctor are 0.32 percentage points more likely to become a doctor, or 41% more likely, than boys who live farther away on the same sheet.

4.2 All Occupations Separately

We next consider neighbor transmission effects across a broader set of occupations, including occupations available to women in the time period. To simplify the analysis and to ensure that we have a sufficient number of individuals in each occupation for statistical precision, we focus on the 50 largest non-farm occupations in terms of male employment in 1910, and the 25 largest non-farm occupations for women.

Own household and next door estimates are plotted for boys and girls for each occupation separately in Figure 1. In both cases, we scale the coefficients by the mean among children with no individuals in the occupation in their household or next door so that the plotted coefficients can be interpreted as proportional changes relative to the untreated mean.[15] Both are separately sorted by effect size. Solid filled markers indicate that the coefficient is statistically significant at the 5% level. We find that, for all 50 of the largest occupations for men, having an individual in the same household significantly predicts that the child is more likely to enter that occupation. Own household transmission for girls is positive and statistically significant for most occupations, but the effect sizes are much lower than for boys. Girls choose the occupation of a parent one to five times as often as an untreated peer.

Focusing on the next-door coefficients, we see that most occupations show positive transmission effects from next-door neighbors. In contrast to the own-household coefficients, however, not all are statistically different from zero and three (stationary firemen, doorkeepers, and shipping and receiving clerks) are negative, although not statistically significant. The highest transmission effect is for brickmasons—both for next-door and own household exposure—although the next several most transmissive occupations differ between the two lists. Doctors also rank highly for next-door transmission, in eighth place. Girls show more mixed evidence on next-door neighbor transmission. While the majority of the top-25 occupations have positive transmission, only six are statistically significant. Although the results for girls are suggestive, we lack the same level of statistical precision as we do for boys.

[15] We plot the un-transformed coefficients in Appendix Figure A3.

4.3 All Occupations Combined

In Table 4, we present results from the stacked regression specification defined in Equation (2). Panel A shows average effects across the 50 largest occupations for boys, and Panel B shows average effects across the 25 largest occupations for girls. As in Table 3, Columns 1 and 2 do not include location fixed effects (although we still include a fixed effect for each occupation j). Recall that Column 1 includes all census manuscript sheets, while Column 2 includes only sheets that have at least one individual of each occupation j. In contrast to the results when examining only doctors, in the stacked regression coefficients for both own-household effects and next-door neighbor effects become substantially smaller as we include geographically smaller location fixed effects. This is consistent with substantial locational sorting, on average, for the occupations in our sample.

Our preferred specification in Column 4 includes a fixed effect for each census manuscript sheet interacted with a fixed effect for each occupation. In this specification, growing up with an adult in the same household in an average occupation increases the probability that the boy enters that occupation by 3.81 percentage points, or about 115%. Growing up next door to an individual in a particular occupation increases the probability that the boy enters that occupation by 0.34 percentage points on average, or about 10.3%. For both the own-household and next-door coefficients, the scaled estimates in Table 4 are smaller than those for doctors in Table 3, which is consistent with our findings in Figure 1 that transmission effects for doctors are larger than for an average occupation. For girls, shown in Panel B, the effects are also positive and significant, but smaller in percentage point change. Girls are 0.85 percentage points, or about 54%, more likely to choose their own parent's occupation, and 0.1 percentage points, or about 6.4%, more likely to choose their next-door neighbor's occupation. The relative increase is about half that for boys from a much lower base. This smaller effect size likely reflects the changing nature of female employment during this time period, and the lower overall propensity to participate in the labor force at all compared to men.

4.4 Robustness

We next implement several additional strategies to minimize concerns that results are driven by unobservable within-sheet sorting, summarizing results here and presenting full results in Appendix Table A4 for doctors and Appendix Table A5 for the stacked sample.

One concern is that households may be able to sort to live close to amenities, and that exposure to these amenities may vary even within a sheet. For example, households with higher incomes may purchase houses closer to a local park on a residential street. Since enumerators begin a new sheet when moving to a new area, if these amenities are more likely to be located at the beginning or end of a street, then they could be correlated with a household's position on the census manuscript sheet. We show that our results are robust to including sheet position fixed effects (i.e., a fixed effect for the first household on each sheet, the second household, etc.). We also show that results are robust to comparing children only to households three doors away or less; by restricting attention to small subsets of sheets, we compare across households with similar exposure even to highly localized amenities. Group quarters such as dormitories or barracks bring a specific set of local features that may affect both exposure to neighbors and occupational outcomes; results are robust to excluding any households living in group quarters.

Even when taking steps to ensure that households are exposed to the same local amenities, households with adults in the same occupations may still be able to sort to live next to one another. Our results are robust to restricting attention to sheets with only one occurrence of each occupation per sheet, which mechanically prevents this kind of sorting. We also separately model the transmission effect when a member of the child's own household and next door neighbor have the same occupation; the interaction term is negative and of comparable magnitude to the next door neighbor coefficient, suggesting that having a next door neighbor in an occupation does little to encourage an individual to enter that occupation above and beyond the effect of an adult in the child's own household.

Even if households in the same occupation do not sort to live next to one another, households may be able to sort on other characteristics. For example, a household with a high income occupation like a doctor might be able to select to live next to a household with another high income like a lawyer. To ensure that this is not driving our results, we control for a wide battery of household head characteristics, including the household head's occupational income and education scores, age, race, and nativity. We additionally include household head occupation fixed effects in addition to controlling for other household head characteristics. Our results remain robust.

To directly estimate the extent to which sorting on observable household characteristics explains children's occupational choices, we build on the analysis described in Section 3.2. For each occupation, we regress an indicator for choosing that occupation on 1910 household characteristics in a lasso model. Each child therefore has a predicted probability of entering all large occupations that appear on their sheet. In Figure 2, we report coefficients from a regression of the predicted occupation probability on the next-door neighbor indicator to directly compare the neighbor's influence on the predicted vs. realized occupation choice. We find that next-door neighbor occupation has only a very small correlation with predicted occupation, but a much larger correlation with realized occupation for most jobs. This pattern suggests that ex-post occupation choice is influenced by neighbors even relative to what we would predict with family characteristics. Controlling directly for the lasso prediction, as a summary measure of occupation-specific sorting, has little impact on our next door coefficients (see also Appendix Table A6).

Finally, we show that our main results are robust to using data from different years or estimating transmission effects with alternative functional forms. In Appendix Tables A7 and A8, we repeat our baseline analyses but link children to the 1920, 1930, and 1940 censuses. For both own-household and next door coefficients, effect sizes after scaling by the untreated mean are similar across all three decades, suggesting that we do not lose out on additional information by focusing on 1940 outcomes in our baseline results. In Appendix Figure A4, we show that results are robust to logit specifications, which account for concerns about sparse outcomes when using linear probability models.

5 Why Neighbors Matter: Social Ties and Homophily

What drives the strong patterns of occupation transmission? In this section, we present evidence that social ties are a likely determinant of the transmission of economic outcomes within neighborhoods. First, we introduce a measure of social connectedness and document how it varies across demographic dimensions. Then we compare the intensity of social connection to rates of occupation transmission.

Unlike modern settings, in which it may be possible to directly measure social connectedness (e.g. using online social platforms), there are few proxies for connectedness available in historical data. We focus on one useful proxy facilitated by linked censuses: marriage rates between childhood neighbors. We observe children who appear on the same census sheet in 1910 and mark pairs of children who appear as a married couple in the 1940 census. Although we cannot observe day-to-day social interactions of children living in 1910, higher rates of marriage later in life suggests more intense social interaction between the families.

To calculate marriage rates, we take all linked boys on the 1910 sheet and consider every girl living near them as a potential future spouse. We then link forward to the 1940 census and mark each of those pairs that reported being married at that time. We can then calculate a marriage rate for any subset of sheets (e.g. sheets in urban vs. rural areas) or characteristic of the pair (e.g. families with the same vs. different nationality).[16]

5.1 Neighbor Proximity

First, we show that social connections decline with physical distance between houses in a neighborhood. We find that 0.078 percent of potential next-door neighbor pairs are married in the 1940 census. That probability falls to 0.069 percent two doors away, an 11 percent

[16] Because we cannot link all individuals to the 1940 census, we likely underestimate the true marriage rate, but this should affect estimates similarly across different dimensions.

drop, and continues to fall monotonically to 0.053 percent five doors away, a 32 percent lower rate than immediate next-door neighbors (see the first five rows of Figure 3).

We compare this relationship between distance and social connections to the rate of occupation transmission between more distant neighbors on the same sheet. To estimate this, we extend our baseline stacked regression to include indicators for additional households farther away from a particular occupation j on the same census sheet:

$$\text{Occ1940}_{ij} = \alpha + \beta_0 \text{OwnOcc1910}_{ij} + \sum_{d=1}^{5} \beta_d \text{NeighborOcc1910}_{ijd} + \delta_{\text{a}(i)\times\text{occ}(j)} + \gamma_{\text{s}(i)\times\text{occ}(j)} + \epsilon_{ij}, \tag{5}$$

where the indicators $\text{NeighborOcc1910}_{ijd}$ are equal to one if at least one of the households d steps away from the focal child in the 1910 census sheet has an adult that lists occupation j. For example, if $\text{NeighborOcc1910}_{ij3} = 1$ in the doctor section of the stack, there is a doctor exactly three doors away from the focal child. Children who appear on the same census sheet but are more than five houses away are the omitted group.

Figure 4 plots each β_d. We see a monotonic and steep decline in the transmission rate as we move from immediate next-door neighbor towards five houses down the street.[17] In our baseline results, children 2–5 households away from the adult neighbor are part of the set of counterfactual children; while these children have smaller exposure effects than the children living next door, the effects are still positive, and so our baseline results should be seen as a lower bound on the magnitude of exposure effects. Comparing Figures 3 and 4, there is a clear correlation between social connectedness and occupation transmission. The fact that both figures exhibit a similar pattern of decay suggests that social ties are an important mechanism through which neighborhoods affect economic outcomes.[18]

[17] To estimate Equation (5), we require census sheets with at least six households (the household with an adult in occupation j and five neighbors). Sheets with several large households may have fewer than six households on the sheet, and so looking at the effects on children farther away mechanically reduces the sample size. For this reason, our main analysis in Section 4 focuses only on immediate next-door neighbors.

[18] In Appendix B, we introduce an alternative method to measure neighborhood microgeography, relying on recorded street addresses rather than census manuscript sheet order. We find similar patterns of decay

5.2 Neighborhood Characteristics

To further study how the connectedness of neighborhoods affects the magnitude of occupation transmission, we examine three types of neighborhood heterogeneity: urban and rural, high and low immigrant share, and high and low out-of-state share. Neighborhoods that are less transient and have more permanent residents (and so have a smaller share of immigrants and individuals born out-of-state) are likely to have stronger social connections, which may facilitate easier transmission of occupations across neighbors.

In Figure 3, we validate these conjectures about the strength of social connections in different types of neighborhoods using the differences in marriage rates. While the overall marriage rate for next-door children is 0.078 percent, there is striking heterogeneity by neighborhood type. The marriage rate is six times higher in rural neighborhoods compared to urban neighborhoods. Similarly, the marriage rate is 76 percent higher in neighborhoods with below-median immigrant share compared to above-median immigrant share. Finally, the marriage rate in neighborhoods with below-median out-of-state birthplace share is 86 percent higher than above-median out-of-state birthplace share.

Turning to occupation transmission, we estimate next-door neighbor regressions for boys using the stacked specification in Equation (2). We split the sample by county-level urban or rural status, immigrant share, and out-of-state share and report scaled coefficients in Figure 5. We find that the transmission effect is significantly larger in neighborhoods that are in rural counties, that have lower immigrant shares, and that have more residents who were born in the state (low out-of-state share). These patterns are consistent with more homogeneous places with stronger connectedness facilitating occupational transmission.[19]

using street addresses. Using street addresses also allows us to explore effects for multi-family dwelling, like duplexes. Individuals who reside in the same multi-family dwelling likely have more social interactions than even individuals living in an adjacent dwelling. We find that occupational transmission effects are 2.35 times larger for doctors and 1.32 times larger for the average occupation when a neighbor lives in a different household in the same dwelling than when the neighbor lives at an adjacent address (see Appendix B).

[19]In Appendix Table A9, we explore heterogeneity among other neighborhood characteristics, including the share of non-Whites, whether or not there is an institution of higher education in the same county, and region of the U.S.

5.3 Individual Characteristics

Given the heterogeneity across neighborhoods, it is plausible that there is heterogeneity across characteristics of the neighbors themselves. In particular, we test whether homophily among neighbors increases the transmission rate. From Figure 3, next-door neighbor children are far more likely to marry when their parents share the same birth state, the same birth country, the same race, or have children of the same age.

To test occupation transmission across these different types of pairs, we modify Equation (2) to allow the effects to vary if the focal child and next-door neighbor have the same characteristics:

$$\begin{aligned}\text{Occ1940}_{ij} =& \alpha + \beta_0 \text{OwnOcc1910}_{ij} + \beta_1 \text{NextDoorOcc1910}_{ij} \times \text{SameCharacteristic}_{ij} \\ &+ \beta_2 \text{NextDoorOcc1910}_{ij} \times \text{DifferentCharacteristic}_{ij} + \text{Occ}_j \times \text{Age}_i \\ &+ \text{Occ}_j \times \gamma_{s(i)} + \epsilon_{ij}, \end{aligned} \tag{6}$$

where $\text{SameCharacteristic}_{ij}$ is set equal to one if i has a next-door neighbor in occupation j who is the same as i along various demographic and economic characteristics. Similarly, $\text{DifferentCharacteristic}_{ij}$ equals one if i has a next-door neighbor in occupation j who is different along a characteristic. These estimates are reported in Table 5.

In Column 2, boys are more likely to enter the occupation of their next-door neighbor when the neighbor is born in the same state, 0.49 percentage points more likely for neighbors from the same state versus 0.24 percentage points more likely for neighbors from different states. We present F-test statistics showing that these two coefficients are statistically significantly different from one another. We see similar patterns in Column 3 when we examine differences by birth country (with all native-born grouped together).

In Column 4, we examine whether the race of the next-door neighbor affects the magnitude of the exposure effect. Children with same-race next-door neighbors are 0.38 percentage points more likely to enter into the occupation of that neighbor. Children are 0.83 percentage points *less* likely to go into the occupation of their different-race next-door neighbors. Importantly, this negative effect must be interpreted relative to the appropriate counterfactual group. These coefficients compare the occupational choice of a different-race child relative to the choices of children who live farther away, but still on the same census sheet. Racial segregation in 1910 was high, so these comparison children are more likely to share the same race as the person in the target occupation. For example, a non-White child living next to a White doctor is less likely to become a doctor than other (mostly White) children who grew up on the same street.[20]

In Column 5, we test whether the presence of a child of the same age (±1) in the next-door household affects the magnitude of the transmission effect. If children are more likely to interact with other children of the same age on their streets, then they are also more likely to be exposed to the occupations of the parents of these similar-age children. We do find that next-door neighbors are more predictive of children's future occupations when they have a child of the same age in the household relative to next-door neighbors without a same age child, although the differences in magnitude are modest (0.34 percentage points versus 0.32 percentage points) and not statistically significant.[21]

We explore dimensions of gender homophily in Table 6. For both boys and girls, we find significant transmission effects from both male and female neighbors. Consistent with findings of gender homophily in modern educational settings (Carrell et al., 2010; Dee, 2007), however, the transmission effect for boys is three times larger if the neighbor is male relative to female, while the transmission rate for girls is almost twice as large if the neighbor is female relative to male. Taken together, the heterogeneity by neighborhood and neighbor characteristics are consistent with occupational transmission being larger when social ties

[20]In unreported analysis, we restrict to sub-samples of children of the same race, and find that non-White children have next-door effects statistically indistinguishable from zero when living next to a White adult of the target occupation. Because of strong racial segregation in 1910, there are very few cases of mixed race neighborhoods, making it difficult to estimate cross-race effects with much statistical power.

[21]In Appendix Table A10, we also show differences by household income, household education, and last name homophily. One concern is that extended families co-locate, which would potentially bias our next-door neighbor estimates. However, we find that neighbor transmission effects are nearly identical even when the neighbors do not have the same last name.

are stronger.

6 Exposure Effects on Economic Outcomes

Given that children gravitate toward the occupations of their near-neighbors, we next explore how this impacts long-run economic outcomes. We note that some of the highly transmissible occupations—like doctors and lawyers—have high average income (see Figure 1), but several low-income occupations—such as brickmasons, tailors, and waiters—are also highly transmissible. Exposure to these various occupations could directly affect adult income and educational attainment, or it could simply reallocate children across occupations with similar economic characteristics. For example, shifting a child from law to medicine may have limited implications for their economic outcomes, whereas shifting children from day laborers to electricians could meaningfully improve their life-time earnings. In this section, we examine effects of occupation exposure on long-run income and education for both boys and girls.

6.1 Occupation Transmission by Income and Education Scores

We first show that high-income occupations are more transmissible on average. We run the stacked regression from Equation (2) on subsets of occupations grouped by occupation income scores and education scores, and plot the next-door neighbor effects scaled by the untreated mean in Figure 6.[22] Transmission effects for high-income occupations are slightly larger, at around 12%, while effects for high-education occupations are slightly lower, at 9%. When interacting income and education level, we see that transmission rates are relatively constant except for occupations that have high education and low income—such as teachers

[22]Because of the more limited employment opportunities and lower labor force participation for women in the time period, we focus this section on our sample of boys, but provide analogous estimates for girls in Appendix Table A11. High-income occupations are defined as those in our sample that have an above-median occscore, defined in IPUMS based on the median income for each occupation in the 1950 census. High-education occupations are defined as those in our sample that have an above-median edscore, defined in IPUMS based on the median education attainment for each occupation in the 1950 census.

and clergymen—that have transmission effects that are less than 5%.[23] Although these highly educated neighbors do not have as much influence on direct occupation choice, they do have meaningful impacts on other economic outcomes, as shown below.

6.2 Effects on Adult Income

We now turn our attention to the overall effects of occupation exposure on future income. We follow the same approach as in Equation (1) but for each of the top 50 male occupations, we estimate the effect of living next door to someone of a particular occupation in 1910 on the child's wage income in 1940.[24] As seen in Figure 7, many neighbors' occupations have a significant impact on boys' adult income and there is substantial heterogeneity. Living next to a lawyer or doctor during childhood is associated with a $45-50 (1940$) increase in annual earnings in 1940, relative to other children on the census sheet. Compared to average annual income of $1,071 in 1940 among similarly-aged (30-50) working men, this represents a 4.2-4.7% increase in annual income. As usual, these estimates are relative to other children who were living in the same neighborhood in 1910, thus holding fixed socioeconomic factors that vary across neighborhoods.

There are also occupations that significantly *reduced* next-door children's wage income in 1940 relative to other children on the sheet. Living next to a porter, a truck driver, or a laborer—all of which are low-income occupations—led to annual income reductions at least half as large as the gain from living next to a lawyer or doctor. These significant negative effects may not be driven by direct occupation transmission *per se*, but instead by the lack of exposure to counterfactual occupations on the street. Children who grow up next to

[23] We explore other occupational characteristics, such as heterogeneity by occupational self-employment rates, occupations with apprenticeship structures, or whether occupations are growing or in decline but find few meaningful differences. Our main neighbor spillover finding is likely not driven by direct transmission of businesses, neighbor-provided apprenticeships, or structural labor market changes (see Appendix Table A12.)

[24] The 1940 census only collected information on wage income. Many workers in high-skill occupations such as doctors, lawyers, managers, and real estate agents report no wage income, but report that they had non-wage income (but not the amount). If anything, this under-reporting attenuates the results for high skill occupations.

adults in lower income occupations than the average on their sheet miss out on exposure to neighbors in relatively higher income occupations. Hence, incomes fall not because low-income neighbors actively steer children into low-income work, but because they displace exposure to higher-income occupations.

Girls also have a few professions that substantially increase income. Living next to a milliner, bookkeeper, or stenographer during childhood increases annual earnings by $15-20, a 7-10% increase relative to average annual wage income ($204) and a 2-3% increase relative to average annual wage income among employed women. These results likely differ in magnitude from men because of gender differences in the extensive margin decision to participate in the labor force during this time period, which is influenced by neighbors' occupations for girls, as seen in Figure 1b. As with boys, living next to laborers or private laundresses led to lower annual income relative to other girls on the same sheet.

One puzzle raised by Figure 7 is that several high-education, low-income occupations (e.g., teachers and clergymen) are not strongly transmitted to neighbors (Figure 1) yet are associated with significant income gains. This pattern suggests that neighbors may influence later-life income through multiple channels: (1) a direct channel, in which exposure to neighbors increases entry into higher-paying occupations, and (2) an indirect channel, in which exposure encourages greater educational attainment, thereby expanding access to higher-paying opportunities. We further disentangle these mechanisms by estimating effects on educational attainment and looking at occupation spillovers across occupations in similar educational categories.

6.3 Effects on Educational Attainment

In Figure 8, we find heterogeneous impacts on boys' educational attainment across occupations. Living next to a doctor or lawyer in 1910 increased average years of schooling by nearly 0.4 years relative to other children on the census sheet. However, many of the low-income, high-education occupations also increase educational attainment substantially.

The occupations that have the largest effects on neighbor children's educational attainment are occupations that require formal education, while living next to a neighbor in some trade occupations, such as lumberman, miner, teamster, or laborer, significantly reduced children's educational attainment. Figure 8b shows similarly large effects for many professional occupations with educational requirements for girls. Those living next to nurses, bookkeepers, and teachers attain about 0.2 more years of education than other girls on the street, but many other occupations also lead to educational gains for girls. Living next to some occupations decreases educational attainment, including laundresses, laborers, and kindred workers.

6.4 Cross-Occupational Spillovers

Higher educational attainment likely opens doors to additional occupations besides the specific job held by the next-door neighbor. We therefore estimate an occupation-by-occupation spillover between each of the 25 largest male occupations. To do this, we estimate Equation (1) but iterate through each of the top 25 occupations as both the outcome and the independent variables.[25] The scaled effect of growing up in the same household and next door to a neighbor in any given occupation on the probability of being in each of the top 25 occupations is presented in Figure 9. The occupations are sorted by education score, with the most educated occupations at the left and bottom.

Several clear patterns emerge. First, the largest, positive effects are concentrated along the diagonal, suggesting high levels of occupation-specific transmission. Interestingly, the middle portion of the heat map shows only very specific occupation transmission, with few spillovers to nearby professions. This reflects the specific nature of training necessary for many trade occupations like barbers and blacksmiths. Living next to a plumber does not influence boys to become brickmasons even though they are in the same industry and have similar education requirements.

Second, it is clear that there are often positive spillovers to occupations with similar ed-

[25]We focus on the top 25 occupations for simplicity in presentation. We have explored cross-occupation spillovers for the top 50 occupations and find similar patterns.

ucation requirements and negative spillovers to occupations where the educational requirements differ substantially. The positive, significant spillovers are concentrated in the bottom left (high-education occupations) and top right (low-education occupations) quadrants. For example, living next door to a doctor as a child significantly increases the probability that the child enters a high-paying, high-skilled occupation that requires educational training, such as teacher or manager, but reduces the likelihood that the boy becomes a laborer, kindred worker, miner, or brickmason (low-education occupations). The implications are similar when looking at teachers, a low-income, high-education occupation. Growing up next to a teacher increases the probability of being in a high-paying, high-education occupation such as doctor or manager by almost as much as the effect on becoming a teacher.

In sum, these results suggest that the benefit of growing up next to someone in a high-education occupation is both specific, leading children to follow them into a particular high-paying career, but also general, leading children to attain more education and explore broader career options, many of which lead to higher later-life income.

7 Conclusion

Childhood neighborhoods have important implications for future economic success (Chetty et al., 2026). Using neighborhood microgeography reconstructed from historic, door-to-door census enumeration, we show that part of this phenomenon can be explained by the occupation of the neighbors a child grew up next to. Among boys in 1910, living next door to someone in a particular occupation increased the likelihood that they worked in that occupation in 1940 by 10% relative to other boys who were living on the same street. Girls similarly had positive transmission of occupations, albeit concentrated among different occupations and at lower rates.

This neighbor-to-neighbor transmission of occupation varies across job type, with larger exposure effects from neighbors in high-income occupations. The influence from childhood neighbors' careers also has long-term economic impacts. If a person's childhood neighbors were in particular occupations (such as doctor, lawyer, teacher, or clerical worker), the child experienced higher income and educational attainment as an adult, relative to other children who grew up on the same street. Part of this effect comes through the direct channel of occupation matching, but growing up next door to neighbors in more educated occupations also increases educational attainment and the likelihood of working in a broad set of well-paying jobs. Positive transmission effects on education are striking for girls, who in this time period were beginning to enter professional occupations at higher rates and perhaps benefited uniquely from networking with older professional women in their neighborhood. At the same time, some occupations such as truck driver, laborer, and household worker led to reductions in income and educational attainment for neighboring children. These countervailing impacts on economic outcomes suggest a natural scarcity in positive economic influence within social networks. Even within small groups of similar children living on the street, idiosyncratic placement of neighbors seems to have meaningful impacts on later-life opportunity.

Neighbor transmission effects are larger in neighborhoods that are more socially connected and among neighbors who share common characteristics such as similar-aged children, place of origin, and race. These patterns emphasize the role of social connectedness and are consistent with both information and exposure channels. Interacting with someone in a high-income or high-education occupation can make the returns of that occupation salient or remove information barriers that keep people from being eligible to work in those occupations. Children might not know that a particular occupation exists in their choice set unless they know someone who has pursued that career. Further research could also consider alternative channels, such as the direct transmission of occupations through employment offers, apprenticeships, or business bequests.

Our results suggest a rationale for policies that expose children to social connections that can facilitate upward mobility. Neighborhoods are a natural setting for children to build these relationships, but other settings like after-school programs, camps, teams, churches, and community organizations can likely facilitate those social ties as well. Integrative policies that seek to mix children from different backgrounds and family work experience are promising, but given the role of social homophily in the strength of these spillovers, there may be limits to how much these policies can impact aggregate upward mobility across the economy as a whole.

References

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, Myera Rashid, and Santiago Pérez, "Census Linking Project: 1900-1910 crosswalk," April 2022.

Agresti, Barbara F., "Measuring residential segregation: In nineteenth-century American cities," *Sociological Methods & Research*, May 1980, *8* (4), 389–399.

Aigner, Dennis J., "Regression with a binary independent variable subject to errors of observation," *Journal of Econometrics*, March 1973, *1* (1), 49–59.

Aliprantis, Dionissi, "Human capital in the inner city," *Empirical Economics*, November 2017, *53* (3), 1125–1169.

Andrews, Michael J., "Bar talk: Informal social interactions, alcohol prohibition, and invention," *Working Paper*, 2019.

___ **and Chelsea Lensing**, "Cup of Joe and knowledge flow: Coffee shops and invention," 2024.

___ **and Yiling Zhao**, "Home economics and women's gateway to science," *Working Paper*, 2025.

Andrews, Rodney, Marcus Casey, Bradley L. Hardy, and Trevon D. Logan, "Location matters: Historical racial segregation and intergenerational mobility," *Economics Letters*, September 2017, *158*, 67–72.

Arzaghi, Mohammad and J. Vernon Henderson, "Networking off Madison Avenue," *The Review of Economic Studies*, 2008, *75* (4), 1011–1038.

Atkin, David, M. Keith Chen, and Anton Popov, "The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley," Working Paper 30147, National Bureau of Economic Research June 2022.

Bayer, Patrick, Marcus Casey, W. Ben McCartney, John Orellana-Li, and Calvin Zhang, "Distinguishing causes of neighborhood racial change: A nearest-neighbor design," *American Economic Review*, November 2025, *115* (11), 3999–4039.

___ **, Stephen L. Ross, and Giorgio Topa**, "Place of work and place of residence: Informal hiring networks and labor market outcomes," *Journal of Political Economy*, 2008, *116* (6), 1150–1196.

Bell, Alex and Neviana Petkova, "The long-term impacts of mentors: Evidence from experimental and administrative data," February 2023, (4868302).

___ **, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen**, "Who becomes an inventor in America? The importance of exposure to innovation," *The Quarterly Journal of Economics*, 2019, *134* (2), 647–713.

Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F. Katz, and Christopher Palmer, "Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice," *American Economic Review*, May 2024, *114* (5), 1281–1337.

Billings, Stephen B., David J. Deming, and Jonah Rockoff, "School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg," *The Quarterly Journal of Economics*, February 2014, *129* (1), 435–476.

___ , **Mark Hoekstra, and Gabriel Pons Rotger**, "The scale and nature of neighborhood effects on children," *Journal of Public Economics*, December 2024, *240*, 105260.

Breda, Thomas, Julien Grenet, Marion Monnet, and Clémentine Van Effenterre, "How effective are female role models in steering girls towards STEM? Evidence from French high schools," *The Economic Journal*, 2023, *133* (653), 1773–1809.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley E.B. Wilbert, "Breakthroughs in historical record linking using genealogy data: The Census Tree Project," *Explorations in Economic History*, September 2025, *98*, 101717.

___ , **Joseph Price, Zachary Ward, and Haley E.B. Wilbert**, "Family trees and falling apples: Historical intergenerational mobility estimates for women and men," Working Paper 31918, National Bureau of Economic Research November 2023.

Carrell, Scott E., Marianne E. Page, and James E. West, "Sex and science: How professor gender perpetuates the gender gap," *The Quarterly Journal of Economics*, 2010, *125* (3), 1101–1144.

Catalini, Christian, "Microgeography and the direction of inventive activity," *Management Science*, September 2018, *64* (9), 4348–4364.

Chetty, Raj and Nathaniel Hendren, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *The Quarterly Journal of Economics*, 2018, *133* (3), 1107–1162.

___ **and** ___ , "The impacts of neighborhoods on intergenerational mobility II: County-level estimates," *The Quarterly Journal of Economics*, 2018, *133* (3), 1163–1228.

___ , **John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter**, "The opportunity atlas: Mapping the childhood roots of social mobility," *American Economic Review*, 2026, *116* (1), 1–51.

___ , **Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barbera, Monica Bhole, and Nils Wernerfelt**, "Social capital I: Measurement and associations with economic mobility," *Nature*, 2022, *608*, 108–121.

___ , **Nathaniel Hendren, and Lawrence F. Katz**, "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment," *American Economic Review*, 2016, *106* (4), 855–902.

___ , ___ , **Patrick Kline, and Emmanuel Saez**, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, November 2014, *129* (4), 1553–1623.

Chung, Bobby W., "Peers' parents and educational attainment: The exposure effect," *Labour Economics*, June 2020, *64*, 101812.

Chyn, Eric, "Moved to opportunity: The long-run effects of public housing demolition on children," *American Economic Review*, October 2018, *108* (10), 3028–3056.

___ , **Robert Collinson, and Danielle H. Sandler**, "The long-run effects of America's largest residential racial desegregation program: Gautreaux," *The Quarterly Journal of Economics*, August 2025, *140* (3), 2213–2267.

Cook, Lisa D., Trevon D. Logan, and John M. Parman, "Racial segregation and Southern lynching," *Social Science History*, 2018, *42* (4), 635–675.

Corak, Miles and Patrizio Piraino, "The intergenerational transmission of employers," *Journal of Labor Economics*, 2011, *29* (1), 37–68.

Dal Bó, Ernesto, Pedro Dal Bó, and Jason Snyder, "Political dynasties," *The Review of Economic Studies*, January 2009, *76* (1), 115–142.

Dee, Thomas S., "Teachers and the gender gaps in student achievement," *The Journal of Human Resources*, 2007, *42* (3), 528–554.

Duncan, Dustin T., Ichiro Kawachi, and Foreword by Ana V. Diez Roux, eds, *Neighborhoods and health*, second edition ed., Oxford, New York: Oxford University Press, April 2018.

Eriksson, Katherine and Zachary Ward, "The residential segregation of immigrants in the United States from 1850 to 1940," *The Journal of Economic History*, December 2019, *79* (4), 989–1026.

Fairlie, Robert W. and Alicia Robb, "Families, human capital, and small business: Evidence from the characteristics of business owners survey," *ILR Review*, January 2007, *60* (2), 225–245.

Fruehwirth, Jane Cooley and Jessica Gagete-Miranda, "Your peers' parents: Spillovers from parental education," *Economics of Education Review*, December 2019, *73*, 101910.

Greenberg, Kyle, Matthew Gudgeon, Adam Isen, Corbin L. Miller, and Richard W. Patterson, "Intergenerational transmission of occupation: Lessons from the United States Army," NBER Working Paper 33009, National Bureau of Economic Research September, revised November 2024.

Grinblatt, Mark, Matti Keloharju, and Seppo Ikheimo, "Social influence and consumption: Evidence from the automobile purchases of neighbors," *The Review of Economics and Statistics*, 2008, *90* (4), 735–753.

Haltiwanger, John C., Mark J. Kutzbach, Giordano E. Palloni, Henry Pollakowski, Matthew Staiger, and Daniel Weinberg, "The children of HOPE VI demolitions: National evidence on labor market outcomes," *Journal of Public Economics*, November 2024, *239*, 105188.

Hellerstein, Judith K., Melissa McInerney, and David Neumark, "Neighbors and coworkers: The importance of residential labor market networks," *Journal of Labor Economics*, 2011, *29* (4), 659–695.

Hvide, Hans K. and Paul Oyer, "Dinner table human capital and entrepreneurship," Working Paper 24198, National Bureau of Economic Research January 2018.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman, "Moving to opportunity in Boston: Early results of a randomized mobility experiment," *The Quarterly Journal of Economics*, 2001, *116* (2), 607–654.

Kawano, Laura, Bruce Sacerdote, William L. Skimmyhorn, and Michael Stevens, "On the determinants of young adult outcomes: Impacts of randomly assigned neighborhoods for children in military families," Working Paper 32674, National Bureau of Economic Research July 2024.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz, "Experimental analysis of neighborhood effects," *Econometrica*, January 2007, *75* (1), 83–119.

Laitila, Janne, Jerry Montonen, and Samuel Solomon, "Childhood peers, occupational choice, and the intergenerational persistence of the elite," SSRN Scholarly Paper 4981877, Social Science Research Network, Rochester, NY October 2024.

Laliberté, Jean-William, "Long-term contextual effects in education: Schools and neighborhoods," *American Economic Journal: Economic Policy*, May 2021, *13* (2), 336–377.

Lentz, Bernard F. and David N. Laband, "Why so many children of doctors become doctors: Nepotism vs. human capital transfers," *The Journal of Human Resources*, 1989, *24* (3), 396–413.

Library of Congress, U.S., "Fire insurance maps at the Library of Congress: A resource guide," n.d. accessed July 12, 2025.

Logan, Trevon D. and John M. Parman, "The national rise in residential segregation," *The Journal of Economic History*, 2017, *77* (1), 127–170.

___ **and** ___ , "Segregation and homeownership in the early twentieth century," *American Economic Review*, May 2017, *107* (5), 410–414.

___ **and** ___ , "Segregation and mortality over time and space," *Social Science & Medicine*, February 2018, *199*, 77–86.

Manski, Charles F., "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542.

McCartney, W. Ben and Avni M. Shah, "Household mortgage refinancing decisions are neighbor influenced, especially along racial lines," *Journal of Urban Economics*, March 2022, *128*, 103409.

Mertz, Mikkel, Maddalena Ronchi, and Viola Salvestrini, "Female representation and talent allocation in entrepreneurship: The role of early exposure to entrepreneurs," IZA Discussion Paper 17801, Institute of Labor Economics (IZA) 2025.

Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou, "Mothers, peers, and gender-role identity," *Journal of the European Economic Association*, 2020, *18* (1), 266–301.

Pischke, Steve, "Lecture notes on measurement error," 2007.

Polyakova, Maria, Petra Persson, Katja Hofmann, and Anupam B. Jena, "Does medicine run in the family—evidence from three generations of physicians in Sweden: Retrospective observational study," *BMJ*, December 2020, *371*, m4453.

Porter, Catherine and Danila Serra, "Gender differences in the choice of major: The importance of female role models," *American Economic Journal: Applied Economics*, 2020, *12* (3), 226–254.

Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley, "Combining family history and machine learning to link historical records: The Census Tree data set," *Explorations in Economic History*, 2021, *80*, 101391.

Quincy, Sarah, "Income shocks and housing spillovers: Evidence from the World War I veterans' bonus," *Journal of Urban Economics*, November 2022, *132*, 103494.

Redding, Stephen J. and Daniel M. Sturm, "Neighborhood effects: Evidence from wartime destruction in London," Working Paper 32333, National Bureau of Economic Research April 2024.

Rosenbaum, James E., "Changing the geography of opportunity by expanding residential choice: Lessons from the Gautreaux program," *Housing Policy Debate*, January 1995, *6* (1), 231–269.

Sanborn Map Company, "Collection: Sanborn maps," Various Years. Library of Congress Digital Collections.

Shah, Avni and W. Ben McCartney, ""I'll have what she's having": Neighborhood social interactions lead to policy spillovers," *Journal of the Association for Consumer Research*, October 2023, *8* (4), 403–415.

Sharkey, Patrick and Gerard Torrats-Espinosa, "The effect of violent crime on economic mobility," *Journal of Urban Economics*, November 2017, *102*, 22–33.

Staiger, Matthew, "The intergenerational transmission of employers and the earnings of young workers," *Working Paper*, October 2023.

Tan, Hui Ren, "How widespread are social network effects? Evidence from the early twentieth-century United States," *Journal of Labor Economics*, January 2022, *40* (1), 187–237.

U.S. Bureau of the Census, Department of Commerce and Labor, "Thirteenth Census of the United States, April 15, 1910, instructions to enumerators," 1910.

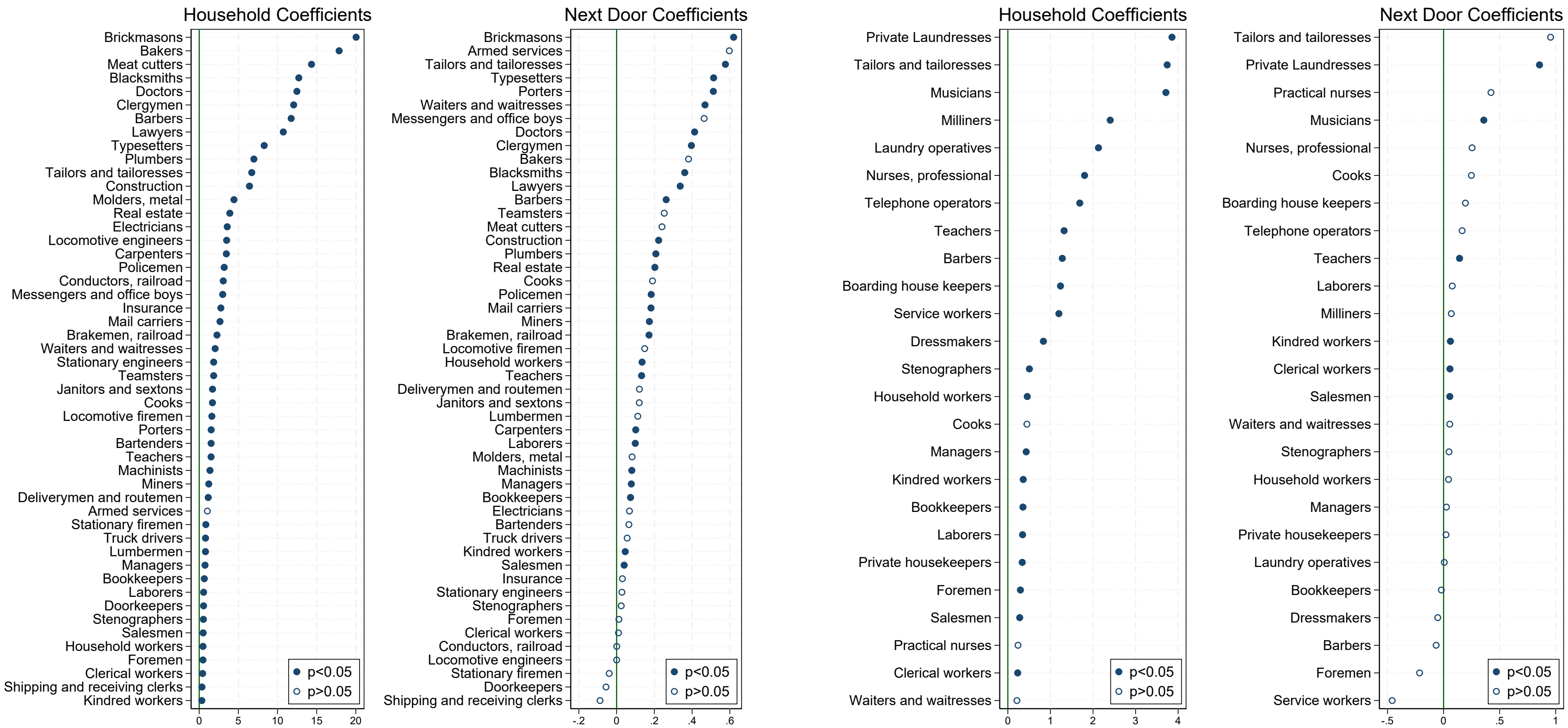
Ventura, Maria, "Following in the family footsteps: Incidence and returns of occupational persistence," CEP Discussion Paper 2121, Centre for Economic Performance, LSE 2025.

Tables and Figures

Figure 1: Household and Next Door Coefficients, Individual Occupations

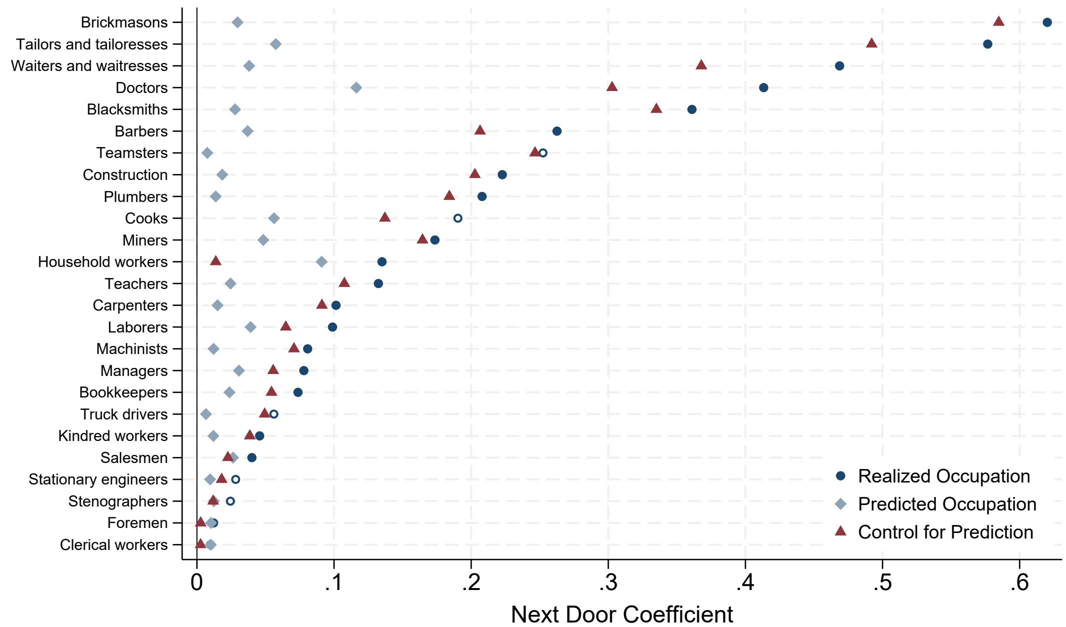
(a) 50 Largest Occupations for Men

(b) 25 Largest Occupations for Women



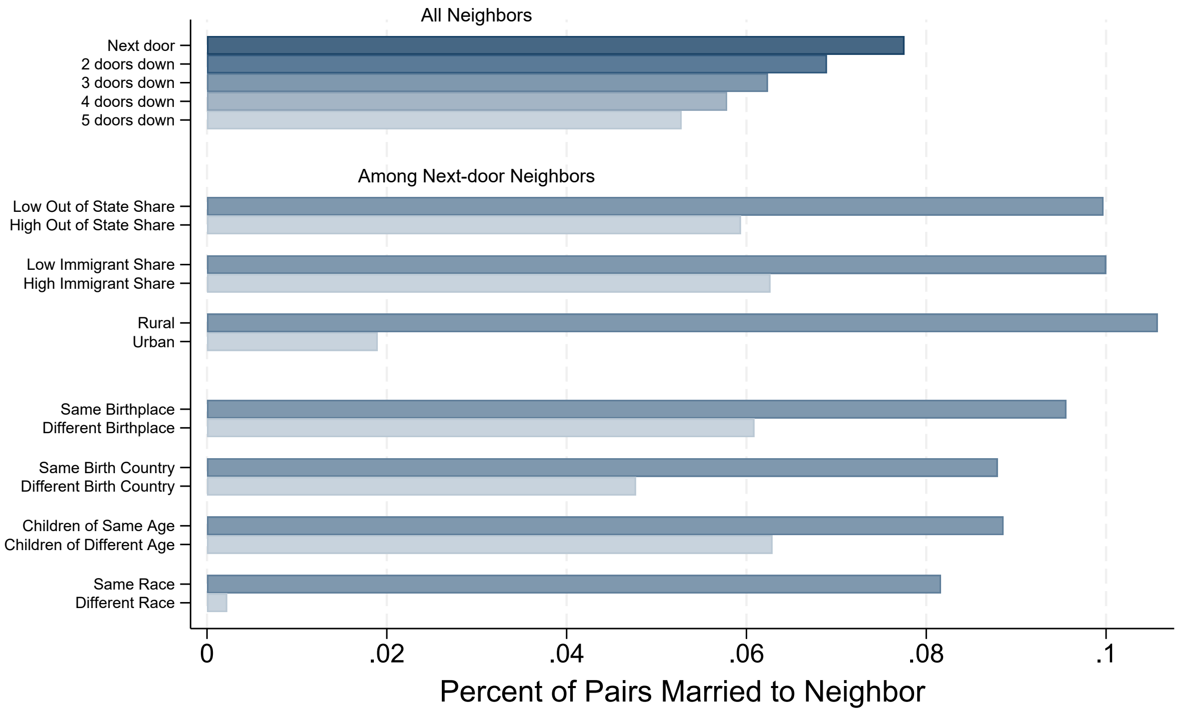
Notes: Estimates from regression Equation (1) for each occupation run separately are reported. Sample restricted to boys between the ages of 5 and 18 in 1910 that can be linked to the 1940 census. The outcome is a binary measure that equals one if the child reports working in that occupation in 1940. The "Household Coefficients" panel reports the coefficient on the binary indicator for having someone in your household in that occupation in 1910. The "Next Door Coefficients" panel reports the coefficient on the binary indicator for having someone next door in that occupation in 1910. Both coefficients are scaled by the untreated mean of the outcome. All regressions also include sheet and age fixed effects. Markers are filled if the coefficient is statistically significant at the 5% level. Standard errors clustered at the household level.

Figure 2: Predicted vs. Realized Occupation Regressions



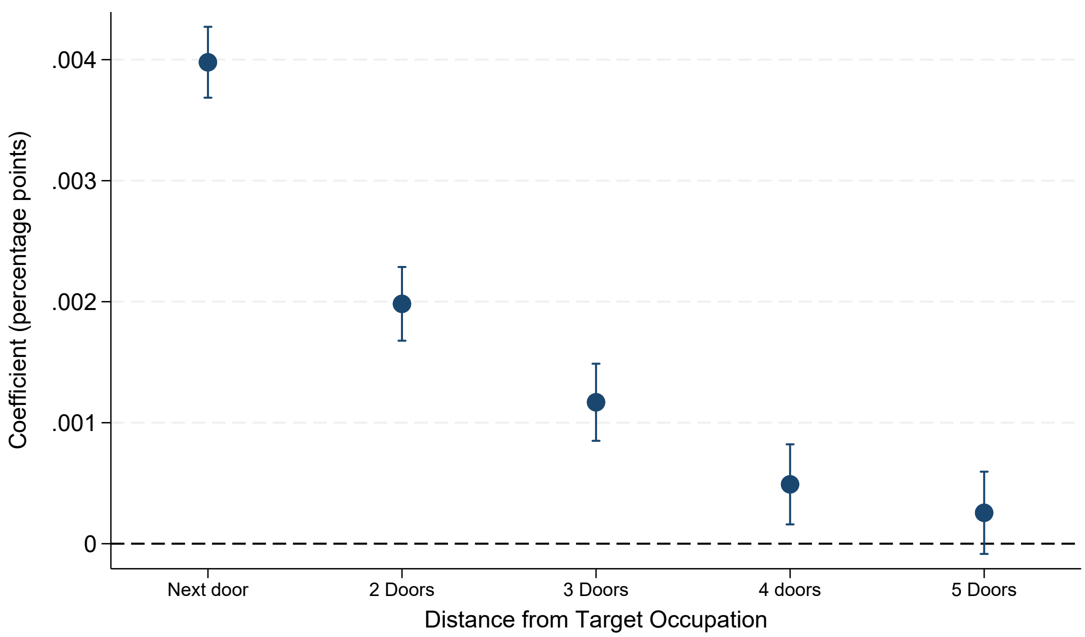
Notes: The coefficient on next-door neighbor, scaled by the untreated mean, are reported for three separate regression specifications for the 25 largest occupations for men. Sample restricted to boys between the ages of 5 and 18 in 1910 that can be linked to the 1940 census. The light blue diamonds represent the coefficient when the outcome is the predicted occupation probability from the lasso regression of household characteristics regressed on next-door neighbor occupation. The dark blue dots represent the coefficient when the outcome is the realized occupation (as in Figure 1a. The red triangle represent the coefficients when the outcome is the realized occupation when we control directly for the predicted probability.

Figure 3: Heterogeneity in Social Connectedness — Marriage Rates of Childhood Neighbors



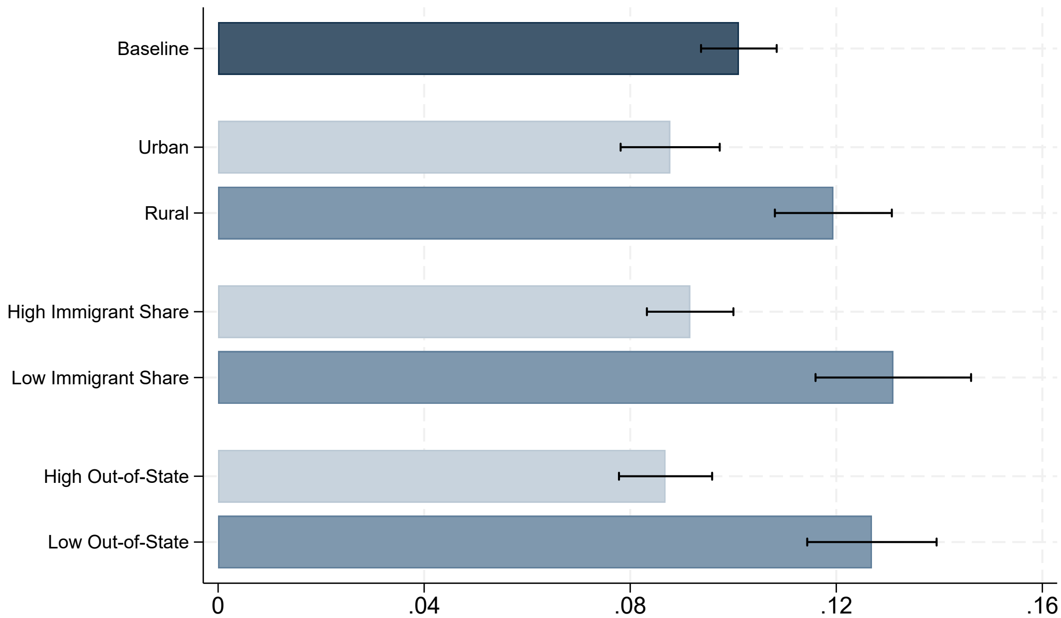
Notes: Marriage rates constructed at the number of doors away level in the top panel, and for next-door neighbors in the bottom panel. For every potential boy-girl pair in the given group on a 1910 sheet of the given geographic proximity (e.g., one door away, two doors away), we calculate the share that are married in 1940. We then compare the share of next-door pairs that are married differentially by neighborhood characteristics and by homophily between the households. To be counted as married both the husband and the wife must be linked to the 1940 Census, meaning these estimates may be a lower bound of the true effect.

Figure 4: Door Distance Regressions (Average Across 50 Largest Male Occupations)



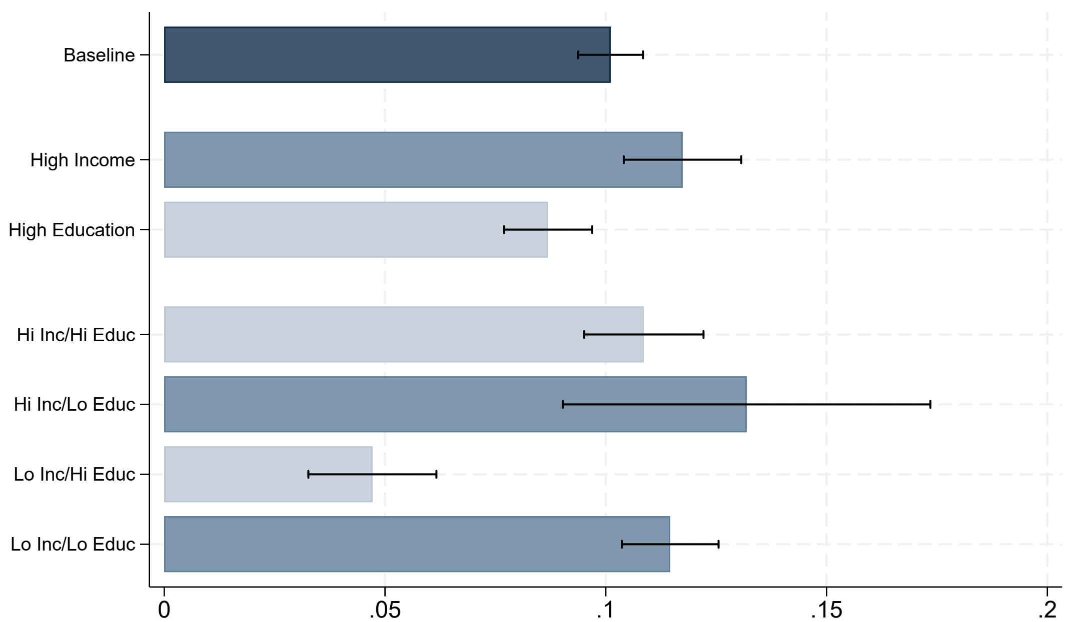
Notes: Coefficients from the stacked regression specification in Equation (5) for the 50 largest occupations for boys are plotted. Sample restricted to boys between the ages of 5 and 18 in 1910 that can be linked to the 1940 census. The outcome is a binary measure that equals one if the boy chooses the exposed occupation in 1940. This is regressed on binary measures that equal one if the boy's family lived next door, two doors away, etc. from a particular occupation in 1910, as well as a binary measure that equals one if the child had that occupation in their own household (not plotted). Regression includes census sheet fixed effects. 95% confidence intervals included with standard errors clustered at the household level.

Figure 5: Heterogeneity in Occupation Transmission by Neighborhood



Notes: Each bar reports the next door coefficient scaled by the untreated mean from the stacked regression specification described in Equation (2). Each regression sample is restricted to groups of sheets that fit the description on the y-axis. Sample of interest is boys 5-18 years old in 1910 that are linked to the 1940 census. Error bars represent the 95% confidence interval (scaled using the the delta method), where standard errors are clustered at the household level.

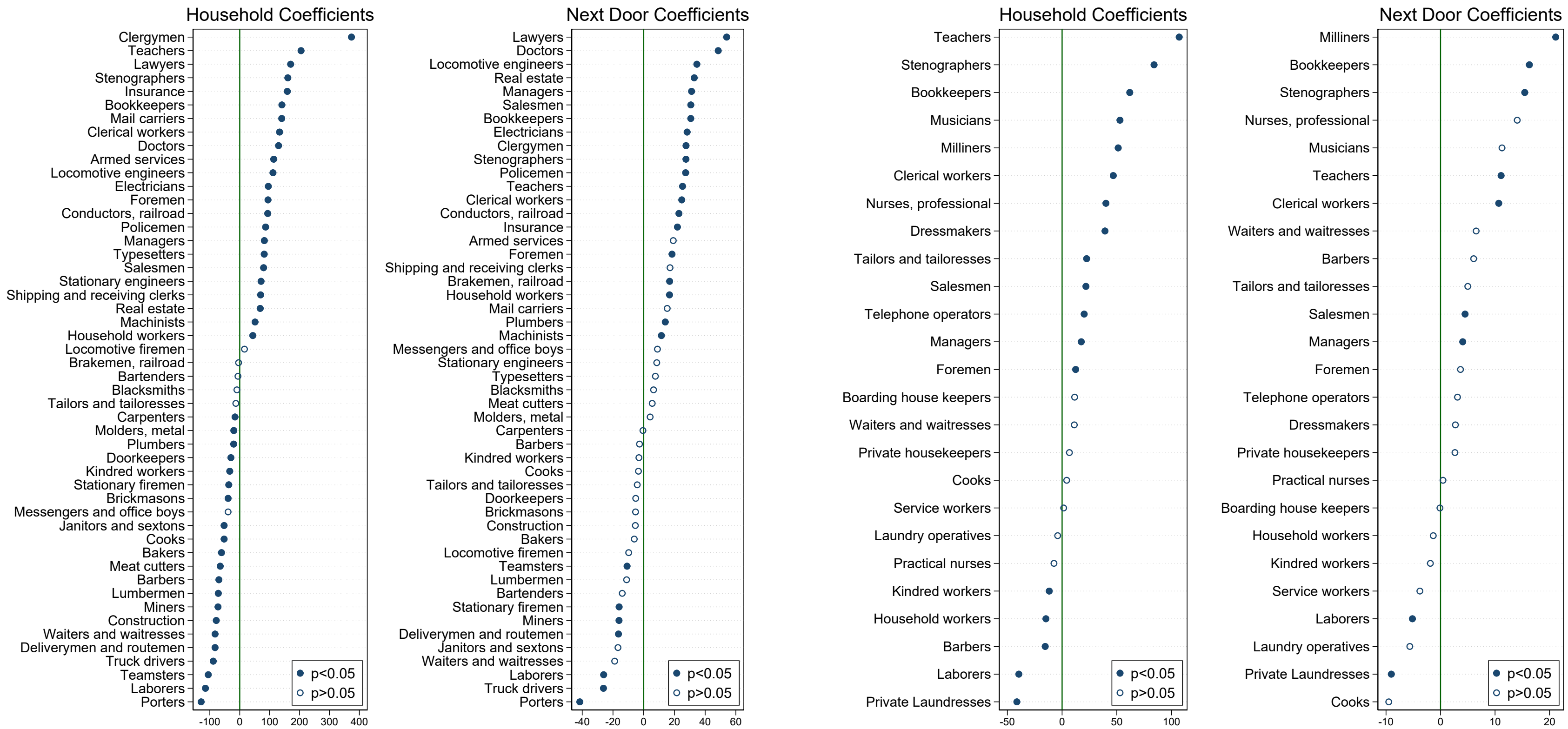
Figure 6: Heterogeneity by Occupation Type



Notes: Each bar reports the next door coefficient scaled by the untreated mean from the stacked regression specification described in Equation (2). Each regression sample is restricted to the subset of occupations in the stack denoted on the y-axis. Sample of interest is boys 5-18 years old in 1910 that are linked to the 1940 census. Error bars represent the 95% confidence interval (scaled using the the delta method), where standard errors are clustered at the household level. The occupation groups are divided using the median income and education occupation scores for each occupation from IPUMS.

Figure 7: Household and Next Door Coefficients on Income

(a) 50 Largest Occupations for Men (b) 25 Largest Occupations for Women

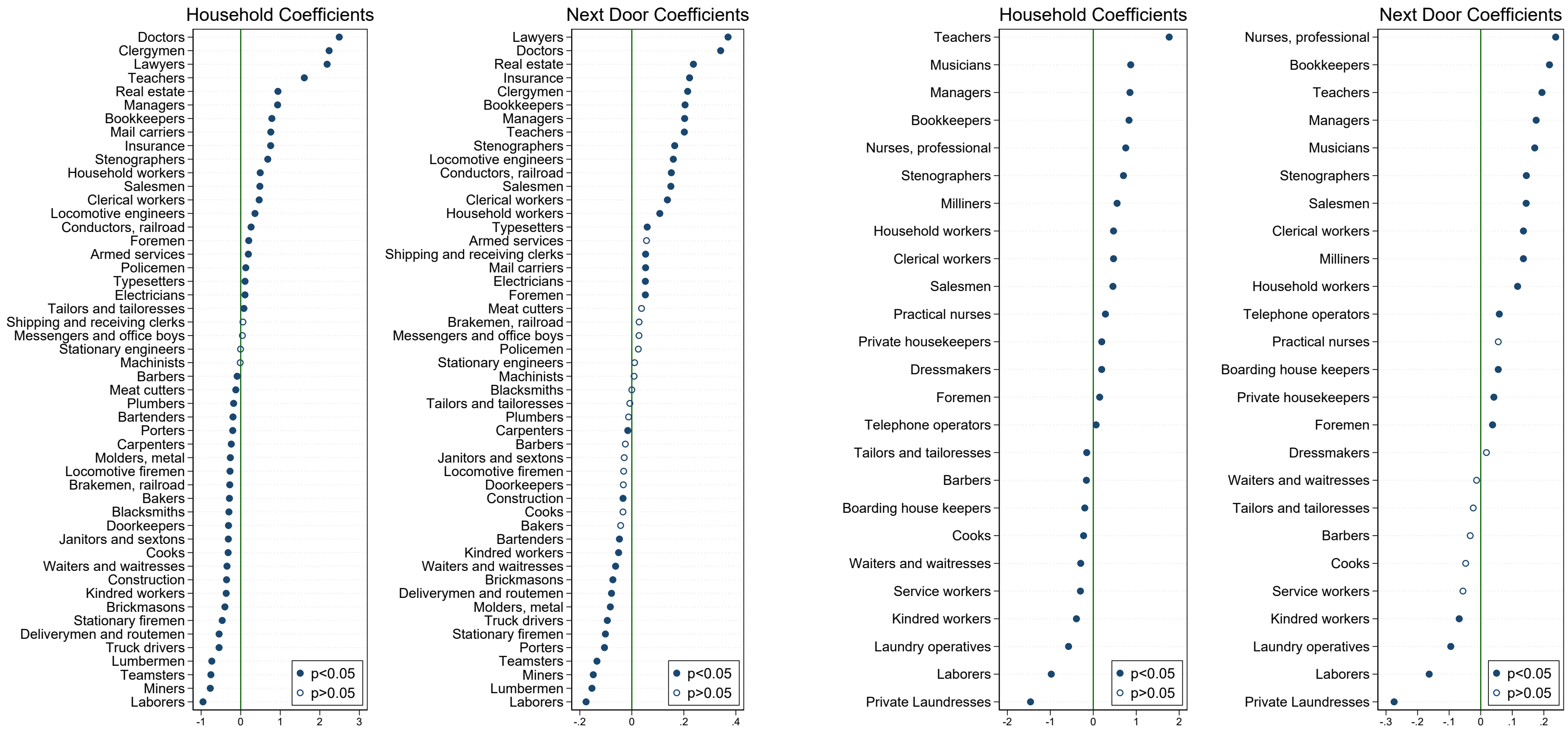


Notes: Estimates from regression Equation (1) for each occupation run separately are reported, where the outcome is the individual's wage income in 1940. Sample restricted to boys between the ages of 5 and 18 in 1910 that can be linked to the 1940 census. The "Household Coefficients" panel reports the coefficient on the binary indicator for having someone in your household in that occupation in 1910. The "Next Door Coefficients" panel reports the coefficient on the binary indicator for having someone next door in that occupation in 1910. All regressions include sheet and age fixed effects. Markers are filled if the coefficient is statistically significant at the 5% level. Standard errors clustered at the household level.

Figure 8: Household and Next Door Coefficients on Education

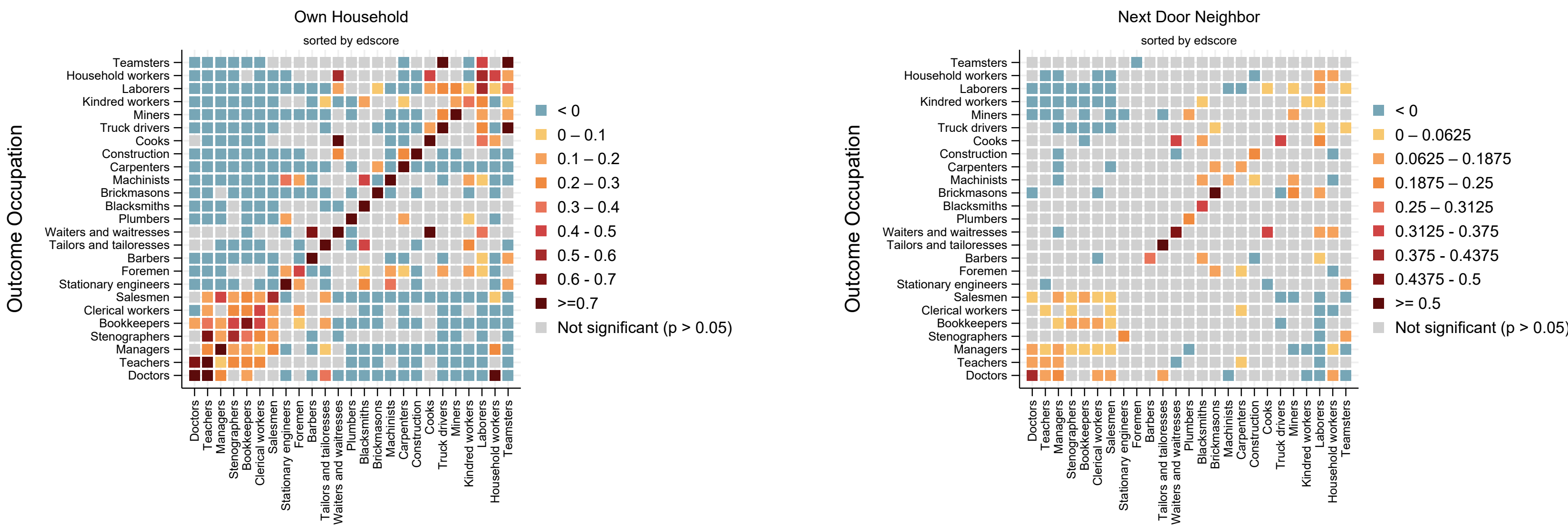
(a) 50 Largest Occupations for Men

(b) 25 Largest Occupations for Women



Notes: Estimates from regression Equation (1) for each occupation run separately are reported, where the outcome is the individual's educational attainment recorded in 1940. Sample restricted to boys between the ages of 5 and 18 in 1910 that can be linked to the 1940 census. The "Household Coefficients" panel reports the coefficient on the binary indicator for having someone in your household in that occupation in 1910. The "Next Door Coefficients" panel reports the coefficient on the binary indicator for having someone next door in that occupation in 1910. All regressions include sheet and age fixed effects. Markers are filled if the coefficient is statistically significant at the 5% level. Standard errors clustered at the household level.

Figure 9: Occupation Spillover Matrices



Notes: These figures report the coefficients for occupation spillovers for own household (left panel) and next-door neighbor (right panel) transmission. Coefficients are scaled by the untreated mean of the outcome. All regressions also include sheet and age fixed effects. The x-axis corresponds to the occupation that the child is exposed to in 1910 and the y-axis corresponds to the potential occupation that a child could report in 1940. Each square is obtained from a separate regression. Each square in the matrix is colored to represent the sign, magnitude, and significance level of the coefficient from a regression of the y-axis outcome on the x-axis exposure as described in the text. For example, living next door to a doctor has a large positive effect on becoming a doctor, an insignificant effect on becoming a plumber, and a significant negative effect on becoming a laborer.

Table 1: Summary Statistics

	All 1910 children (1)	1910 - 1940 Linked children (2)	1910 Boys (3)	1910 - 1940 Linked boys (4)	1910 Girls (5)	1910 - 1940 Linked girls (6)
Age in 1910	11.380	11.244	11.366	11.228	11.395	11.268
Male	0.503	0.603	1.000	1.000	0.000	0.000
Nonwhite	0.130	0.047	0.128	0.060	0.132	0.027
Rural	0.603	0.646	0.611	0.634	0.595	0.666
Northeast	0.252	0.226	0.250	0.240	0.253	0.204
Midwest	0.315	0.378	0.316	0.368	0.314	0.395
South	0.371	0.324	0.371	0.323	0.371	0.325
West	0.063	0.072	0.063	0.070	0.062	0.076
Head of household						
Married	0.884	0.914	0.883	0.907	0.885	0.923
Foreign born	0.311	0.277	0.311	0.292	0.310	0.255
Income score	21.975	22.107	21.899	22.170	22.051	22.011
Education score	12.342	12.867	12.124	12.712	12.563	13.103
Total	26,161,014	10,529,180	13,146,449	6,346,719	13,014,565	4,182,461

Notes: This table shows summary statistics for the sample data. Column 1 describes all children age 5-18 in the 1910 census. Column 2 describes the children linked from the 1910 census to the 1940 census. Columns 3-4 further restrict the sample to boys only, while columns 5-6 restrict to girls only.

Table 2: Balance Test by Household Characteristics

Household characteristics:	Own household (1)	Next door neighbors (2)	Further neighbors (3)	Difference: (next door - further) (4)	Std. error of difference (5)
Panel A: Doctors					
Head is a home owner	0.717	0.566	0.547	0.019	0.003
Head occupation score (100s of 1950 dollars)	69.857	25.357	24.217	1.140	0.088
Head education score (share in occ with some college in 1950)	82.244	18.120	15.593	2.527	0.131
Head is white	0.987	0.977	0.973	0.004	0.001
Head is a US native	0.886	0.828	0.826	0.002	0.002
Lasso prediction: Child is a doctor in 1940	0.034	0.008	0.007	0.001	0.000
Observations	29,529	52,109	223,511		
Panel B: Teachers					
Head is a home owner	0.732	0.608	0.592	0.015	0.002
Head occupation score (100s of 1950 dollars)	20.604	21.372	21.197	0.175	0.041
Head education score (share in occ with some college in 1950)	23.216	12.037	11.219	0.818	0.054
Head is white	0.967	0.967	0.966	0.001	0.001
Head is a US native	0.836	0.808	0.805	0.003	0.001
Lasso prediction: Child is a teacher in 1940	0.013	0.010	0.010	0.000	0.000
Observations	132,367	219,050	804,187		
Panel C: Brickmasons					
Head is a home owner	0.532	0.496	0.503	-0.006	0.003
Head occupation score (100s of 1950 dollars)	27.086	23.017	23.041	-0.024	0.068
Head education score (share in occ with some college in 1950)	6.332	10.242	10.742	-0.500	0.082
Head is white	0.968	0.969	0.968	0.001	0.001
Head is a US native	0.602	0.660	0.684	-0.024	0.003
Lasso prediction: Child is a brickmason in 1940	0.011	0.004	0.004	0.000	0.000
Observations	45,325	65,801	266,940		

Notes: This table tests treatment balance by comparing household characteristics of children living in the same house, next door, or further away on the same census sheet to doctors, teachers, and brickmasons. Columns 1-3 present means for children in mutually exclusive categories based on the closest adult with that occupation. Column 4 shows the difference between the next door and further neighbors, and Column 5 shows the heteroskedasticity-robust standard error of the difference. The 'Lasso Prediction' is estimated from separate regressions of a 1940 occupation choice indicator regressed on a battery of household and individual characteristics. Observation counts report the number of boys in each group with a non-missing lasso prediction.

Table 3: Baseline Doctor Results

Dependent variable: Doctor occupation in 1940	All sheets	At least one doctor occupation per sheet: No geographic FE	City - enumeration district FE	Sheet FE
	(1)	(2)	(3)	(4)
Own household	0.0983***	0.0954***	0.0963***	0.0970***
	(0.0018)	(0.0019)	(0.0018)	(0.0018)
Next door neighbor	0.0072***	0.0036***	0.0033***	0.0032***
	(0.0005)	(0.0006)	(0.0006)	(0.0007)
R-squared	0.011	0.046	0.138	0.327
Untreated mean	0.0038	0.0078	0.0078	0.0078
Observations	6,335,662	305,063	305,063	305,063

Notes: Sample includes all male children who were age 5-18 in 1910 full-count census and can be linked to the 1940 census. Column 1 includes all sheets in 1910 census, while columns 2-4 include only sheets with at least one adult in the target occupation. All columns include age fixed effects and Column 4 adds census sheet fixed effects.

*p<0.1, **p<0.05, ***p<0.01.

Table 4: Baseline Stacked Regression Results

Dependent variable: Target occupation in 1940	All sheets	At least one person in target occupation per sheet: No geographic FE	City - enumeration district FE	Sheet FE
	(1)	(2)	(3)	(4)
Panel A: Boys				
Own household	0.0548***	0.0467***	0.0424***	0.0381***
	(0.0001)	(0.0002)	(0.0002)	(0.0002)
Next door neighbor	0.0198***	0.0112***	0.0070***	0.0034***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
R-squared	0.038	0.048	0.087	0.323
Untreated mean	0.0100	0.0331	0.0331	0.0331
Observations	316,783,104	26,649,896	26,649,896	26,649,896
Panel B: Girls				
Own household	0.0158***	0.0116***	0.0095***	0.0085***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Next door neighbor	0.0087***	0.0041***	0.0020***	0.0010***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
R-squared	0.011	0.018	0.083	0.356
Untreated mean	0.0072	0.0156	0.0156	0.0156
Observations	104,389,296	11,135,892	11,135,892	11,135,892

Notes: This table reports estimates from the stacked regression specification, with Panel A showing results for the 50 largest male occupations, and Panel B for the 25 largest female occupations. Sample includes all children who were age 5-18 in 1910 full-count census and can be linked to the 1940 census. Column 1 includes all sheets in 1910 census, while columns 2-4 include only sheets with at least one adult in the target occupation. All columns include age by occupation fixed effects and Column 4 adds sheet by occupation fixed effects.

*p<0.1, **p<0.05, ***p<0.01.

Table 5: Heterogeneity by Neighbor Similarity

Dependent variable: Target occupation in 1940	Baseline	Birthplace	Birth country	Race	Same age
	(1)	(2)	(3)	(4)	(5)
Own household	0.0381***	0.0381***	0.0381***	0.0381***	0.0381***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Next door neighbor	0.0034***				
	(0.0001)				
Treatment 1: Same characteristic		0.0049***	0.0040***	0.0038***	0.0034***
		(0.0002)	(0.0002)	(0.0001)	(0.0003)
Treatment 1 average		*0.0966*	*0.1581*	*0.2286*	*0.0336*
Treatment 2: Different characteristic		0.0024***	0.0021***	-0.0083***	0.0032***
		(0.0002)	(0.0002)	(0.0006)	(0.0001)
Treatment 2 average		*0.1439*	*0.0818*	*0.0085*	*0.1941*
F-test		104.68	62.52	403.92	0.47
Two-tailed p-value		0.0000	0.0000	0.0000	0.4947
R-squared	0.323	0.323	0.323	0.323	0.323
Untreated mean	0.0331	0.0332	0.0332	0.0332	0.0333
Observations	26,649,896	26,599,180	26,599,180	26,599,180	26,649,896

Notes: The sample is all boys in the 50 occupations of the baseline stacked regressions. In Columns 2-5, treatments compare head of household of the focal child to the target occupation holder next door. In Column 5, same age is defined as any child in the target occupation house being within one year of the focal child. All columns include age by occupation and sheet by occupation fixed effects.

*p<0.1, **p<0.05, ***p<0.01.

Table 6: Heterogeneity by Gender Homophily

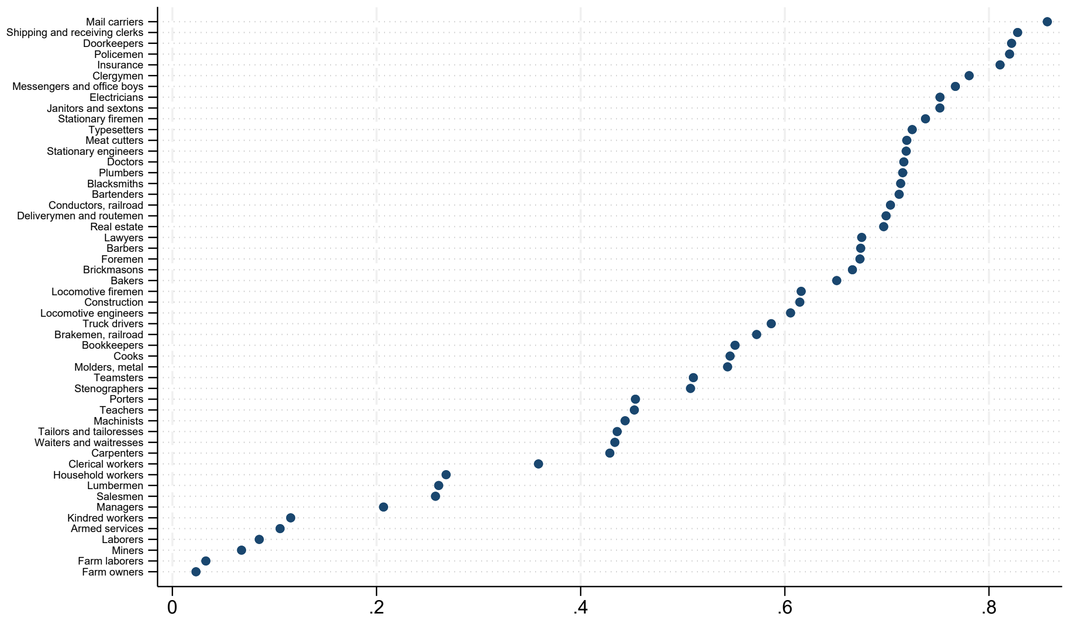
Dependent variable: Target occupation in 1940	Boys baseline	Boys homophily	Girls baseline	Girls homophily
	(1)	(2)	(3)	(4)
Own household	0.0381***		0.0085***	
	(0.0002)		(0.0002)	
Next door neighbor	0.0034***		0.0010***	
	(0.0001)		(0.0001)	
Male own household		0.0431***		0.0051***
		(0.0002)		(0.0002)
Male next door neighbor		0.0037***		0.0008***
		(0.0001)		(0.0002)
Female own household		0.0095***		0.0149***
		(0.0003)		(0.0003)
Female next door neighbor		0.0012***		0.0014***
		(0.0003)		(0.0002)
R-squared	0.323	0.323	0.356	0.356
Untreated mean	0.0331	0.0331	0.0156	0.0156
Observations	26,649,938	26,649,938	11,135,904	11,135,904

Notes : The sample in Columns 1 and 2 consist of the boys in the stacked regression of 50 largest male occupations. The sample in Columns 3 and 4 consists of the girls in the stacked regression of 25 largest female occupations. Columns 2 and 4 separate the treatment variables by the gender of the adult in the target occupation. All columns include age by occupation and sheet by occupation fixed effects.

*p<0.1, **p<0.05, ***p<0.01.

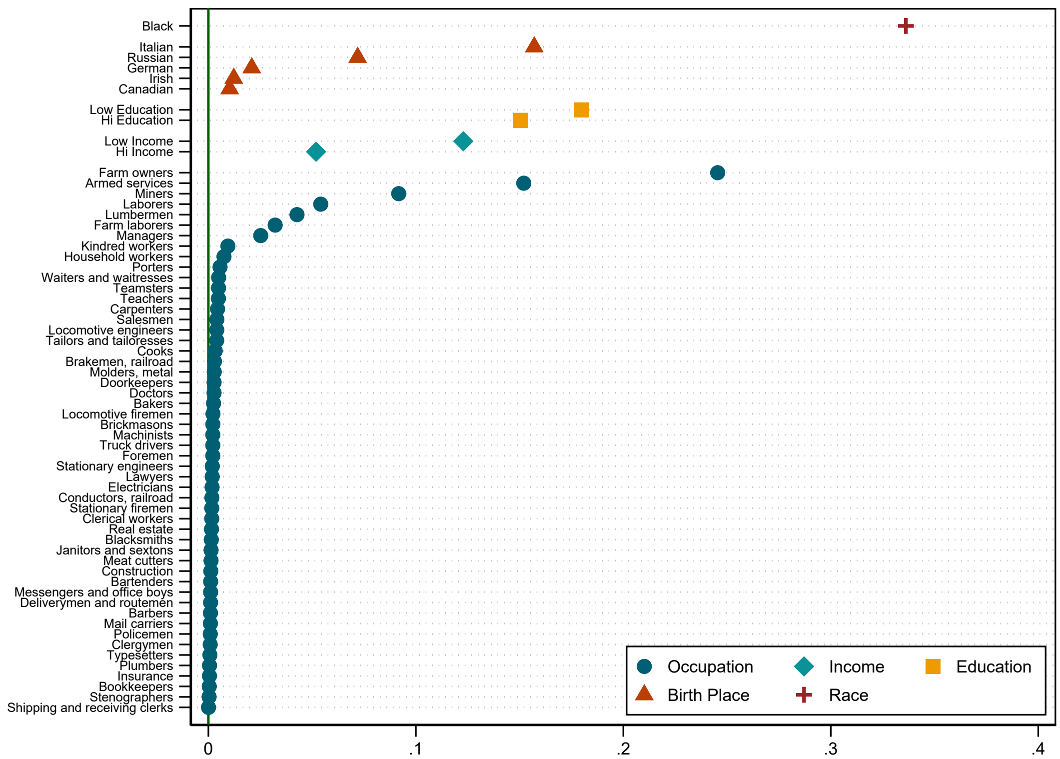
A Additional Tables and Figures

Figure A1: Share of Sheets with Only a Single Occurrence of Each Occupation



Notes: Among the census manuscript pages with at least one occurrence of each occupation, the share of pages in which the occupation appears exactly once for each of the 50 largest occupations for boys.

Figure A2: Logan-Parman (2017) Measure of Segregation

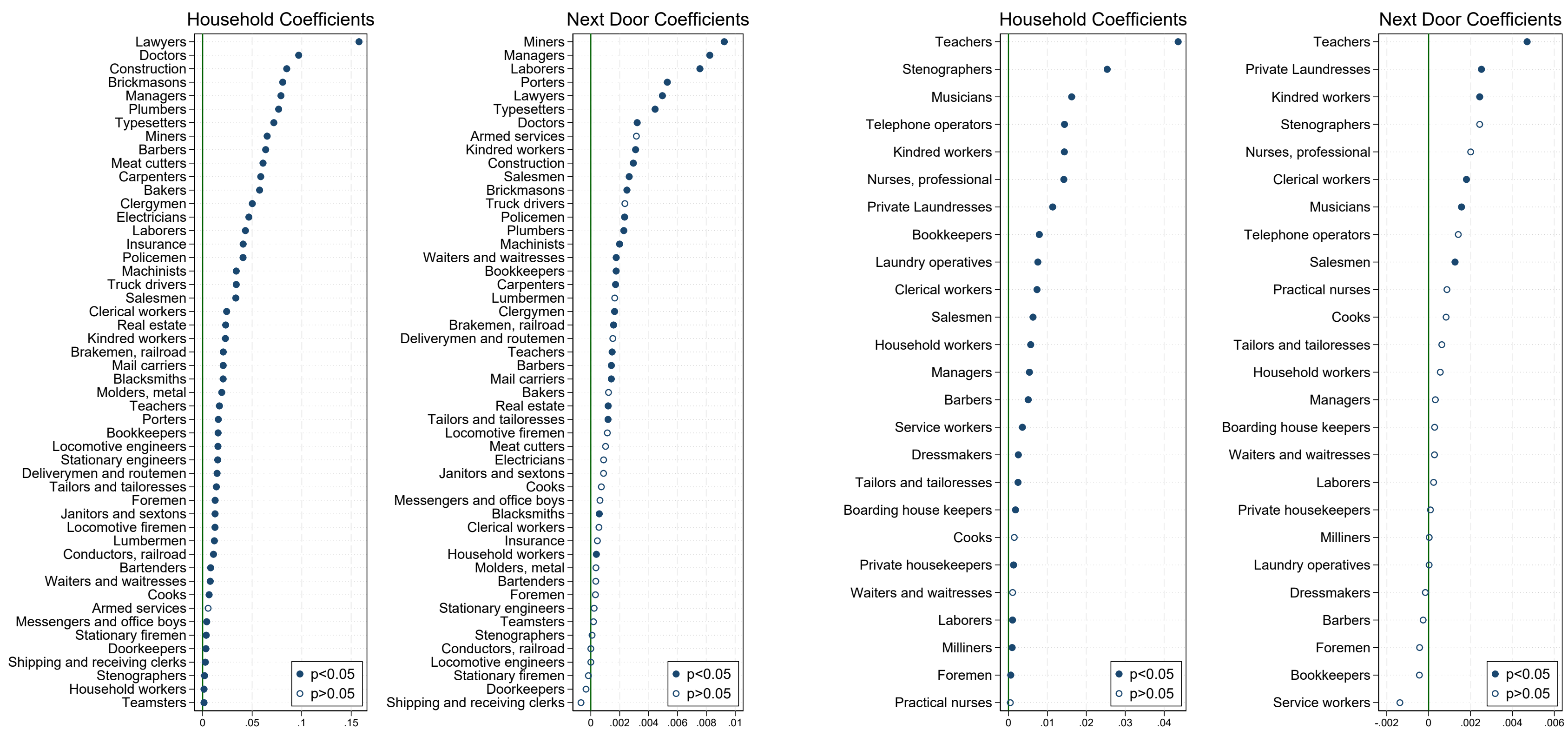


Notes: Residential segregation of occupations and other groups, using the Logan and Parman (2017a) measure of next-door neighbor-based segregation. The measure calculates how likely an adult in one occupation is to live next to an adult in a different occupation relative to what would be expected when randomly allocating occupations across households. For a given occupation j, this is given by $\eta_j = \frac{E(\bar{x}_j) - x_j}{E(\bar{x}_j) - E(\underline{x_j})}$, where x_j is the observed number of pairs of adjacent households in which one household has occupation j and the other does not, $E(\bar{x}_j)$ is the expected x_j if households sorted randomly given the total number of j in the population, and $E(\underline{x_j})$ is the expected x_j if households were perfectly segregated by occupation. $\eta_j = 0$ therefore corresponds to no residential segregation for occupation j, while $\eta_j = 1$ corresponds to perfect segregation. We plot the Logan and Parman (2017a) measure for each of the 50 largest occupations for boys in solid blue circles. To put occupational segregation into perspective, we also the Logan and Parman (2017a) segregation measure for race (red crosses), birthplace (red triangles), educational attainment (yellow squares), and income (blue diamonds) of household heads.

Figure A3: Household and Next Door Coefficients, Individual Occupations, Without Scaling

(a) 50 Largest Occupations for Men

(b) 25 Largest Occupations for Women

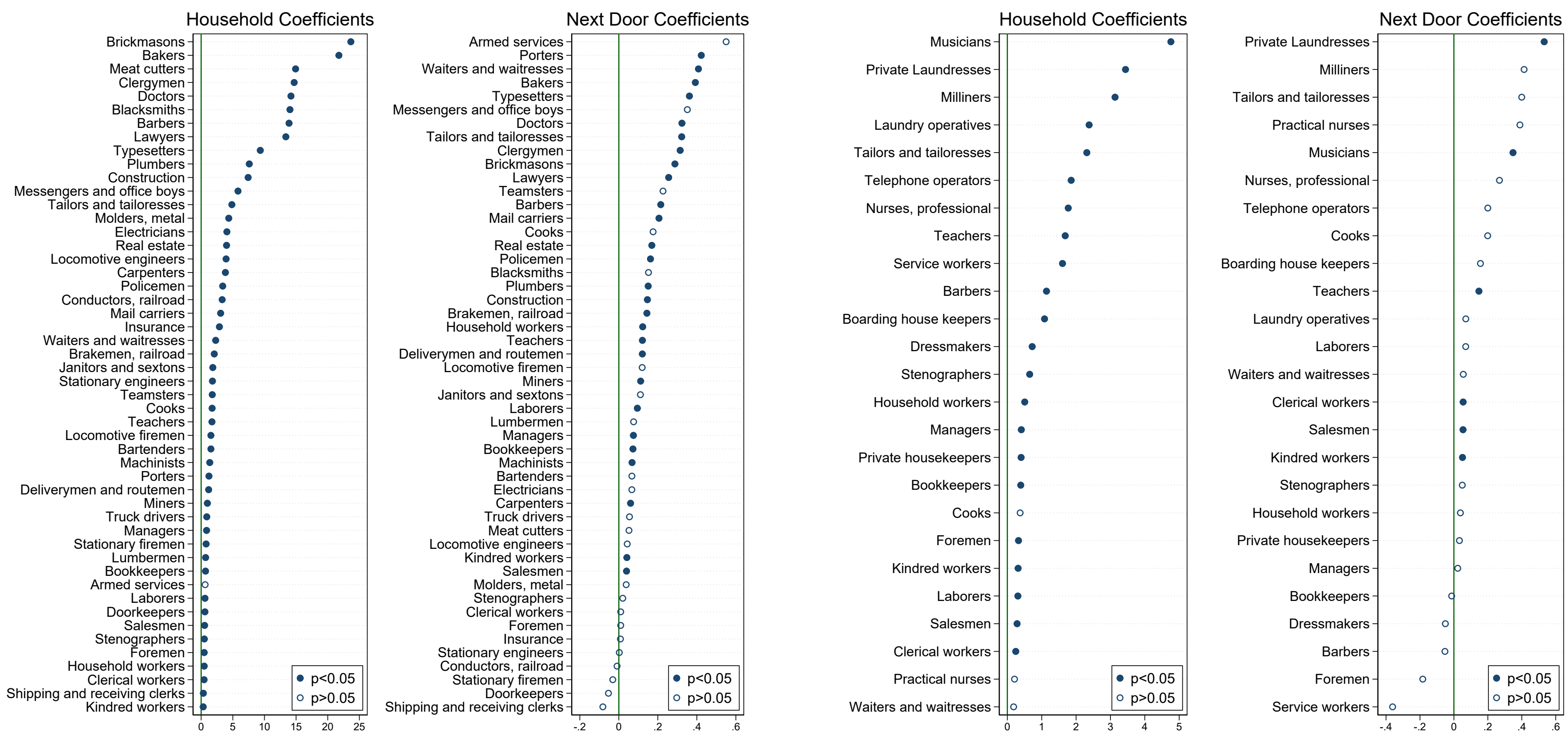


Notes: Estimates from regression Equation (1) for each occupation run separately are reported. Sample restricted to boys between the ages of 5 and 18 in 1910 that can be linked to the 1940 census. The outcome is a binary measure that equals one if the child reports working in that occupation in 1940. The "Household Coefficients" panel reports the coefficient on the binary indicator for having someone in your household in that occupation in 1910. The "Next Door Coefficients" panel reports the coefficient on the binary indicator for having someone next door in that occupation in 1910. Unlike the main table, coefficients are not scaled by the untreated mean of the outcome. All regressions also include sheet and age fixed effects. Markers are filled if the coefficient is statistically significant at the 5% level. Standard errors clustered at the household level.

Figure A4: Household and Next Door Coefficients, Logit Specification

(a) 50 Largest Occupations for Men

(b) 25 Largest Occupations for Women



Notes: Estimates from the logit estimation corresponding to regression Equation (1) for each occupation run separately are reported. Sample restricted to boys between the ages of 5 and 18 in 1910 that can be linked to the 1940 census. The outcome is a binary measure that equals one if the child reports working in that occupation in 1940. The "Household Coefficients" panel reports the coefficient on the binary indicator for having someone in your household in that occupation in 1910. The "Next Door Coefficients" panel reports the coefficient on the binary indicator for having someone next door in that occupation in 1910. Unlike the main table, coefficients are not scaled by the untreated mean of the outcome. All regressions also include sheet and age fixed effects. Markers are filled if the coefficient is statistically significant at the 5% level. Standard errors clustered at the household level.

Table A1: Stacked Regression Results while Handling Farmers in Different Ways

Dependent variable: Target occupation in 1940	Baseline occupations (no farm owners or farm laborers)	Baseline occupations plus farm owners & laborers	No farm owners	No farm laborers	Only farm owners	Only farm laborers	Only farm owners & farm laborers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Own household	0.0381***	0.0408***	0.0341***	0.0454***	0.1028***	0.0105***	0.0498***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0007)	(0.0004)	(0.0004)
Next door neighbor	0.0034***	0.0041***	0.0031***	0.0045***	0.0112***	0.0013***	0.0073***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0008)	(0.0003)	(0.0003)
R-squared	0.323	0.368	0.315	0.376	0.335	0.250	0.376
Untreated mean	0.0331	0.0346	0.0336	0.0342	0.0690	0.0402	0.0484
Observations	26,649,896	33,117,292	29,644,912	30,122,276	3,472,379	2,995,015	6,467,394

Notes: Regression specification and boys sample as described in main stacked sample table. Column 1 excludes farm owners and laborers as in the baseline analysis, Column 2 adds farm owners and laborers, Column 3 excludes farm owners, Column 4 excludes laborers, Columns 5-6 includes only farm owner and/or farm laborers. All regressions include age by occupation and sheet by occupation fixed effects.

*p<0.1, **p<0.05, ***p<0.01.

Table A2: Balance Test by Household Characteristics — Not Mutually Exclusive Groups

Household characteristics:	Own household (1)	Next door neighbors (2)	Further neighbors (3)	Difference: (next door - further) (4)	Std. error of difference (5)
Panel A: Doctors					
Head is a doctor	0.817	0.001	0.000	0.001	0.000
Head is a home owner	0.674	0.541	0.540	0.001	0.003
Head occupation score (100s of 1950 dollars)	68.199	24.705	24.273	0.432	0.086
Head education score (share in occ with some college in 1950)	81.042	17.828	15.762	2.066	0.125
Head is white	0.930	0.946	0.971	-0.025	0.001
Head is a US native	0.836	0.801	0.824	-0.023	0.002
Lasso prediction: Child is a doctor in 1940	0.033	0.008	0.007	0.001	0.000
Observations	29,529	52,109	223,511		
Panel B: Teachers					
Head is a teacher	0.127	0.002	0.000	0.002	0.000
Head is a home owner	0.682	0.576	0.587	-0.011	0.001
Head occupation score (100s of 1950 dollars)	19.053	20.653	21.320	-0.667	0.040
Head education score (share in occ with some college in 1950)	22.513	11.987	11.394	0.593	0.054
Head is white	0.891	0.924	0.963	-0.040	0.001
Head is a US native	0.774	0.774	0.805	-0.030	0.001
Lasso prediction: Child is a teacher in 1940	0.012	0.010	0.010	0.000	0.000
Observations	132,367	219,050	804,187		
Panel C: Brickmasons					
Head is a brickmason	0.759	0.002	0.000	0.002	0.000
Head is a home owner	0.497	0.473	0.495	-0.023	0.003
Head occupation score (100s of 1950 dollars)	25.737	22.200	23.059	-0.859	0.067
Head education score (share in occ with some college in 1950)	5.810	9.875	10.793	-0.917	0.078
Head is white	0.910	0.931	0.964	-0.033	0.001
Head is a US native	0.563	0.631	0.682	-0.051	0.003
Lasso prediction: Child is a brickmason in 1940	0.010	0.004	0.004	0.000	0.000
Observations	45,325	65,801	266,940		

Notes: This table tests treatment balance by comparing household characteristics of children living in the same house, next door, or further away on the same census sheet to doctors, teachers, and brickmasons. Columns 1-3 present means for children in each category, allowing for cases where a doctor may be in the same house and next door. Column 4 shows the difference between the next door and further neighbors, and Column 5 shows the heteroskedasticity-robust standard error of the difference. The 'Lasso Prediction' is estimated from separate regressions of a 1940 occupation choice indicator regressed on a battery of household and individual characteristics. Observation counts report the number of boys in each group with a non-missing lasso prediction.

Table A3: Balance Test by Household Characteristics — One Occupation per Sheet

Household characteristics:	Own household (1)	Next door neighbors (2)	Further neighbors (3)	Difference: (next door - further) (4)	Std. error of difference (5)
Panel A: Doctors					
Head is a doctor	0.810	0.000	0.000	0.000	0.000
Head is a home owner	0.711	0.560	0.542	0.018	0.003
Head occupation score (100s of 1950 dollars)	69.203	24.971	24.070	0.901	0.091
Head education score (share in occ with some college in 1950)	81.412	17.385	15.362	2.023	0.135
Head is white	0.985	0.974	0.970	0.004	0.001
Head is a US native	0.884	0.826	0.824	0.003	0.002
Lasso prediction: Child is a doctor in 1940	0.033	0.008	0.007	0.001	0.000
Observations	23,288	43,179	202,096		
Panel B: Teachers					
Head is a teacher	0.121	0.000	0.000	0.000	0.000
Head is a home owner	0.730	0.600	0.586	0.014	0.002
Head occupation score (100s of 1950 dollars)	20.160	20.770	20.947	-0.177	0.047
Head education score (share in occ with some college in 1950)	21.718	11.007	10.770	0.238	0.059
Head is white	0.962	0.962	0.962	0.000	0.001
Head is a US native	0.828	0.802	0.803	-0.001	0.002
Lasso prediction: Child is a teacher in 1940	0.012	0.010	0.010	0.000	0.000
Observations	82,291	142,653	612,230		
Panel C: Brickmasons					
Head is a brickmason	0.753	0.000	0.000	0.000	0.000
Head is a home owner	0.517	0.492	0.497	-0.005	0.003
Head occupation score (100s of 1950 dollars)	26.937	22.960	23.015	-0.055	0.073
Head education score (share in occ with some college in 1950)	6.389	10.306	10.800	-0.493	0.089
Head is white	0.966	0.966	0.964	0.002	0.001
Head is a US native	0.615	0.666	0.685	-0.020	0.003
Lasso prediction: Child is a brickmason in 1940	0.010	0.004	0.004	0.000	0.000
Observations	34,235	52,444	231,546		

Notes: This table tests treatment balance by comparing household characteristics of children living in the same house, next door, or further away on the same census sheet to doctors, teachers, and brickmasons. The sample is restricted to sheets with a single appearance of the occupation, so Columns 1-3 present means for children in mutually exclusive categories. Column 4 shows the difference between the next door and further neighbors, and Column 5 shows the heteroskedasticity-robust standard error of the difference. The 'Lasso Prediction' is estimated from separate regressions of a 1940 occupation choice indicator regressed on a battery of household and individual characteristics. Observation counts report the number of boys in each group with a non-missing lasso prediction.

Table A4: Doctor Results: Robustness

Dependent variable: Doctor in 1940	Include sheet position fixed effects	Within 3 doors of household head	Drop group quarters	One doctor per sheet	Interaction	Head of household controls	Head occupation fixed effects
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Own household	0.0969***	0.0984***	0.0979***	0.0912***	0.0953***	0.0709***	0.0284***
	(0.0018)	(0.0022)	(0.0018)	(0.0020)	(0.0018)	(0.0021)	(0.0026)
Next door neighbor	0.0032***	0.0040***	0.0031***	0.0024***	0.0358***	0.0025***	0.0027***
	(0.0007)	(0.0011)	(0.0007)	(0.0007)	(0.0093)	(0.0007)	(0.0006)
Own household × Next door					-0.0335***		
					(0.0093)		
Head occscore						0.0001**	
						(0.0000)	
Head edscore						0.0003***	
						(0.0000)	
Head age						-0.0002***	-0.0002***
						(0.0000)	(0.0000)
White						-0.0039*	-0.0001
						(0.0021)	(0.0011)
US native						-0.0028***	-0.0047***
						(0.0010)	(0.0007)
Education flag						0.0100***	
						(0.0013)	
R-squared	0.327	0.406	0.327	0.319	0.328	0.329	0.055
Untreated mean	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Observations	304,552	143,143	300,420	268,504	305,063	305,050	305,040

Notes: Column 1 includes sheet position in the 1910 census fixed effects to the column 4 regression of the baseline doctor table. Column 2 restricts our sample to only include households within 3 houses of a doctor in 1910. Column 3 restricts our sample to not include group quarters in 1910. Column 4 restricts the sample to census sheets with only one doctor in 1910. Column 5 includes an interaction between own household and next door neighbor. Column 6 adds controls for head of household characteristics in 1910, and column 7 adds head of household occupation fixed effects to column 6. All columns include age and sheet fixed effects and the sample is restricted to sheets with at least one doctor.

**p*<0.1, ***p*<0.05, ****p*<0.01.

Table A5: Stacked Regression Results: Robustness

Dependent variable: Target occupation in 1940	Include sheet position fixed effects	Within 3 doors of household head	Drop group quarters	One target occupation per sheet	Interaction	Head of household controls	Head occupation fixed effects
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Own household	0.0381***	0.0385***	0.0385***	0.0325***	0.0361***	0.0564***	0.0213***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Next door neighbor	0.0034***	0.0033***	0.0033***	0.0018***	0.0138***	0.0196***	0.0086***
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0005)	(0.0001)	(0.0001)
Own household × Next door					-0.0116***		
					(0.0005)		
Head occscore						0.0001***	
						(0.0000)	
Head edscore						0.0000	
						(0.0000)	
Head age						-0.0001***	0.0000***
						(0.0000)	(0.0000)
White						-0.0171***	-0.0219***
						(0.0003)	(0.0003)
US native						-0.0017***	-0.0001
						(0.0001)	(0.0001)
Education flag						0.0034***	
						(0.0002)	
R-squared	0.323	0.365	0.323	0.301	0.323	0.064	0.058
Untreated mean	0.033	0.039	0.033	0.026	0.033	0.039	0.033
Observations	26,612,050	15,304,988	26,328,024	17,838,936	26,649,896	26,648,648	26,648,172

Notes: Column 1 includes sheet position in the 1910 census fixed effects to the column 4 regression of the baseline stacked table. Column 2 restricts our sample to only include households within 3 houses of a target occupation in 1910. Column 3 restricts our sample to not include group quarters in 1910. Column 4 restricts the sample to census sheets with only one adult in the target occupation in 1910. Column 5 includes an interaction between own household and next door neighbor. Column 6 adds controls for head of household characteristics in 1910, and column 7 adds head of household occupation fixed effects to column 6. All columns include age by occupation and sheet by occupation fixed effects and the sample is restricted to sheets with at least one person in the target occupation.

**p*<0.1, ***p*<0.05, ****p*<0.01.

Table A6: Control for Lasso Predicted Occupation

Dependent variable: Target occupation in 1940	Doctors only:		50 largest male occupations (stack):	
	(1)	(2)	(3)	(4)
Own household	0.0970***	0.0721***	0.0381***	0.0209***
	(0.0018)	(0.0020)	(0.0002)	(0.0002)
Next door neighbor	0.0032***	0.0024***	0.0034***	0.0023***
	(0.0007)	(0.0007)	(0.0001)	(0.0001)
Predicted occupation probability		0.9204***		0.7058***
		(0.0441)		(0.0038)
R-squared	0.327	0.329	0.323	0.326
Untreated mean	0.008	0.008	0.033	0.033
Observations	305,063	305,063	26,649,896	26,649,896

Notes: Columns 1 and 3 reproduce the main next door neighbor results in the preferred specification as presented in the doctor and stacked tables. Columns 2 and 4 add a control variable for the predicted occupation probability from a lasso regression of occupation choice on household characteristics.

**p* <0.1, ***p* <0.05, ****p* <0.01.

Table A7: Doctor Results in 1920, 1930, and 1940 Censuses

	1910 children linked to at least one census			1910 children linked to all three censuses		
Dependent variable: Doctor in…	1920	1930	1940	1920	1930	1940
	(1)	(2)	(3)	(4)	(5)	(6)
Own household	0.0164***	0.0852***	0.0970***	0.0199***	0.0944***	0.1092***
	(0.0007)	(0.0016)	(0.0018)	(0.0011)	(0.0024)	(0.0026)
Next door neighbor	0.0008***	0.0025***	0.0032***	0.0008*	0.0031***	0.0035***
	(0.0003)	(0.0006)	(0.0007)	(0.0005)	(0.0010)	(0.0010)
R-squared	0.266	0.321	0.327	0.359	0.383	0.387
Untreated mean	0.0014	0.0065	0.0078	0.0014	0.0065	0.0079
Observations	410,387	323,153	305,063	177,548	177,548	177,548

Notes: Regression specification and sample as described in main doctor table, except that the sample in Columns 1-3 consists of boys that are linked to the corresponding outcome census, and Columns 4-6 consists of boys that are linked to all censuses from 1910 to 1920, 1930, and 1940. The outcome in each column is an indicator for the individual listing doctor as their occupation in the 1920, 1930, or 1940 census as indicated in the column header.

**p* <0.1, ***p* <0.05, ****p* <0.01.

Table A8: Stacked Regression Results in 1920, 1930, and 1940 Censuses

Dependent variable: Target occupation in…	1910 children linked to at least one census			1910 children linked to all three censuses		
	1920	1930	1940	1920	1930	1940
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Boys						
Own household	0.0365***	0.0450***	0.0381***	0.0375***	0.0471***	0.0410***
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Next door neighbor	0.0032***	0.0037***	0.0034***	0.0032***	0.0036***	0.0034***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)
R-squared	0.318	0.324	0.323	0.405	0.385	0.377
Untreated mean	0.0262	0.0329	0.0331	0.0257	0.0327	0.0330
Observations	37,437,816	28,531,024	26,649,896	15,473,423	15,473,423	15,473,423
Panel B: Girls						
Own household	0.0169***	0.0107***	0.0085***	0.0140***	0.0090***	0.0080***
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Next door neighbor	0.0016***	0.0012***	0.0010***	0.0011***	0.0011***	0.0009***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)
R-squared	0.338	0.351	0.356	0.426	0.395	0.393
Untreated mean	0.0255	0.0180	0.0156	0.0200	0.0146	0.0149
Observations	23,472,128	13,424,049	11,135,892	6,885,592	6,885,592	6,885,592

Notes: Table reports coefficients from the stacked regressions where outcomes are indicators of the child reporting the target occupation in the census year listed in column header. The sample in Columns 1-3 consists of children that are linked to the corresponding outcome census, and Columns 4-6 consists of children that are linked to all censuses from 1910 to 1920, 1930, and 1940. The Panel A sample consists of boys exposed to the 50 largest occupations for men in 1910, and Panel B consists of girls exposed to the 25 largest occupations for women in 1910. All regressions include age by occupation and sheet by occupation fixed effects.

**p*<0.1, ***p*<0.05, ****p*<0.01.

Table A9: Heterogeneity by Additional Neighborhood Characteristics

Dependent variable: Target occupation in 1940	High non-white share	Low non-white share	University in county	No university in county	Northeast	Midwest	South	West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Own household	0.0396***	0.0370***	0.0360***	0.0402***	0.0361***	0.0385***	0.0425***	0.0338***
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0004)	(0.0006)
Next door neighbor	0.0038***	0.0030***	0.0029***	0.0038***	0.0027***	0.0032***	0.0049***	0.0025***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0004)
R-squared	0.336	0.312	0.328	0.319	0.330	0.308	0.337	0.315
Untreated mean	0.0338	0.0327	0.0324	0.0340	0.0335	0.0318	0.0365	0.0296
Observations	11,676,929	14,972,664	13,604,760	13,044,685	9,198,233	9,743,871	5,464,553	2,243,240

Notes: Regression specification and boys sample as described in main stacked sample table. Each column reports results within different subsets of the census sample based on neighborhood characteristics or region.

**p*<0.1, ***p*<0.05, ****p*<0.01.

Table A10: Heterogeneity by Additional Characteristics — Unique Last Names and SES

Dependent variable: Target occupation in 1940	Baseline	Last name	Income score	Education score
	(1)	(2)	(3)	(4)
Own household	0.0381***	0.0381***	0.0378***	0.0376***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Next door neighbor	0.0034***			
	(0.0001)			
Treatment 1: Same characteristic		0.0216***	0.0060***	0.0077***
		(0.0010)	(0.0002)	(0.0002)
Treatment 1 average		*0.0041*	*0.1447*	*0.1453*
Treatment 2: Different characteristic		0.0030***	-0.0006***	-0.0030***
		(0.0001)	(0.0002)	(0.0002)
Treatment 2 average		*0.2330*	*0.0918*	*0.0913*
F-test		308.56	799.33	2142.69
Two-tailed p-value		0.0000	0.0000	0.0000
R-squared	0.323	0.323	0.323	0.323
Untreated mean	0.0331	0.0332	0.0332	0.0332
Observations	26,649,896	26,598,318	26,599,180	26,599,180

Notes: Regression specification and boys sample as described in main stacked sample table. In columns 2-4, treatments compare head of household of the focal child to the target occupation holder next door. In columns 3-4 the treatments evaluate if both head of household of focal child and target occupation next door are above/below median in income score and education score. The number of observations in column 2 is slightly less due to missing last names in digitized census records.

*p<0.1, **p<0.05, ***p<0.01.

Table A11: Heterogeneity by Occupation Income and Education

Dependent variable: Target occupation in 1940	Baseline	High income	High education	High income: High education	High income: Low education	Low income: High education	Low income: Low education
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Boys							
Own household	0.0381***	0.0537***	0.0439***	0.0609***	0.0396***	0.0242***	0.0334***
	(0.0002)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0004)	(0.0002)
Next door neighbor	0.0034***	0.0040***	0.0037***	0.0053***	0.0014***	0.0017***	0.0036***
	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0002)
R-squared	0.323	0.337	0.323	0.334	0.290	0.301	0.322
Untreated mean	0.0331	0.0339	0.0421	0.0485	0.0103	0.0358	0.0314
Observations	26,649,896	8,481,845	10,662,698	5,356,208	3,125,637	5,306,490	12,861,562
Panel B: Girls							
Own household	0.0085***	0.0045***	0.0114***	0.0045***		0.0162***	0.0057***
	(0.0002)	(0.0003)	(0.0003)	(0.0003)		(0.0004)	(0.0002)
Next door neighbor	0.0010***	0.0002	0.0012***	0.0002	No female occupations in this group	0.0019***	0.0008***
	(0.0001)	(0.0002)	(0.0002)	(0.0002)		(0.0003)	(0.0002)
R-squared	0.356	0.316	0.339	0.316		0.342	0.380
Untreated mean	0.0156	0.0098	0.0204	0.0098		0.025	0.011
Observations	11,135,892	1,822,587	5,461,296	1,822,587		3,638,709	5,674,596

Notes: Regression specification and sample as described in main stacked sample table. Each column reports results within groups of occupations based on whether the occupation is above or below the median income and education score, calculated separately for male and female workers. At least one person is in the target occupation for each column.
$^*p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$.

Table A12: Additional Heterogeneity by Occupation Type

Dependent variable: Target occupation in 1940	Baseline	High self-employment	High apprentice share	Growing occupations	Shrinking occupations
	(1)	(2)	(3)	(4)	(5)
Own household	0.0381***	0.0604***	0.0509***	0.0343***	0.0426***
	(0.0002)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Next door neighbor	0.0034***	0.0037***	0.0018***	0.0031***	0.0036***
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
R-squared	0.323	0.340	0.286	0.323	0.321
Untreated mean	0.0331	0.0294	0.0140	0.0405	0.0243
Observations	26,649,896	9,852,195	4,852,186	14,477,857	12,172,040

Notes: Regression specification and boys sample as described in main stacked sample table and reported in Column 1. Columns 2-3 include occupations with high self-employment and high apprentice share respectively. Columns 4 and 5 split occupations by the percent change in employment between 1910 and 1940 and categorize them as growing or shrinking.
$^*p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$.

B Reconstructing Neighborhood Microgeography from Historical Censuses

Figure B1 shows an example page from the 1910 census, covering parts of West Division and Grace Streets in Chicago, IL.[26] Our goal is to use these census manuscript sheet data to reconstruct the microgeography of a neighborhood. To do this, we use the order in which households are listed on each sheet to determine when two households live immediately next door to one another.

This method to construct the neighborhood microgeography relies on the fact that prior to 1970, censuses were enumerated door-to-door in a predictable manner, and consequently households that are listed immediately above or below one another should be the immediate next households one would visit when walking down the street. Figure B2 provides an illustration from the 1910 census enumerator instructions showing how the enumerators were required to interview households and confirming that enumerators were required to proceed in a way that preserves microgeographic proximity.

B.1 Validating the Census Sheet Order using Street Addresses and Historical Fire Insurance Maps

To validate the census sheet-based method of constructing neighborhood microgeography, we consider an alternative method of determining neighbors using street addresses. In the historical censuses, street names were written vertically on the left hand side of census manuscript sheets. House numbers are then filled into the first column. Figure B1 shows both of these for a neighborhood in Chicago. One limitation of the street addresses is that they are only recorded in urban areas and when available. Difficulties in transcribing street names and numbers result in additional missing street addresses. In total, about 65% of

[26]These sheets contain portions of a neighborhood just east of Goose Island and the North Branch Canal, just south of W. Scott Street (called Vedder Street in the 1910 census). Grace Street no longer exists; the buildings have been demolished and the block is now vacant.

households in the 1910 census have a street address. The unavailability of street addresses for a large fraction of households is one reason why we use the census sheet order method to determine neighborhood microgeography for our baseline results.

For a subset of census sheets that contain street addresses, we use the reported street addresses in combination with historical maps from the Sanborn Fire Insurance Company (Sanborn Map Company, Various Years) to validate that the sheet order does in fact accurately reflect neighborhood microgeographic proximity. The Sanborn maps, in conjunction with the information in the census manuscript pages, reveal several cases in which using the census sheet order in fact provides a more accurate representation of microgeographic proximity than relying on street addresses alone, as we detail below.

Figure B3 shows an example of a Sanborn map for a portion of northern Chicago between Lincoln Park and the Chicago River. Each colored square represents a page in the map book with a map and house numbers for each portion of a neighborhood. For more information on the Sanborn maps, see Library of Congress (n.d.).

In Figure B4, we zoom in on a portion of the census sheet presented in Figure B1 along with its corresponding Sanborn map. By matching census pages to the maps, we can "follow the enumerator down a street." The blue numbers map a household in the census sheet to the corresponding house in the map. The first household reported on this portion of Grace Street after turning the corner from West Division is house number 1206; the map shows that it is the second physical structure on the street, but 1200 Grace Street (the first structure) is non-residential and so does not contain a household enumerated by the census. The census sheet reveals that 1206 Grace Street is a multi-family home containing four households. Because census enumerators typically only write the house number for the first household in multi-family homes, the last three households in 1206 Grace Street have missing addresses in the transcribed census data; hence, without additional imputation rules for missing addresses, the street address method would not be able to detect the geographic proximity between the four households living at 1206 Grace Street. The census sheet correctly orders the next three households, at 1210, 1212, and 1216 Grace Street; these are the last four households on the census sheet.[27]

Figure B5 follows the same enumerator a few census sheets later. On the first line of this census sheet, the enumerator finished the last house on a small section of Vedder Street (Vedder intersects Grace Street), a multi-family home at 743 Vedder Street. Following the enumerator instructions, the enumerator then turned a corner and proceeded down North Halsted Street. The street address method, which determines microgeographic proximity by identifying close street numbers on the same street, is unable to identify the close proximity between the last house on Vedder Street and the first house on Halsted Street; the sheet order method, in contrast, preserves this proximity.

Figure B5 highlights the fifth household on this census sheet, 1233 Halsted Street. As can be seen on the census sheet, the written address is difficult to read. In the transcribed census records, this address is mis-recorded as 1223 Halsted. If we used the street address method, we would therefore erroneously place this household further down its street and hence measure next-door neighbors with error; the census sheet method again avoids this source of measurement error because it does not rely on transcribed addresses to determine microgeographic proximity.

We next use the street addresses and Sanborn Fire Insurance maps to quantify how often the census sheet order correctly identifies next-door neighbors. To do this, we first restrict the sample to focal child on the same census sheet as a doctor. We then choose 1 percent of doctors at random (located on 383 distinct sheets) to verify by hand. About one quarter of these doctor-focal child pairs have addresses and can be matched to a Sanborn Fire Insurance map. We then hand check each observation and construct a measure of neighbors based on the addresses on the census image and the Sanborn map. We then locate the address of the doctor and focal child on the Sanborn map and determine if they are next-door neighbors. We cross-tabulate this with our census sheet-based measure of next-door neighbors below in

[27] 1212 Grace Street is also a multi-family home. We discuss how we use multi-family homes in more detail below.

Table B1.

For this subsample, we find that the census sheet order measure of next-door neighbors ("neighbor is a doctor") is consistent with the hand-checked approach 91 percent of the time $((230+32)/(230+5+22+32))$. Among those we hand-checked, if they did not have a doctor next door the census sheet ordering accurately captures this 91.3 percent of the time $(230/(230+22))$. If they did have a doctor next door (based on the hand-checked maps), the census sheet ordering accurately captures this 86.5 percent of the time $(32/(5+32))$.

Figure B6 presents additional suggestive evidence, using the full sample with recorded street addresses, that sheet order typically does a good job of capturing microgeographic proximity. Panel A shows that next-door households are much more likely to have a house number that is within two of one another (e.g., 1210 Grace Street and 1212 Grace Street). Panel B shows that households that are closer to one another on a census sheet are more likely to be on the same street; note that if there were no measurement error in the census sheets, the share of next-door households on the same street would still be less than 100% since census enumerators are instructed to turn corners, as shown in Figure B5.

B.2 Supplemental Sheets

A potential source of measurement error with the census sheet method arises if a family is not home at the time the enumerator first canvasses the neighborhood. As the enumerator instructions detail:

> "In case a family is out at the first visit, or in case the only persons at home are...not able to supply the required information about the members of the family, you must enumerate this family at a later visit. But no space should be left blank for this family upon the schedule you are filling at the time of your first visit unless you have positive and reliable information as to the number of persons in the family so that you will know exactly how many lines to leave blank...
>
> Use an extra sheet or sheets of the population schedule for enumerating those

> families who were out at the time of your first visit or those individuals for whom no spaces were left blank or no names were entered on the sheet of schedule you were filling at that time. At the head of these extra sheets write the word "Supplemental" and number them, finally, as the last sheets used in your work" (U.S. Bureau of the Census, 1910, p. 22-23).

Unfortunately, when the census manuscript pages were transcribed, no digital record was kept of which pages were supplemental. If supplemental sheets are used, this leads to measurement error in next-door neighbors for two reasons. First, households recorded on adjacent rows of a supplemental sheets often do not in fact reside next door to one another. Second, when a household's immediate neighbor appears on the supplemental sheet, the next household listed on the manuscript sheet would not actually be the next-door neighbor; the household would mistakenly be identified as next door to a neighbor two doors away rather than the true next-door neighbor. To the best of our knowledge, the issue of supplemental sheets has not been acknowledged in previous work using census sheets to determine neighborhood proximity. Notice that this issue will not arise when using the street address method, since we can use street addresses recorded on the supplemental sheets to place the supplemental households adjacent to their actual neighbors.

To ensure that the presence of supplemental sheets are not affecting our results, in Tables B2 and B3 we repeat our baseline doctor and stacked regression results, respectively, but drop sheets likely to be supplemental. Since supplemental sheets are included at the end of each enumeration district, we discard the last one, two, and five sheets from each district. All results are similar to our baseline results. We additionally find that no occupations were especially overrepresented on the likely supplemental sheets.

B.3 Results using Street Addresses

As discussed in Section 2, our indicator for living next door to a doctor may be subject to misclassification error in some cases, potentially biasing the OLS estimates. Misclassification could be due to skips or other inconsistencies in the enumeration process. Unlike continuous regressors, misclassification of binary regressors is necessarily non-classical. If the dummy is one (zero), measurement error can only be negative (positive), meaning that the error is mechanically negatively correlated with the true dependent variable. However, as Pischke (2007) notes, it is possible to recover the magnitude of bias by using the extent of misclassification.

Following the notation in Pischke (2007), consider the estimated $\hat{\beta}$ coefficient associated with some treatment (d_i) such as living next door to a doctor. If we only observe a version of the treatment ($\tilde{d}_i$) that misclassifies some observations then

$$plim\hat{\beta} = \beta \left[P(d_i = 1|\tilde{d}_i = 1) - P(d_i = 1|\tilde{d}_i = 0) \right]$$

We are able to estimate these probabilities using the subsample of doctors that we hand checked against the Sanborn Fire Insurance maps. In this sample, we find that among the people we classify as neighbors of doctors ($\tilde{d}_i = 1$), 59.3 percent ($32/(22 + 32)$ from Table B1) actually live next door to doctors based on the address and map information. Among people we classify as living more than one door away from doctors ($\tilde{d}_i = 0$), 2.12 percent actually live next door to doctors. Based on these measures, our estimates of the effect of living next door to a doctor are attenuated to 57.2 percent of the true effect. The causal effect is about 1.75 times larger than our estimate for doctors.[28]

One approach to dealing with misspecification error is to instrument our indicator for being next door to a doctor on the census sheet with an indicator for being next door using the street addresses. But because our treatment of interest is binary and the associated measurement error is non-classical, this commonly used IV approach to correct classical measurement error does not apply in our setting. The IV estimate will, however, still be useful as a bounding exercise, as recommended by Pischke (2007) and Aigner (1973). We

[28]Across each of the different occupations, as long as $P(d_i = 1|\tilde{d}_i = 1) > P(d_i = 1|\tilde{d}_i = 0)$ our point estimates will be biased downward.

can use the subset of census sheets with reported street addresses to construct an alternative measure of proximity.

To determine the extent to which measurement error affects our estimates that rely on census sheet ordering, we restrict attention to sheets where every household has an address and only single family homes exist on the sheet, to ensure that we are capturing neighbors with the street address ordering. This allows us to focus on neighborhoods where the enumerator actually recorded addresses and any remaining missing values are due to transcription or digitization error. We then determine that two households are next door to one another if they have the closest street number of the same parity (that is, both even or both odd) on the same street and enumeration district.[29]

Next door indicators defined using street addresses are likely also measured with error, but for reasons that we argue are likely to be orthogonal to the error in the sheet-ordered dummies (e.g. transcription errors of house numbers and street names). We exploit this fact in Table B4. In Column 1, we present first stage estimates, showing that the $NextDoorOcc1910_i$ indicator defined using street addresses strongly predicts the $NextDoorOcc1910_i$ indicator defined using the census sheet order. Column 2 repeats our baseline results that identify next door neighbors using the census sheet order, but restricted to single family homes on sheets with addresses. For this subsample, we find results that are somewhat larger than our baseline estimates. In Column 3, we repeat this baseline exercise but instead define next door neighbors using the street address measure; we keep the same fixed effects structure as in our baseline specification. For doctors (in Panel A), we find nearly identical results when defining next door neighbors with either method. In the stacked regression (in Panel B), we find point estimates that are about 38% larger when using street addresses to identify next door neighbors instead of the census sheet order. Finally, in Column 4, we present IV estimates that use the next door indicator defined using the street address as an instrument for the next door indicator defined using census sheet order. Consistent with the

[29] Bayer et al. (2025) use a similar method to identify next door neighbors using modern data on parcels from CoreLogic.

preceding discussion, we find larger transmission effects in the IV regression. Transmission effects for doctors are about 30% larger after instrumenting with the street address measure of proximity, and transmission effects are about 88% larger in the stacked regression after instrumenting.

The street address-based measures of neighbors also yield patterns of proximity consistent with the neighbor definition based on census sheet order. Using street addresses in the single family home sample, we order households on the same street and parity. We determine how many addresses away a focal child lives from someone in the target occupation for each of the 50 largest male, non-farm occupations. We then replicate the regression in Equation (5), but use the address-based definition of proximity. As seen in Figure B7, we observe a similar, nearly monotonic decline in transmission rates, consistent with patterns of social connectedness.

B.4 Multi-Family Homes

Although some addresses are missing due to transcription and digitization error, many household are missing addresses because of the data collection process. When enumerators came to an address with more than one household living there, they typically wrote down the street address only for the first household. We are able to impute missing addresses in these cases. If a household has a missing address, but there is a household above it and a household below it on the same sheet that contain non-missing addresses, we assign that household the address of the household above it. These are likely to be multi-family homes. The imputed addresses—and particularly the ability to identify multi-family homes—provide an alternative method to measure close geographic proximity. Two families that reside in the same building likely had more social interactions even than individuals who live in adjacent buildings.

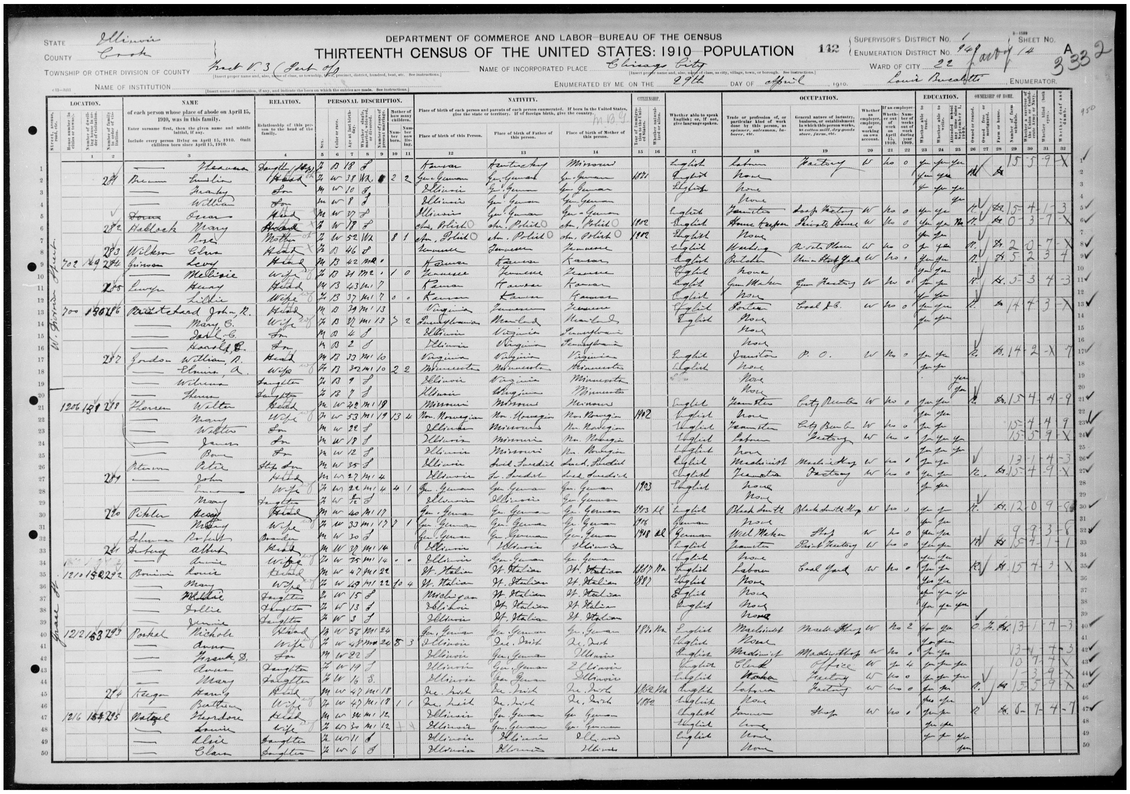
In Table B5, we show that occupational transmission effects are indeed larger when two households live in the same multi-family home. For this exercise, we exclude any multi-family homes with more than six households at the same address, as these likely represent a failure to record or transcribe an address rather than a multi-family home. We estimate

$$\begin{aligned}\text{Occ1940}_{ij} = \alpha &+ \beta_0\text{OwnOcc1910}_{ij} + \beta_1\text{SameAddressOcc1910}_{ij} + \beta_2\text{AdjacentAddressOcc1910}_{ij} \\ &+ \delta_{\text{age}(i)\times\text{occ}(j)} + \gamma_{\text{sheet}(i)\times\text{occ}(j)} + \epsilon_{ij}, \qquad (7)\end{aligned}$$

where $\text{SameAddressOcc1910}_{ij}$ is an indicator equal to one if an individual in occupation j lives at the same address as child i but in a different household and $\text{AdjacentAddressOcc1910}_{ij}$ is an indicator equal to one if an individual in occupation j lives at an adjacent address on the same street. In this specification, the comparison group is children who do not have individuals in occupation j in their same building or an adjacent address. Because enumerators must record multiple households for an address with a multi-family home, fewer of these addresses can fit on each census sheet. For this exercise, $\gamma_{\text{sheet}(i)\times\text{occ}(j)}$ is therefore a street-by-parity-by-enumeration-district-by-occupation fixed effect, comparing children on the same side of a particular street within each enumeration district.

A child who grows up with a doctor in a different household at the same address is about 0.74 percentage points, or 74%, more likely to become a doctor than other children on the same street. The occupational transmission effect from doctors in a different household at the same address is about 2.35 times larger than the occupational transmission effect from doctors at an adjacent address. We find a broadly similar story in the stacked regression. A child who grows up with an individual in a particular occupation is on average 0.29 percentage points (10%) more likely to go into that occupation than others on the same street, and 1.32 times more likely to go into that occupation than a child with a neighbor at an adjacent address in that occupation.

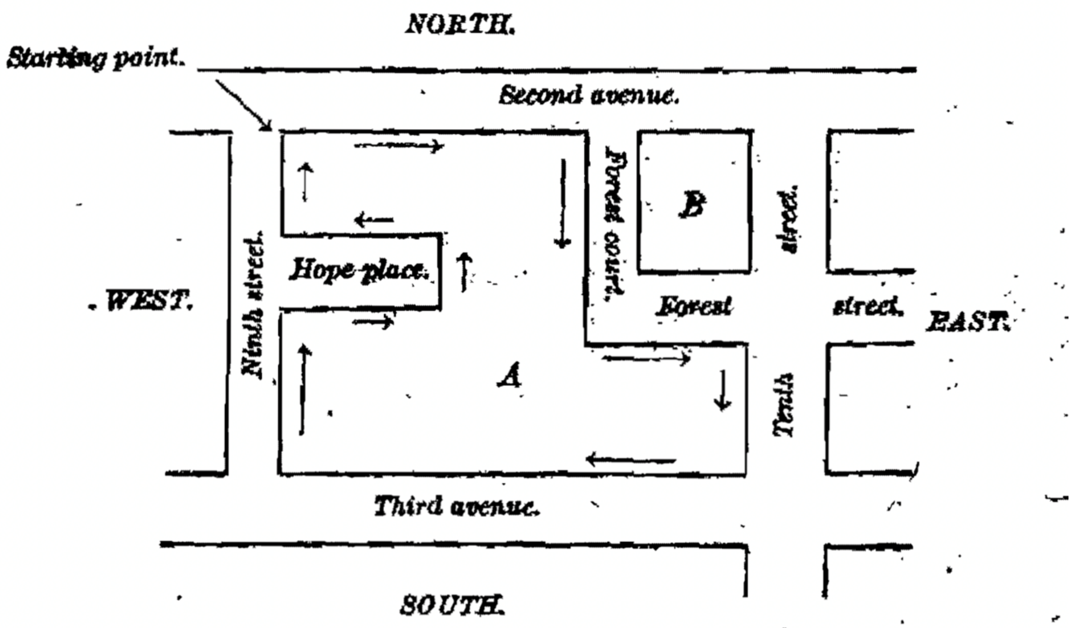
Figure B1: Example 1910 Census Manuscript Page, West Division and Grace St., Chicago



Notes: This is a representative census sheet collected by hand for the 1910 census.

Figure B2: 1910 Census Enumerator Instructions

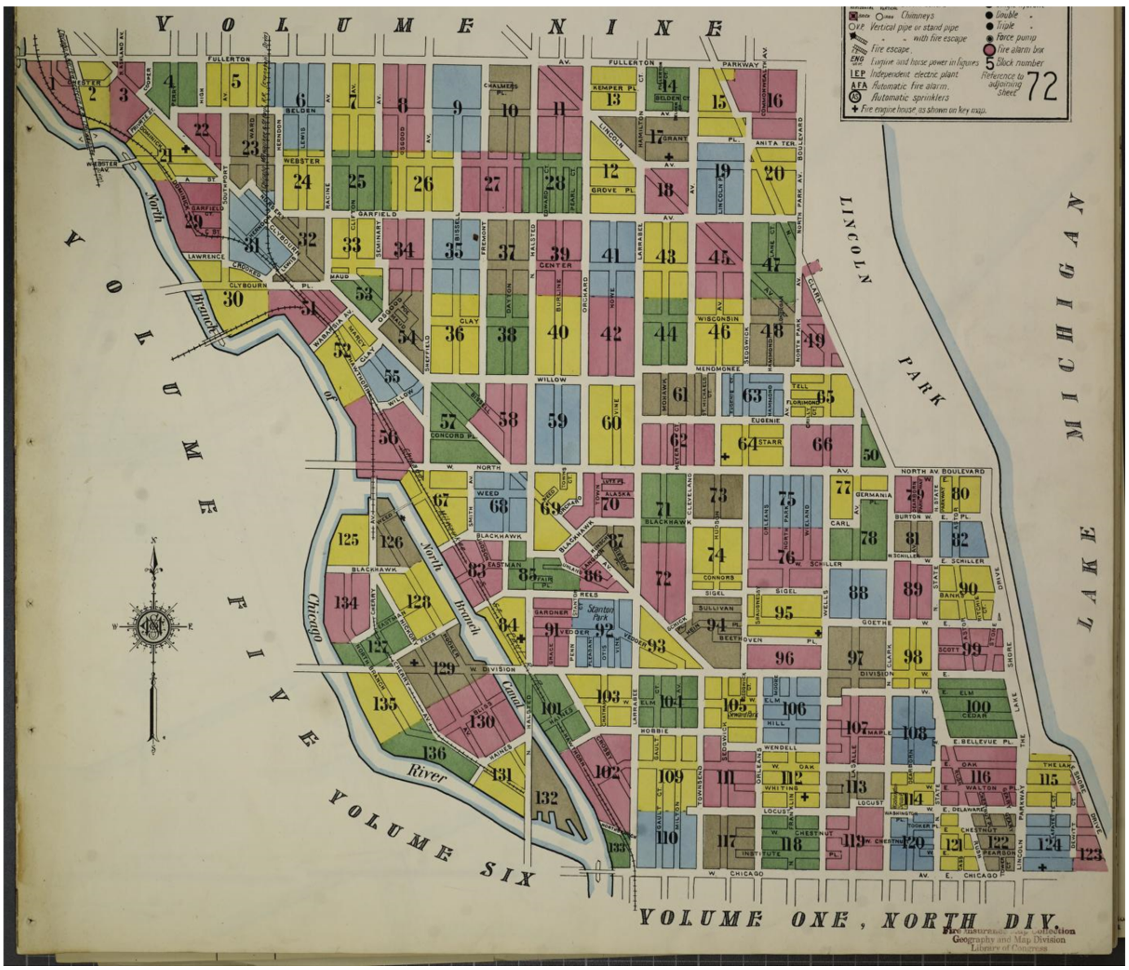
69. The arrows in the following diagram indicate the manner in which a block containing an interior court or place is to be canvassed:



(Note that block marked A is to be fully canvassed before work is undertaken in block B.)

Notes: This diagram shows an example route around a city block of a 1910 census enumerator. The predictable enumeration pattern generates the relationship between census sheet ordering and physical proximity on the historical streets.

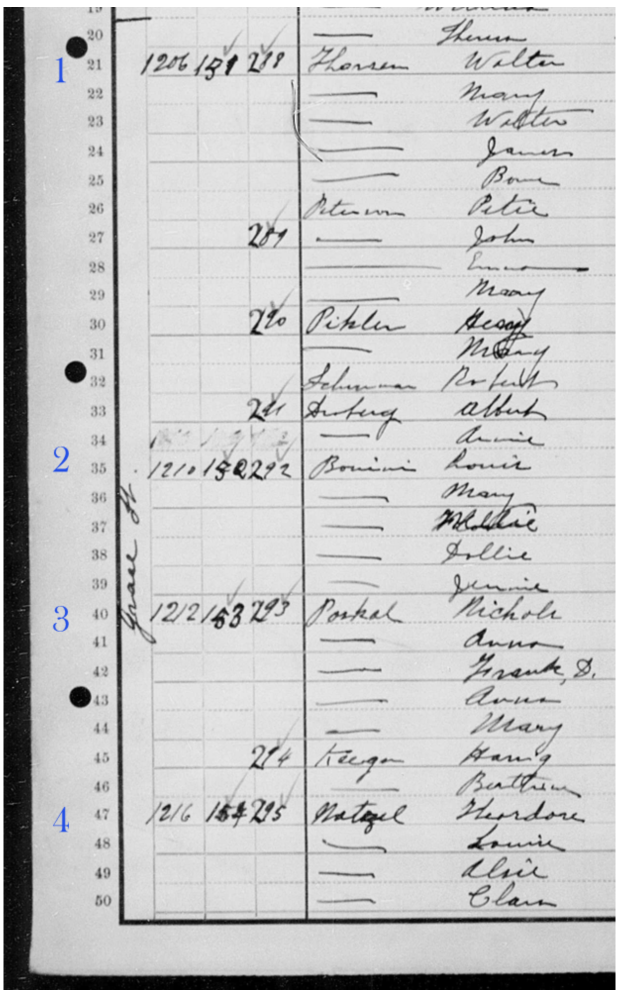
Figure B3: Example of Sanborn Fire Insurance Map for a Portion of Chicago, IL



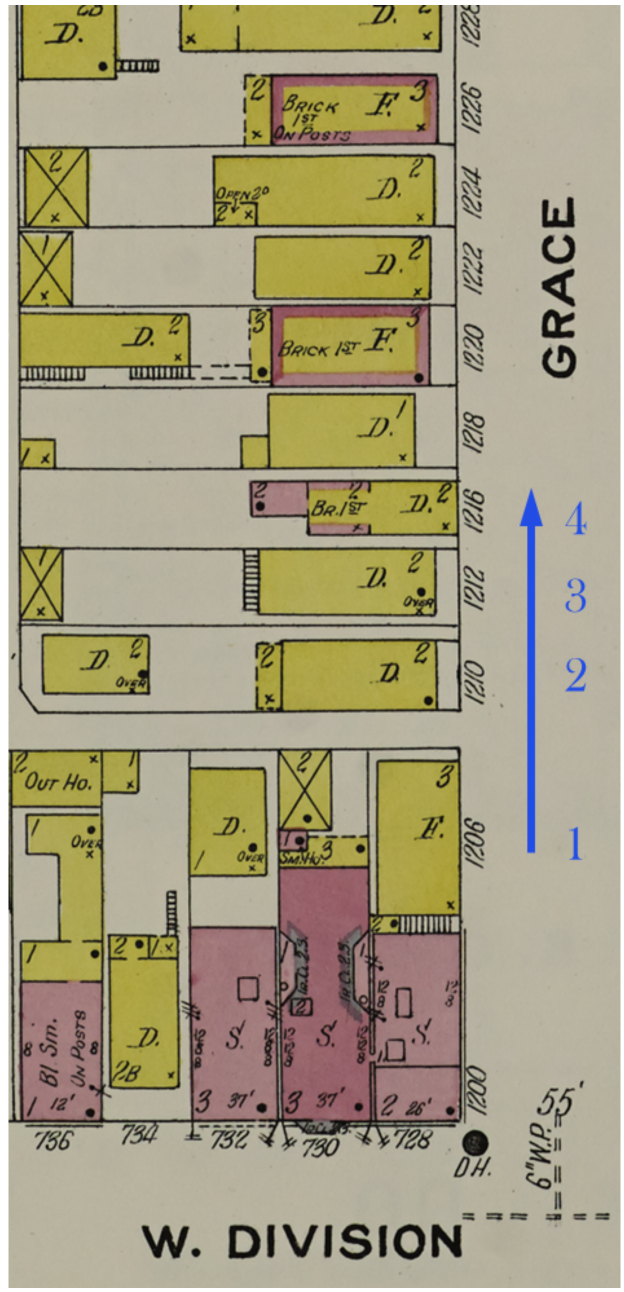
Notes: This figure shows the cover page for the Sanborn Fire Insurance Company maps for a portion of Chicago, Illinois. Each region then has a series of detailed maps, with individual buildings and plots.

Figure B4: Detailed View of Grace Street, Chicago, IL

(a)



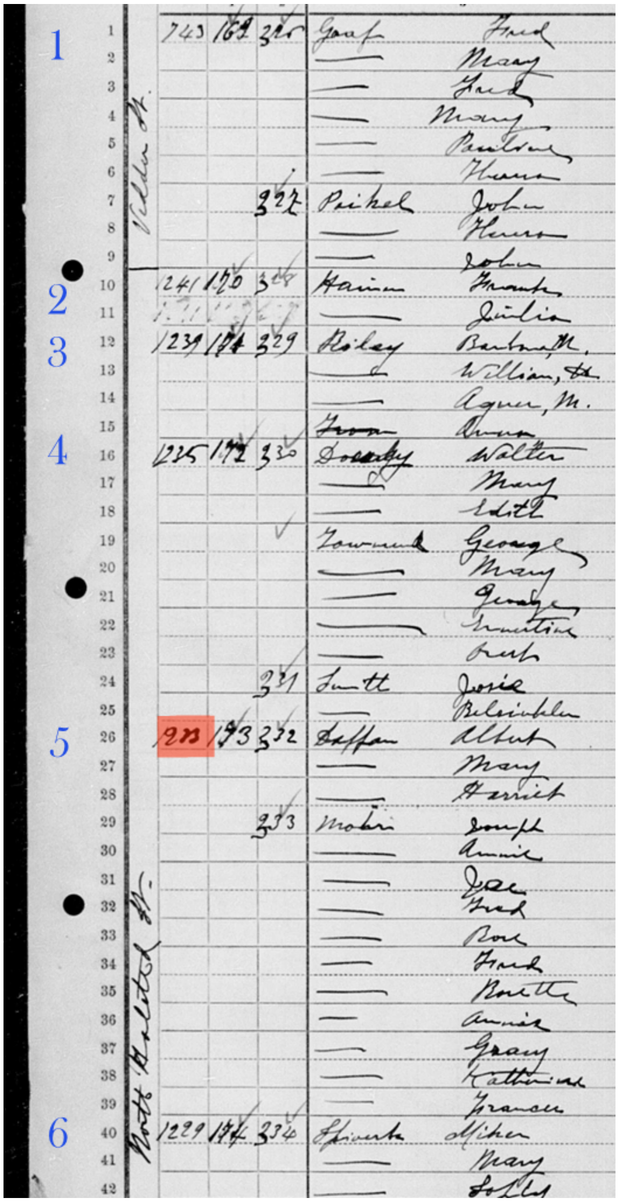
(b)



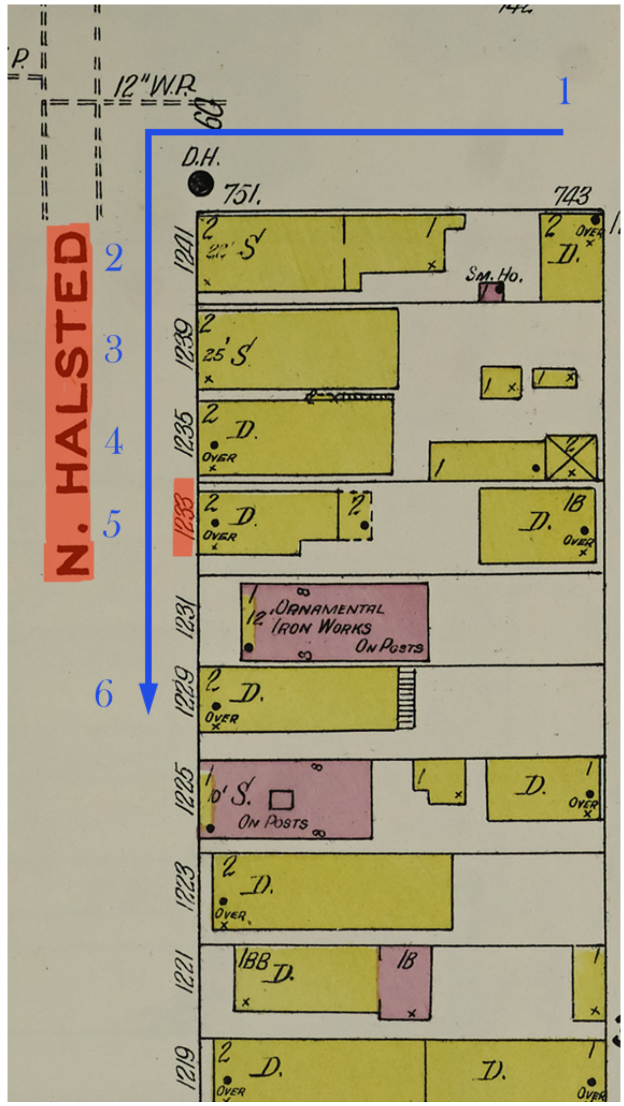
Notes: This figure shows for Grace Street the households from the portion of the census sheet (from Figure B1) and the buildings/addresses from the Sanborn Fire Insurance Map that correspond.

Figure B5: Detailed View of North Halsted Street, Chicago, IL

(a)

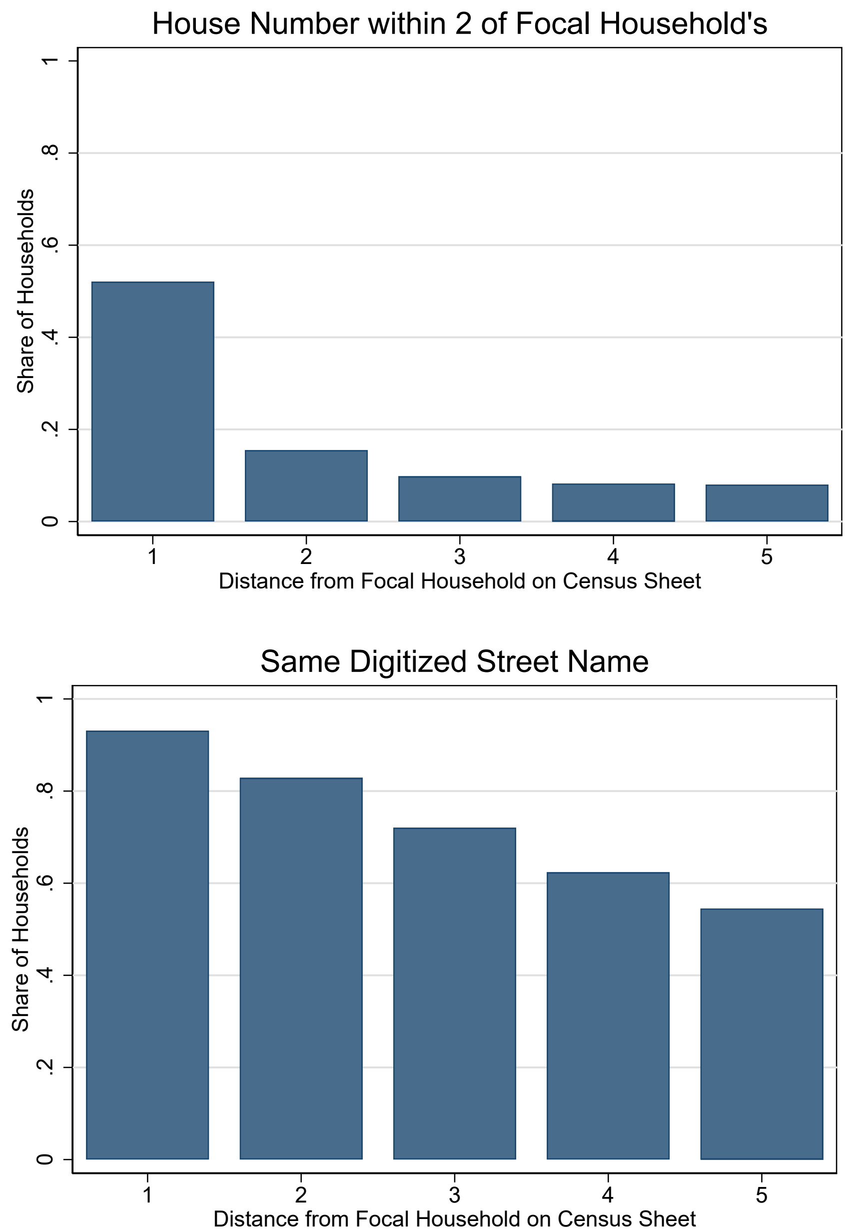


(b)



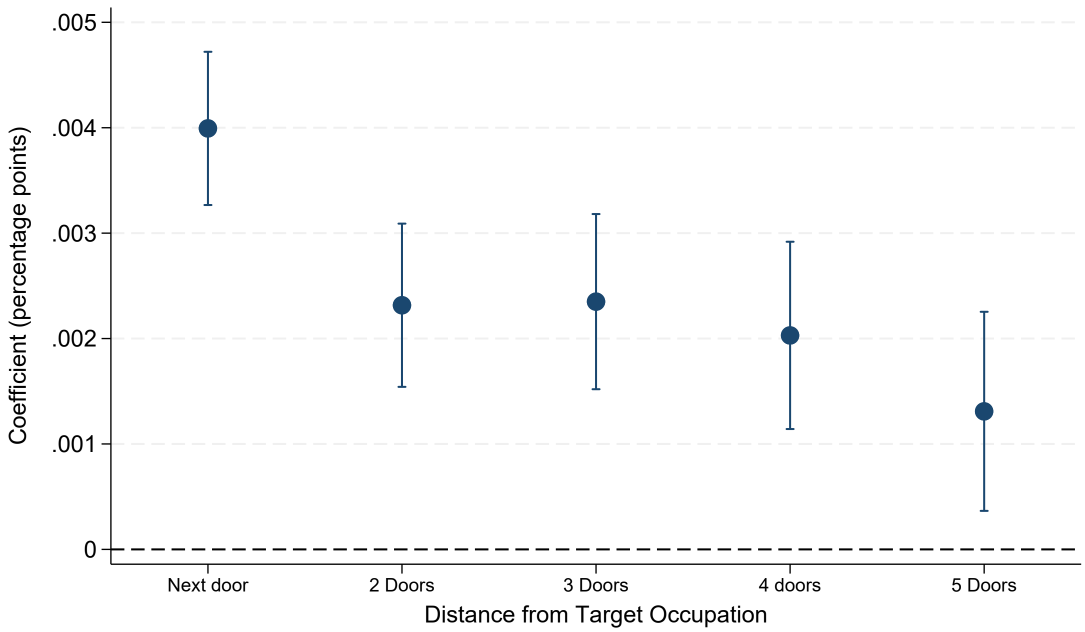
Notes: This figure shows for North Halsted Street the households from the portion of the census sheet (from Figure B1) and the buildings/addresses from the Sanborn Fire Insurance Map that correspond.

Figure B6: Sheet Proximity and Address Measures



Notes: Observations are at the household level from the 1910 full-count census. Sample restricted to enumeration districts where 80 percent of households have digitized address information. In the top panel, for each household, the share of households where the house number is within 2 digits of the house number is plotted by the number of households apart based on the sheet ordering definition. In the bottom panel, for each household, the share of households that are on the same street is plotted by the number of households apart based on the sheet ordering definition.

Figure B7: Door Distance Regressions (Average Across 50 Largest Male Occupations)



Notes: Coefficients from the stacked regression specification in Equation (5) for the 50 largest occupations for boys are plotted, where distance to neighbor is measured using street addresses rather than census sheet ordering. Sample restricted to boys between the ages of 5 and 18 in 1910 that can be linked to the 1940 census and that live on census sheets that strictly contain single family homes (one household per address). The outcome is a binary measure that equals one if the boy chooses the exposed occupation in 1940. This is regressed on binary measures that equal one if the boy's family lived next door, two doors away, etc. from a particular occupation in 1910, as well as a binary measure that equals one if the child had that occupation in their own household (not plotted). Regression includes census sheet fixed effects. 95% confidence intervals included with standard errors clustered at the household level.

Table B1: Sheet-based Measurement Error Relative to Hand-checked Sample

		Doctor Next Door (Map-Based Definition) 0	1
Doctor Next Door	0	230	5
(Sheet-Based Definition)	1	22	32

Notes: Of the 1 percent random sample of doctors on sheets with focal children, a total of 289 doctor-child pairs had both address information (for the doctor and child) and a corresponding Sanborn map for the cross-validation. The columns indicate whether or not the doctor would be labeled as a next-door neighbor based on the addresses and maps. The rows indicate whether or not the doctor would be labeled as a next-door neighbor based on the census sheet ordering. Agreement between the two measures is along the diagonal. The sample restrictions were imposed before using any address implementation discussed in section B.3 above. As such, if the child or doctor lived in a multi-family home where address information was not recorded for all households in the unit, the map-based approach would still record the doctor-child pair as next-door neighbors while the sheet-based approach would not.

Table B2: Doctor Results while Dropping Possible Supplemental Sheets

Dependent variable: Doctor in 1940	Baseline	Number of sheets dropped		
		Last sheet	Last two sheets	Last five sheets
	(1)	(2)	(3)	(4)
Own household	0.0970***	0.0969***	0.0966***	0.0957***
	(0.0018)	(0.0018)	(0.0019)	(0.0020)
Next door neighbor	0.0032***	0.0030***	0.0029***	0.0032***
	(0.0007)	(0.0007)	(0.0007)	(0.0008)
R-squared	0.327	0.326	0.325	0.323
Untreated mean	0.0078	0.0078	0.0079	0.0080
Observations	305,063	298,785	288,941	250,262

Notes: Columns 1-4 include only sheets with at least one adult in the target occupation. All columns include sheet and age fixed effects.
**p*<0.1, ***p*<0.05, ****p*<0.01.

Table B3: Stacked Regression Results while Dropping Possible Supplemental Sheets

Dependent variable: Target occupation in 1940	Baseline	Number of sheets dropped		
		Last sheet	Last two sheets	Last five sheets
	(1)	(2)	(3)	(4)
Own household	0.0381***	0.0382***	0.0382***	0.0383***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Next door neighbor	0.0034***	0.0033***	0.0033***	0.0033***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
R-squared	0.323	0.322	0.322	0.322
Untreated mean	0.0331	0.0331	0.0331	0.0331
Observations	26,649,896	26,091,776	25,241,104	22,507,744

Notes: Columns 1-4 include only sheets with at least one adult in the target occupation. All columns include age by occupation and sheet by occupation fixed effects.
*p<0.1, **p<0.05, ***p<0.01.

Table B4: Street Address Analysis

Outcome:	Next door (sheet)	In target occupation in 1940:		
Specification:	First stage	OLS	OLS	IV
	(1)	(2)	(3)	(4)
Panel A: Doctors				
Own household	-0.0954***	0.1153***	0.1150***	0.1157***
	(0.0052)	(0.0049)	(0.0049)	(0.0049)
Next door neighbor		0.0057***		0.0074**
(sheet definition)		(0.0021)		(0.0033)
Next door neighbor	0.7672***		0.0057**	
(street definition)	(0.0060)		(0.0025)	
R-squared	0.735	0.437	0.437	0.049
Untreated mean	0.0945	0.0107	0.0110	0.0107
Observations	59,205	59,205	59,205	59,205
Panel B: 50 largest male occupations (stack)				
Own household	-0.0880***	0.0370***	0.0370***	0.0374***
	(0.0007)	(0.0004)	(0.0004)	(0.0004)
Next door neighbor		0.0024***		0.0045***
(sheet definition)		(0.0003)		(0.0004)
Next door neighbor	0.7230***		0.0033***	
(street definition)	(0.0008)		(0.0003)	
R-squared	0.733	0.391	0.391	0.004
Untreated mean	0.0994	0.0323	0.0319	0.0323
Observations	6,076,859	6,076,859	6,076,859	6,076,859

Notes: The sample is limited to census sheets where each address on the sheet is a single-family home, meaning there is only one household living at each address. We also limit the sample to only sheets with at least one adult in the target occupation. All columns include sheet fixed effects. Standard errors are clustered at the household level. Column 1 regresses the sheet-based measure of next-door on the street-based measure of next-door, and whether or not the individual has someone in their household in the target occupation. This shows the correlation between the two measurements and corresponds to the first stage associated with column 4. Column 2 uses census sheet order to define neighborhood proximity. Column 3 uses street address order. Column 4 reports the second stage of an instrumental variables regression with street address order instrumenting for census sheet order. Panel A uses doctors as the target occupation. Panel B uses the 50 largest male occupations.

*p<0.1, **p<0.05, ***p<0.01.

Table B5: Multi-Family Homes

Dependent variable: Target occupation in 1940	Doctors:		50 largest male occupations (stack):	
	(1)	(2)	(3)	(4)
Own household	0.1085***	0.1084***	0.0331***	0.0331***
	(0.0049)	(0.0049)	(0.0005)	(0.0005)
Same address, different household	0.0074*	0.0080**	0.0029***	0.0033***
	(0.0038)	(0.0039)	(0.0004)	(0.0004)
Adjacent address		0.0034***		0.0025***
		(0.0012)		(0.0002)
R-squared	0.309	0.309	0.296	0.296
Untreated mean	0.010	0.010	0.029	0.025
Observations	189,668	189,668	17,785,976	17,785,976

Notes: Sample is from the main doctor and stack regressions for boys. We further restrict to streets that had at least some street numbers digitized, and restrict to street-enumeration districts that had no more than 6 households in an address. Finally, we restrict to sheets that have at least one target occupation on it. Order of addresses is determined by enumerated house number and street name instead of sheet order. All regressions include age fixed effect and street name by address parity by enumeration district fixed effects.

* $p<0.1$, ** $p<0.05$, *** $p<0.01$.