Like A Good Neighbor: The Role of Neighbors in Career Choice*

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

March 5, 2025

PRELIMINARY AND INCOMPLETE, PLEASE DO NOT CITE OR CIRCULATE

Abstract

There is growing evidence that where a child grows up has long-run, economic mobility consequences. We explore the role of neighbors in this place-based transmission. Using linked census records for over 6 million boys and 4 million girls we reconstruct historic census sheet microgeography to estimate how growing up next door to someone in a particular occupation affects a child's probability of working in that occupation as an adult, relative to other children who grew up further down the street. Living next door to someone as a child increases the probability of having the same occupation as them 30 years later by about 10 percent. High income, high education occupations are more transmissible, and homophily across ethnicity, race, and child age strengthens transmission, consistent with information and exposure channels. This occupational transmission has real economic consequences. Children who grow up next to neighbors in high income occupations, such as doctors or lawyers, see gains in income and education, even relative to other children living on the same street, suggesting that neighborhood networks significantly contribute to economic mobility.

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

*Thanks to Christian Ames, Thomas Barden, Carver Coleman, Max Dunbar, Zach Flynn, Jimena Kiser, Josh Nicholls, Abbie Sanders, Rebekah Todd, and Conner Wilkinson for outstanding research assistance, as well as audiences at Clemson, Northwestern, NYU, UC Merced, Towson, Yale, Mannheim, University of Southern Denmark, and Rutgers for helpful comments.

†University of Maryland Baltimore County, Email: mandrews@umbc.edu
‡Northwestern University, Email: ryan.hill@kellogg.northwestern.edu
§Brigham Young University, 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 (Chetty et al., 2018, 2014). The "moving to opportunity" literature finds that some neighborhoods have a causal role in improving future economic outcomes (Bergman et al., n.d.; Chetty and Hendren, 2018a,b; Chetty et al., 2016; Chyn, 2018; Haltiwanger et al., 2020; Katz et al., 2001; Kawano et al., 2024; Kling et al., 2007). But why do high quality neighborhoods matter? Previous work has focused on the role that neighborhoods play in providing access to education, health, and safety amenities and infrastructure that augment investments in children (Laliberté, 2021). 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. More specifically, we study the influence of nearby adult neighbors on the future career choices of children, a key decision impacting lifetime earnings.

Existing work finds strong intergenerational correlations between parents and children's later career outcomes (Bell et al., 2019; Corak and Piraino, 2011; Fairlie and Robb, 2007; Hvide and Oyer, 2018). But it is plausible that other adult connections in the neighborhood beyond parents also play a role in shaping children's futures. As Chetty and Hendren (2018b) and Chetty et al. (2022) show, the demographic makeup of neighborhoods and the level of connection between high-SES and low-SES individuals are some of the strongest predictors of neighborhood intergenerational mobility effects. Many neighborhood effects are highly localized (Aliprantis, 2017; Billings et al., 2022; Redding and Sturm, 2024), and local, neighborhood employment networks are strong (Bayer et al., 2008; Hellerstein et al., 2011; Tan, 2022). The ability to easily interact face-to-face, as is possible between nearby neighbors, facilitates the spread of ideas (Andrews, 2019; Andrews and Lensing, 2024; Arzaghi and Henderson, 2008; Atkin et al., 2022; Catalini, 2018; Moretti, 2021). Other studies find that geographically proximate peers influence children's schooling choices (Avdeev et al., 2023; Barrios-Fernández, 2022; Matta and Orellana, 2022).

To identify the impact of nearby adult neighbors on occupation choice, we exploit within-street variation in exposure to different careers. Our empirical strategy can be described with a simple thought experiment. Suppose that Max lives next door to Dr. Smith. Carl lives five doors away from Dr. Smith on the same street. Is Max more likely to become a doctor than Carl? We scale this thought experiment to regressions that estimate exposure effects for all children who lived near doctors (and many other occupations) in the 1910 census. Importantly, we use narrow geographic fixed effects to focus our analysis on the comparison of children living on the same census manuscript sheet (which is typically a subset of a single street), but are exposed to next door neighbors with different careers. Our analysis relies on the identification assumption that while selection into neighborhoods may not be random, selection of immediate next door neighbors (sorting of households within a particular subset of street) is as-good-as-randomly assigned. We assess this assumption using a variety of tests in our data. This strategy is analogous to work exploring the effects of neighbor's race on neighborhood choice (Bayer et al., 2022).

We study the effect of adult neighbors on career choice using historical US Census data. We take advantage of two key features of the de-anonymized historical census data and related data sets. First, we use modern machine learning methods and user-contributed linkages that allow us to track over 10 million children across censuses, from their childhood neighborhood into their adult careers. These approaches offer a large improvement over alternative linking methods, especially for girls. Second, we exploit the fact that, prior to 1970, historical U.S. censuses were collected by enumerators going door-to-door. By examining the ordering of households on census manuscript pages we can reconstruct the microgeography of a neighborhood.

To begin, we illustrate our approach by studying the effects of growing up next door to one exemplary occupation: doctors. 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, n.d.). We find that 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. To put this magnitude into perspective, having a doctor as a next door neighbor is about one-thirtieth as predictive that a child will become a doctor as is having a doctor in the child's own household; having a doctor in the same household makes a child 12.4 times more likely to become a doctor than other children on the same sheet. We show that these conclusions are robust to several alternative specifications and samples of the data, including restricting attention to sheets with only one doctor and using sheets with transcribed street addresses as an alternative measure of proximity to correct for measurement error in proximity.

Next, we extend this approach to examine the top 50 largest, non-farm occupations for men in the 1910 census and the top 25 largest occupations for women.[1] We first estimate each occupation separately, similar to our doctor analysis. While there is substantial heterogeneity across occupations, the point estimate on next door neighbors is positive for all but four of the male occupations, is statistically significantly greater than zero in 26 out of the 50 occupations, and is never statistically significantly negative. Because fewer women were participating in the labor force in the early 1900s, we have less power to estimate precise transmission effects for women, but we find similar positive (albeit lower in magnitude) effects for girls. We then combine information on all 50 occupations into a single "stacked regression" in which each child appears as an observation in the regression multiple times for each occupation that appears on his census sheet. A boy is about 10% more likely to enter into the average occupation when they live next door to an individual in that occupation than are other children on the same census sheet. Using all 50 occupations, an individual in the child's own household is about an order of magnitude more predictive than the next door neighbor. A girl is about 6% more likely to enter their neighbor's occupation compared

[1]Because female employment was highly concentrated in both 1910 and 1940 we restrict our analysis to fewer occupations.

to other girls on the street. As in our results using only doctors, we show that these findings are robust to numerous alternative specifications and subsamples.

The stacked regression approach allows us to examine heterogeneity across occupational characteristics. Here we focus on boys because of the higher demands on statistical power for estimating heterogeneity. Children are more likely to go into the occupations of their next door neighbors when their neighbor is in a high income occupation. This suggests that selective careers are difficult to enter without access to someone with experience in that occupation. These relationships might be key for passing information, opportunities, or general mentoring. Lower skilled occupations may have lower barriers to entry and therefore not require as much direct influence from a mentor.

Importantly, the occupation composition of childhood neighbors has real economic implications. Boys growing up next door to a high-income, high-education worker, such as a doctor or lawyer, have significantly higher education and earnings as adults relative to other boys on the same census sheet. Whereas growing up next door to a porter, truck driver, or laborer actually results in lower income for the children 30 years later. We also find that other high-education occupations, that do not pay as well or experience large exposure effects, such as teacher, also result in income and education gains for neighboring boys. These effects are partially driven by the direct transmission of occupation from adults to neighbor children, but there are spillovers to other, high-income and high-education occupations, suggesting a broader information mechanism, where living next to someone of particular occupations exposes children to more general information about potential occupations or encourages educational investment. For example, growing up next to a teacher increases the probability of working in other high-education occupations, like doctor and lawyer. Meanwhile, growing up next to a doctor increases the probability that a boy chooses a different high-income, high-education occupation, but not a high-income, low-education occupation, even relative to other children on the same 1910 census sheet. Results on economic outcomes are particularly striking for girls. The growth of professionalized occupations, such as teachers, nurses, and stenographers, meant that exposure to these new and promising career paths was an important influence on girls educational attainment and future income.

To understand what makes a childhood neighbor matter we document how heterogeneity across neighborhood and person-specific characteristics alters the magnitude of estimated exposure effects. Consistent with neighborhood connectedness, 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 neighbor's occupation when they share a birthplace, are of the same race, are the same gender, or when the families have similar aged children. Overall, these results suggest that children are influenced in their career choices by their interactions with people in their social networks. This influence is highly correlated with physical proximity and connectedness. This could be because the formation of social links is more likely with nearby neighbors, and also because the nature of the relationship between the children and their neighbors is stronger with closer proximity.

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 a dearth of evidence to causally identify this mechanism. We are able to provide causal estimate 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). Studies of parents' influences, however, are typically unable to disentangle the effects of environment and informational spillovers from genetic endowments and intra-family transfers such as inheriting a family business.[2] By estimating the effect of nearby neighbors, we remove these last two channels and investigate the importance of information and exposure spillovers. This paper also joins a recent literature

[2]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 in the military, respectively.

focused on the role of information, exposure, and mentors in children's (especially girls') decisions to enter high-income and high-education careers (Breda et al., 2023; Mertz et al., 2024; Olivetti et al., 2020; Porter and Serra, 2020).

This work also adds context to a growing economic history literature. Agresti (1980) and Logan and Parman (2017a) pioneered the use of census sheet order to construct new measures of residential segregation by race. 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). Rather than measuring local segregation, we innovate on the ability to use neighborhood microgeography to capture differences in children's neighborhood exposure.[3]

We advance the literature identifying neighbors in historical censuses by utilizing historical maps in conjunction with a close examination of census manuscript images to document several potential errors with existing measures of census sheet proximity. We propose several modifications to these methods to minimize these errors. We also use census street addresses when available as an alternative measure of neighborhood microgeography, as well as using the street address measure of proximity as an instrument for the census sheet measure of proximity to correct for measurement error.[4] To show the plausibility of our identification assumption using our census sheet measure of microgeographic proximity, we use the Logan and Parman (2017a) technique to construct a measure of residential segregation by occupation; to the best of our knowledge, we are the first to conduct this analysis. We find that occupations are much less residentially segregated than either race or immigrant groups.

The paper proceeds as follows. Section 2 discusses the census data, in particular how

[3]Other have used census sheet to identify potential neighbors. Quincy (2022) uses census sheet to identify individuals who receive veterans' bonuses and compare them to nearby neighbors, while Tan (2022) shows that workers' industry composition is more concentrated within census sheets.

[4]Census street addresses have been used to identify the owners of particular residences in, e.g., Akbar et al. (2019); Quincy (2022).

we construct links from children in 1910 to adults in 1940 and how we construct measure of microgeographic proximity. Section 3 lays out our identification strategy and presents evidence for why it is plausible. Section 4 presents our baseline results for doctors, for the 50 largest male occupations separately, the 25 largest female occupations, and for the stacked regressions. Section 5 explores heterogeneous treatment effects in the stacked regression. Section 6 disentangles how exposure to next door neighbors can affect economic outcomes of children when they are adults. Finally, Section 8 briefly concludes.

2 Data

Existing work has relied on occupation case studies (e.g., congressional legislatures (Dal Bó et al., 2009), the military (Greenberg et al., 2024)) or administrative employment records, like the Longitudinal Employer Household Dynamics linked to census records (Staiger, 2023), to document intergenerational transmission of occupation. Outside of the US, Norwegian registry data has been used to explore the propensity of sons to follow the occupation choice of fathers (Hvide and Oyer, 2018). However, even these datasets are not suited to answer the question at hand: does a child's exposure to adult neighbors affect their eventual career choice. To answer this question, we must not only be able to link a child to their parents' occupations, but we must simultaneously be able to observe the occupations of all of their geographically proximate neighbors. Long-running surveys like the Panel Study of Income Dynamics do not contain this information on neighbors and even recent innovations in modern census record linking through the Personal Identification Key (PIK) either restrict analysis to survey samples or the PIKed full-count censuses do not span enough time to observe the children as adults.

For this reason we exploit individual-linked full count census data for 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 including individuals' names, sex, birth year, birth place, occupation, and address.[5] Each individual's information is recorded on a census line with a family identifier, thus allowing us to connect families within 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, initially developed in Price et al. (2021) and described in more detail in Buckles et al. (2023).

2.1 Full Count Census Data

We obtain full count census micro data for 1910 and 1940 from Ancestry. This includes all of the digitized information on a census sheet including state, city, enumeration district, 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. Using the 1910 census enumeration district, sheet, and line number we can approximately reconstruct neighborhoods and create a proxy for the geographic proximity between individuals. We use our proximity measure to identify the occupations of each household's neighbors. We merge the de-identified full count census data to IPUMS records using historical IDs (HISTID) to work with cleaned, categorized measures. Using the 1940 census we can observe the child's own occupational choice as an adult, 30 years later. In some analyses we also use 1920 and 1930 full count census micro data, to explore the dynamics of occupation transmission over time.

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. (2023). The Census Tree builds on family tree data from one of the world's largest internet genealogy platforms, FamilySearch.org. FamilySearch users add and link historic records to the public profiles of their own relatives.

[5]As discussed in detail below, address was not recorded for every individual.

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 would not be possible for a trained research assistant or machine learning algorithm. Since the the Family Tree is an open-edit wiki-style platform, mistakes made by one user can be edited by other users. The Family Tree can be used to create 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 new 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 of the key innovations 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. 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.[6] As seen in Table 1 our linked sample comes from household

[6]This likely understates the true linking rate as some children in 1910 have likely died by 1940.

settings that are representative of the full population, but our linked sample is more white than the full population, which is perhaps unsurprising given the difficulty linking Black individuals during this time period.

Because women historically change their last name upon marriage, traditional linking methods perform poorly when linking women. For this reason researchers will often exclude women or focus on subsamples of women that are easier to follow over time. As such our ability to evaluate outcomes of girls is a relative innovation. However, the patterns and opportunities for employment and education were quite different for boys and girls in 1910. 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 only 31.7 percent (90.7 percent for men). Given these patterns, we view the occupational choice effect of boys and girls as both interesting, but distinct questions. As such, the set of occupations we consider and the estimated effects will be sex-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. We rely on the enumerators' record keeping to identify potential neighbors and adults the focal children might observe or interact with. Importantly for us, Census enumerators were given explicit instructions on how to collect 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)." As specified in the 1910 enumerator instruction guide, the intent was to record families on a census sheet *neighbor-by-neighbor*. This allows us to use the census collection information, such as enumeration district, sheet (i.e., page) number, line number, and in some cases address to identify likely neighbors with fine geographic precision.

We will estimate occupational transmission for many occupations, but for clarity we will describe our data creation process just looking at doctors. First, we flag all households across all 1910 census sheets that include a focal sample child 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 addresses when available) 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 a set of 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"). The last set of indicators equal one if there is a doctor on the same sheet and exactly 2, 3, 4, or 5 households above or below one's own household, each as a separate indicator. Thus when we limit the sample to focal children we can observe if they have a doctor in their own household, one next door, or one further down the sheet. For our baseline analysis we do not restrict the gender of the next door neighbor in the target occupation. However, we also explore gender homophily in section 5. We repeat this process for each target occupation separately. We will 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.[7]

Given census instructions for enumeration, sheet ordering should allow us to accurately identify neighbors and geographic proximity. Focusing on the subset of households where street address has been recorded and transcribed, we find this to be true in general. As seen in Figure 1, The probability of being on the same street falls with sheet ordering and the

[7]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 that live on a street with no other farmers, for instance a lone urban farmer in a neighborhood, are very different from the average farmer. Second, the nature of farming was changing substantially between 1910 and 1940. Third, because farmers often lived on large farms, the sheet ordering is less tied to geographic proximity. Finally, farmers often report no wage income; in the analyses in subsequent sections, we consider heterogeneity across occupations in terms of income, among other dimensions, making it difficult to know how to classify farmers. We present results using farmers in Appendix A.

gap between house numbers (e.g., 32 Mulberry Street vs 34 Mulberry Street) increases as the sheet ordering distance increases between households. Although this is true in general, it is not always the case. If no one was home when the enumerator stopped, that household is revisited later and added to a supplemental sheet at the end of the enumeration district. As such we observe cases where households with addresses are not in order on the census sheet. This will introduce measurement error in our geographic proximity explanatory variables.[8]

Unfortunately, for most households in the census it is impossible to know if the sheet ordering leads us to mis-assign next door and further-away neighbors. XX percent of households in the 1910 census contain no address information. Even among those that do record addresses, we cannot verify the sheet ordering is correct without street maps from the time period to verify that addresses are listed in order. However, we can quantify the level of measurement error for a subset of observations using historical Sanborn Fire Insurance Maps (Sanborn Map Company, Various Years). During this time period, the Sanborn Fire Insurance company constructed detailed maps of streets with house addresses in communities they serviced. Some of these maps are preserved by the Library of Congress.

To quantify the level of measurement error in the sheet order definitions of neighbors we first restrict the sample to focal child on the same census sheet as a doctor. We then choose 1 percent of doctors (on 383 sheets) at random to verify by hand. About one quarter of these census sheets have addresses and can be matched to fire insurance maps from the Library of Congress. We then hand check each observation and construct a measure of neighbors based on the addresses on census image and the fire insurance maps. For this subsample, we find that the sheet-order based next door neighbor measure is correct 91 percent of the time. In the results section we provide further evidence that our estimates are robust to measurement error concerns by excluding individuals on the last sheets in each enumeration district (to eliminate supplemental sheets where the sheet ordering does not reflect neighbor proximity),

[8]Besides laborers, we find that occupation is not strongly predictive of appearing on the last few pages of an enumeration district, which are likely to be supplemental sheets where individuals are recorded who were not home when the enumerator first stopped by.

and by using address information (where available) to construct next door neighbor status.[9]

3 Identification Strategy

Intuitively, we exploit the idea that, while households undoubtedly sort into neighborhoods, including selecting the street they want to live on based in part on how they expect it to affect their children's later economic outcomes, they have little ability to choose their immediate neighbors. In many cases individuals looking to move do not observe potential neighbors' occupations and can only move into available housing. We rely on this quasi-random assignment of neighbors' occupations within a localized neighborhood to identify the effect of adult neighbors on children's future occupation choice. To see this more formally, consider a simple linear model:[10]

$$\text{Occ1940}_{is} = \alpha + \sum_{d=1}^{D} \beta_d \text{AdultOcc1910}_{id} + \epsilon_i, \tag{1}$$

where Occ1940_{is} is an indicator equal to one if a child i that we observe in 1910 has a particular occupation in 1940. AdultOcc1910_{id} is equal to one if the adult living distance d from household i in 1910 is in that occupation. If distance proxies for exposure and interaction, and exposure to adults matter for children's later occupational choices, then β_d should be positive and declining in d. But if households sort into where they want to live, then children may be more likely to go into the same occupation as their more proximate neighbors than the average child for reasons unrelated to exposure effects. For an extreme example, children growing up in coal mining towns are more likely to become coal miners than other children observed in the full count census because that is the primary occupation in their town; they are also more likely to live close to coal miners. As a less extreme case, wealthy households, where many of the adults hold occupations like doctors or lawyers,

[9] See Appendix B for a more detailed description of census sheet neighborhood reconstruction and measurement error.

[10] We estimate a modified version of this, which we describe in detail, in Section 4 below.

may sort into the same high prestige neighborhoods; then children of high income families are more likely to go into the occupations of adults living in their neighborhoods than the average child, even if true causal exposure effects were zero, since high income parents have the resources and familial social capital to encourage their children to enter high income occupations. In these cases,

$$\text{Cov}(\text{AdultOcc1910}_{id}, \epsilon_i) \neq 0.$$

The problem with the examples given above is that we are comparing outcomes for children who live far apart from one another, and so the coefficient on distance to adults with a particular occupation is picking up not only potential exposure effects but also other characteristics of the neighborhood the child's family has sorted into. Our estimates would still be causal if we could control for all of these other neighborhood characteristics. But neighborhoods differ in a multitude of dimensions, most of which are unobservable to the econometrician, and can sometimes be large, with substantial heterogeneity within them. One path forward is to compare children who face a similar level of selection into neighborhoods. For example, families of children living on the same street might sort into neighborhoods for similar reasons, but because of housing supply constraints and information constraints families typically cannot choose how close to be to another household with particular characteristics. By reconstructing the microgeography of neighborhoods at the house-to-house level from the historical censuses, we replicate this comparison between children who reside on the same census manuscript sheet, which is usually only a portion of a typical street block. On average a sheet has only 7.6 households. Within the micro-area covered by a single census manuscript sheet, it is more plausible that relevant neighborhood characteristics and the degree of sorting are held constant. This would imply that

$$\text{Cov}(\text{AdultOcc1910}_{id}, \epsilon_i | s) = 0,$$

where s is the census manuscript sheet. In our preferred regressions specifications below, we include census sheet fixed effects. In many cases, we focus specifically on the influence of an adult living next door to a focal child. Hence all that is required is that, conditional on being on the same census manuscript page, individuals cannot choose their immediate next door neighbor. Then, conditional on appearing on the same census manuscript sheet, the occupation of a focal child's next door neighbor is as good as randomly assigned. Similar identification has been used in housing transaction data in a modern context (Bayer et al., 2022).

3.1 Plausibility of the Identification Assumption

Our identification strategy would fail if households sort *within census manuscript sheet* in ways that are correlated with both their neighbor's occupation and their child's proclivity to work in that occupation. 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.

We think this is unlikely for several reasons. First, we find little evidence that adults having the same occupation are able to sort into adjacent houses. For the 50 largest, non-farm occupations in the 1910 census, among the sheets that contain at least one individual with each occupation, a large fraction of sheets contain only one occurrence of that occupation. With the exception of resource-based occupations that are geographically constrained like mining, most people are the only person in their occupation on their census sheet. (Figure **??**). In the results below, we show that our results are robust to focusing on only the sheets that have one of each occupation, in which case it is by construction impossible for households with the same occupation to sort to be adjacent to one another. Occupation-specific measures of segregation, analogous to measures of racial segregation in Logan and Parman (2017a), show similar results. Across occupations, occupation-specific segregation is close to zero (consistent with as good as random placement) and substantially lower than racial segregation or even ethnic segregation for nearly all occupations.[11] See Appendix C for a detailed analysis of neighborhood sorting.

Second, we show that the effects are not driven by households with similar occupations (and potentially similar aspirations for their children's future careers) sorting next door to each other. As seen in the results below, if we control for occupation characteristics of the household head the effect of the next door neighbor's occupation is virtually unchanged. Results are also similar if we include household head occupation fixed effects. This specification allows us to compare the effect of neighbors occupation holding one's own household occupation fixed. For example, we can look among children of lawyers to see if a lawyer's son that grew up next door to a doctor is more likely to become a doctor than a lawyer's son in the same neighborhood that grew up further down the street from the doctor.[12]

Finally, we repeat our analysis for a broad range of occupations that vary across skill-level, income-level, and prestige, and find significant patterns of transmission. Once we move from particular occupations (such as doctor) to the full set of large occupations the flavor of sorting necessary to invalidate our identifying assumption becomes more nuanced. Remaining threats to identification must be match-specific sorting where families sort into homes next to people in particular occupations because their child has an idiosyncratic proclivity for that particular occupation in a way that is unique and not driven by general characteristics like occupational income or prestige.

Although this level of sophisticated sorting seems unlikely, we are able to rule it out using an alternative strategy that exploits deaths of people in particular occupations by linking the census records to the Database of Human Lifespan (). If our estimated effects were driven by match-specific sorting we would expect similar transmission, even if the neighbor in the focal occupation died before the child could be "exposed" to the career. As we show in the results section, the effects for these children are smaller.

[11]The exception is once again for resources based occupations like farming, mining, and lumbermen.

[12]We are also able to see if a neighbor's occupation predicts the household head's characteristics. Although this is true when comparing all children in the census (consistent with residential sorting) the predictive power falls dramatically once we include finer geographic fixed effects (see Appendix C for details).

4 Results: Doctor Case Study

4.1 Doctors: A Case Study

To introduce our empirical analysis, we begin with an illustrative example, focusing narrowly on boys in 1910 and one particular occupation: Doctors. In the next section, we broaden the analysis to many occupations and both girls and boys.

Here, we present results based on this regression model:

$$\text{Doctor1940}_{is} = \alpha + \beta_0 \text{OwnDoc1910}_i + \sum_{d=1}^{5} \beta_d \text{NeighborDoc1910}_{isd} + \text{Age}_i + \gamma_s + \epsilon_{is}, \quad (2)$$

where Doctor1940_{is} is an indicator for the focal child i on census sheet s listing doctor as their occupation in 1940, OwnDoc1910_i is an indicator for the focal child living with an adult in their household in 1910 that has listed doctor as their occupation, and the indicators $\text{NeighborDoc1910}_{id}$ 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 doctor as their occupation. For example, if $\text{NeighborDoc1910}_{is1} = 1$, then the focal child has at least one doctor that lives one house away (that is, lives next door) in either direction on their side of the street. Age_i is a fixed effect for the focal child's age in the 1910 census. Finally, γ_s is a census sheet fixed effect, and ϵ_{is} is an error term that we cluster at the household level.

We first want to assess whether neighbors have a larger influence on childrens' occupation choices when those neighbors are geographically closer relative to neighbors who live further down the street. Figure 2 shows this relationship (the coefficient for OwnDoc1910_{is} is estimated but not reported for ease of interpretation). For doctors, we find that the immediate next door neighbor is the only neighbor that has a significant positive influence. For this reason, and to simplify the exposition, for the remainder of the paper we focus only on the immediate next door neighbors. Focusing on next door neighbors only may be a conservative test, since households only two doors away from a doctor are now included as part of the untreated group. To the extent exposure effects are operating farther than next door, this will bias against estimating a positive effect of next door neighbors on occupational choice.[13]

We estimate the own household doctor effect and the immediate next door neighbor effect as follows

$$\text{Doctor1940}_{is} = \alpha + \beta_0 \text{OwnDoc1910}_i + \beta_1 \text{NextDoorDoc1910}_{is} + \text{Age}_i + \gamma_s + \epsilon_{is}, \quad (3)$$

where $\text{NextDoorDoc1910}_{is}$ is an indicator equal to one if one of the households next to i has an adult who is a doctor.[14] The neighbors further down the street are now included in the omitted category. The inclusion of census sheet fixed effects (γ_s) 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 explore various levels of location fixed effects in each column of Table 2 to help assess our identification assumptions. Column 1 does not include any location fixed effects and includes all households with a child in the 1910 census. Column 2 also does not include any location fixed effects, but to include a consistent sample in which there is identifying variation in more restrictive specifications, we restrict attention to census sheets with a child in the 1910 census and for which there is at least one household with a doctor. Column 3 adds broad geographic fixed effects for state-city-enumeration district, and Column 4 adds our preferred, more narrow sheet-level fixed effects. Across all four specifications, we see a very consistent positive effect of own household transmission. Boys that grow up with a doctor in the home are 9.54-9.83 percentage points more likely to become a doctor than other boys on the street. We report the mean probability of becoming a doctor for the counterfactual group – boys with no

[13] To estimate Equation 2 for households five doors away from the doctor or less, we require census sheets with at least six households (the household with the doctor and five neighbors). Sheets with several larger households may have fewer than six households on the sheet, and so looking at the effects on children farther away from a doctor mechanically reduces the sample size. Also, if we examine effects up to five doors down, households at the top and bottom of the census sheet are treated differently than households in the middle of the census sheet, since a household at the edge of a sheet only has neighbors in one direction. This is another reason to focus only on next door neighbors.

[14] In other words, NextDoorDoc1910_i is equivalent to $\text{NeighborDoc1910}_{is1}$ from Equation 2.

doctor in the home or next door. Those with doctors in the home are at least 10 times more likely to become a doctor than their peers on the same sheet, a result that matches other estimates of the intra-household transmission rate in other contexts. Our key coefficient of interest is the effect of living next door to a doctor in 1910. For this coefficient, the level of fixed effects is relevant for our identification assumption. The effect with no fixed effects includes not only the causal effect of exposure, but also a selection effect where kids that live in the same neighborhood as other doctors may have been more likely to become doctors even absent exposure. 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 farther 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 the remaining coefficient as the causal exposure effect. Boys that live next door to a doctor are 0.32 percentage points more likely to become a doctor, or 41% more likely than the boys who live further down the street.[15]

4.2 Robustness

We provide three sets of robustness checks, which we briefly summarize here and provide further analysis in the Appendix. First, we probe our main results using more restrictive regression specifications and alternative sample definitions. Second, we consider an alternative method for constructing neighbor proximity using street addresses. Third, we conduct a placebo test focusing on cases where doctors died shortly after the census enumeration.

[15]In Appendix Table 4, we repeat this exercise but link children to the 1920, 1930, and 1940 censuses. For both own-household and next door coefficients, effect sizes are smallest in 1920, of intermediate value in 1930, and largest in 1940. This is consistent with doctors being an occupation that requires substantial investments in human capital that take time. However, if we scale by the untreated mean the relative effects are similar across the three decades, suggesting that we do not lose out on addition information by focusing on 1940 outcomes.

4.2.1 Alternative Specifications and Samples

In the appendix, we show results for a variety of alternative sample restrictions and regression specifications to bolster the credibility of the main results. In Table XX, we restrict to sheets that only have one doctor in order to simplify interpretation by avoiding cases where kids are treated with exposure to multiple doctors in their home and neighborhood. The coefficient on these sheets is smaller by about a third, but still statistically significant from zero.

Next, we present results using additional sets of fixed effects. In all cases, we find similar results: having a doctor in one's own household increases the likelihood that a boy becomes a doctor in 1940 by about ten percentage points, and living next door to a doctor also significantly increases the probability that a boy becomes a doctor in 1940.

4.2.2 Alternative Proximity Measure: Street Ordering

Another potential confounder we consider is the possibility of measurement error in street ordering as described in Section 2. One possible issue is that enumerators skip houses if they initially can't find the family at home and place them instead on a supplemental sheet at the end of their stack of sheets. These skips could attenuate our estimates as households appear closer together on the sheet than they actually are physically on the street. In Appendix Table **??**, we re-run the specification in Column 4 of Table 2 but omit the last several pages from each enumeration district to ensure that results aren't biased by supplemental sheets. We find nearly identical results when dropping the last 1-5 sheets from each enumerator's stack of pages, minimizing concerns about attenuation bias.

A more careful approach to dealing with bias from enumerator misordering is to recalculate our proximity measures using only the recorded street addresses (where available) and ignore the ordering on the enumeration sheets altogether. As discussed in Section 2, our indicator for living next door to a doctor is subject to misclassification error, 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 **?** notes, it is possible to recover the magnitude of bias by extent of misclassification.

Following the notation in **?**, if we estimate the $\hat{\beta}$ 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 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 biased downward by 57.2 percent. The causal effect is about 1.75 times larger than our estimate for doctors.[16]

Because our treatment of interest is binary and the associated measurement error is non-classical, the commonly used IV approach to correct classical measurement error does not apply in our setting. However the IV estimate will still be useful to compute in a bounding exercise as recommended by **?** and **?**. For a subset of census sheets, enumerators list street names and numbers. As such, we can define a measure of proximity using the street-ordering on pages. Next door indicators defined under this definition 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

[16] 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.

this fact in Table 3. In Column 1, we estimate our preferred version of Equation 3 (including census sheet fixed effects) when restricting attention to households that also have recorded street addresses. Both the coefficients for having a doctor in the focal child's own household and next door are slightly larger than the baseline estimates in Column 4 of Table 2, although when we restrict attention to the sheets with street addresses the baseline probability that children will grow up to become doctors is substantially larger as well. Having a doctor in one's own household makes a child about 20.8 times more likely to become a doctor and having a doctor live next door makes a child 59% more likely to become a doctor. In Column 2, we use the street addresses to measure proximity. We consider a household to be next door to the focal household if it has the closest house number (either larger or smaller) on the same street.[17] To completely separate this analysis from the sheet based analysis we include street name-by-enumeration district fixed effects rather than sheet fixed effects to facilitate the within street comparison.

We find results that are qualitatively similar to our results in Column 1, although coefficients for both own household and next door are slightly smaller in magnitude. When using street addresses, growing up next door to a doctor makes a child 0.28 percentage points, or about 46%, more likely to become a doctor. Since the measurement error in both of our measures of household's proximity are believed to be orthogonal, in Column 4 we use the measure of NextDoorDoc1910$_i$ using recorded street addresses as an instrument for NextDoorDoc1910$_i$ using census sheet order. Consistent with the instrument resolving classical measurement error in household proximity, the coefficient on NextDoorDoc1910$_i$ is 2.4 times larger than in our estimates using only the census sheet order in Column 2 and 3.1 times larger than in our estimates using only the street address in Column 3. The OLS and IV estimates bound our estimated effect of growing up next to a doctor on the probability of

[17]One of the reasons we use the sheet ordering as our preferred specification is because addresses are not recorded for about half of the census observations. To ensure that we are capturing neighbors with the street address ordering, we restrict the sample to enumeration districts where at least 80 percent of households have address recorded. This allows us to focus on neighborhoods where the enumerator actually recorded addresses and missing values are due to transcription or digitization error.

becoming a doctor between 0.32 percentage points and 0.87 percentage points. Our measurement error corrections place the estimate squarely between these estimates (0.56 percentage points).

4.2.3 Placebo Check: Doctor Deaths

Thus far, we have operated under the assumption that within a street the placement of doctors relative to children is as-good-as-random. Despite the robustness evidence on sorting that we have presented, lingering identification concerns could arise about selection into immediate proximity to doctors. The ideal experiment to identify the causal effect of occupation exposure would eliminate doctors next to particular children on a street at random, to determine if children that grow up next to doctors are selected differently than other children on the street. In this section, we analyze a natural experiment that is similar in spirit to randomized exposure, exploiting doctor deaths. Using the Database of Human Lifespan (**?**), we tag doctors with documented deaths between 1910 and 1915, the five-year window following the 1910 census where we measure occupation exposure. Then we compare outcomes for boys that live next to a doctor in 1910 who dies before 1915 to boys that live next to a doctor who survived past 1915. This quasi-random shift in the duration of exposure to a doctor neighbor can help solidify the causal link between exposure and outcomes. In both cases, the boy's family similarly selected into homes next to doctors, but we can determine if boys with less exposure experience similar treatment effects. Due to the cross-sectional nature of the census snapshot, we can't measure the exact duration of doctor exposure for children of different ages, but we can assert that a doctor death shortens the exposure period, likely during very formative ages. Table XX shows regression results for the specification in Equation 3, but focusing on sheets with exactly one doctor. We split the regression sample in columns 1 - 3 by whether the doctor survived past 1915, died before 1915 (6% of the sample), or had a missing death date (31% of the sample) respectively. We find that doctor death sharply decreases the probability that a child becomes a doctor both for the doctor's own children as well as the next door neighbor. Children of surviving doctors have an 11 percentage point higher probability of becoming a doctor compared to the untreated mean compared to only 4 percentage points for children of doctors that die. The effect of next door neighbors similarly falls from 0.3 percentage points down to -0.1 percentage points, a coefficient that is not statistically distinguishable to zero. These findings strongly support the causal influence of exposure on occupation choice, and helps to rule out the possibility that effects are biased by selection or sorting in street ordering.

5 Results: Heterogeneity by Occupations

Now that we have established a significant next door neighbor transmission effect in the case of doctors, we can consider a broader set of occupations, including some of the growing professions available to women in the time period. We estimate effects occupation-by-occupation, but also implement a stacked regression approach to understand how occupational characteristics influence exposure effects.

For each occupation we replace the indicators Doctor1940_{is}, OwnDoc1910_{i}, and $\text{NextDoorDoc1910}_{is}$ with analogous indicators for that particular occupation. To simplify the analysis and to ensure that we have a sufficient number of individuals in each occupation, we focus on the 50 non-farmer occupations that have the largest representation among men in the 1910 census, and the 25 largest non-farm occupations for women.

These coefficients are provided for boys and for girls and for each occupation separately in Figure 3a. Both within household and next door neighbor effects are presented. In both cases, we scale the coefficients by the mean among kids with no individuals in the occupation in their household or next door so that the plotted coefficients can be interpreted as percentage changes relative to the untreated mean.[18] Both are separately sorted by effect size and solid filled markers indicate that the coefficient was statistically significant at the the 5% level. We find that, for all 50 of the largest occupations for men, having

[18] We plot the un-transformed coefficients in Appendix Figure A1.

an individual in the same household significantly predicts that the child is more likely to enter that occupation. There is, however, substantial heterogeneity in effect size across occupations. For own-household transmission for boys, doctors are indeed ranked highly, with the fifth largest estimate in percentage terms. But other occupations of various type, such as brickmasons, bakers, meat cutters, blacksmiths, clergymen, barbers, and lawyers also predict that kids are more than ten times more likely to go into the occupation if an adult in their own household is in 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 are likely to choose the occupation of a parent one to four times as often as an untreated peer, with a range of occupations affected, including tailors, nurses, teachers, and musicians among others. The muted transmission effect for girls may be due to lower overall employment among women (in both 1910 and 1940) or expanding occupational opportunities that were available to women in 1940 but not 1910.

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 significantly so. The most predictive occupation is for brickmason, the same as for the own-household coefficients, although the next several largest differ between the two lists. Doctors once again rank highly, in eighth place. Girls show more mixed evidence on neighbor transmission. While the majority of the top-25 occupations have positive transmission, only five are statistically significant. So although the effects for girls are suggestive, we lack the same level of statistical precision for women.

These plots make it clear that some occupations are more likely to transmit to neighboring children. However, it is difficult to draw conclusions about the channels of occupation transmission by inspecting individual cases, so we introduced a stacked regression approach to systematically explore patterns of heterogeneity and causal mechanisms. To do this, we take each occupation-specific regression sample 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 estimate

$$\begin{aligned}\text{Occ1940}_{ijs} =& \alpha + \beta_0 \text{OwnOcc1910}_{ij} + \beta_1 \text{NextDoorOcc1910}_{ij} + \text{Occ}_j \times \text{Age}_i \\ &+ \text{Occ}_j \times \gamma_s + \epsilon_{ijs},\end{aligned} \tag{4}$$

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 street. Since the same household is used multiple times, in this specification we two-way cluster standard errors at the household level and at the individual i level. Our fixed effects for age and census sheet are interacted with occupation to make the same within-occupation comparisons that we made in Figure 3.

We present average effects across occupations from the stacked regression in Table 5a for boys, and Table 5b for girls. Columns use the same location fixed effects as in Table 2. In Columns 1 and 2, we have no 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 for the average occupation. However, as we include finer geographic fixed effects, and compare children living next door to someone in the target occupation to children that are more and more geographically proximate, the estimated effect of $\text{NextDoorOcc1910}_{ij}$ falls.

Our preferred specification in Column 4 includes a fixed effect for each census manuscript sheet, interacted with a fixed effect for each occupation, essentially accounting for locational sorting up to the sub-street level. In this specification, growing up with an adult in the same household in an average occupation increased the probability that the boy enters that occupation by 3.81 percentage points, or about 115%. Growing up next door to an individual with an average occupation increases the probability that the boy enters that occupation by 0.34 percentage points, or about 10.3%. For both the own-household and next door coefficients, the estimates in Table 5 are smaller than those for doctors in Table 2, which is consistent with our findings in Figure 3 that exposure effects for doctors are larger than for an average occupation. For girls, shown in Table 5b 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 neighbor's occupation. The relative increase is about half that for boys off of 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. In Appendix A, we perform the same set of robustness tests as we do for the doctors and find that the results are robust to alternative specifications, using street addresses to account for measurement error, examining earlier outcomes in 1920 and 1930, and exploiting pre-mature deaths as a placebo test.[19]

5.1 What Occupation Characteristics Matter?

To estimate heterogeneity by occupation characteristics we re-estimate the stacked regression from Equation 4 with census sheet fixed effects but restrict the sample to only include stacked panels for occupations that have particular characteristics.[20] In particular we focus on variation in occupational income levels and education requirements, as captured by

[19]As described above, our baseline analysis does not consider farmer owners or farmer laborers. We find larger effect sizes when we include farm owners and laborers among our occupations in the stacked regressions (Table A1).

[20]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 and associated occupations.

occupation income scores and education scores. We plot the next door neighbor effects, scaled by the untreated mean, in Figure ??.[21] Exposure effects for high income effects are slightly larger on average, around 12 percent, while effects for high education occupations are slightly lower, at 9 percent. However, there are stark difference across the four mutually exclusive groups when you cross income and education level. Exposure effects are larger for high income occupations that have higher and lower education requirements, although the point estimate is larger for high income occupations with low educational attainment. Effects are also larger for low-income, low education occupation. However the exposure effects for low paying jobs that require high educational attainment (like teacher or clergyman) are significantly lower, at less than 5 percent. Growing up next to someone in a high income occupation or a low income, low education occupation increases the probability that a child works in that occupation 30 years later relative to other children that grow up on the same street.[22]

6 Exposure Effects on Economic Outcomes

Growing up next to someone in a particular occupation changes a child's likelihood of working in that occupation, but does this change the child's long-run outcomes? Some high-income occupations, such as doctor and lawyer have higher exposure effects, (see Figure 3), but several low-income occupations, such as brickmason, tailor, and waiter are also highly transmissible. It is possible that the type of neighbor you grow up next to could affect a child's earnings and educational attainment as an adult. We consider income and education effects for both boys and girls, whose economic outcomes may have been even more elastic to network effects than boys at the time.

[21] Unscaled estimates, along with estimates for having someone of a particular occupation in your own household are provided in Table 8.

[22] 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, suggesting our main neighbor spillover finding is likely not driven by direct transmission of business, human capital, or structural labor market changes (see Table A3.

6.1 Effects on Adult Income

Using the same approach as in Equation 3 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 wage income in 1940.[23] As seen in Figure 4a, many neighbors' occupations have a significant impact on boy's 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. Relative to average annual income of $1,071 in 1940 among similarly-aged (30-50), working men, this represents a 4.2-4.7 percent increase in annual income. Importantly, this is relative to other children who were living in the same neighborhood in 1910, and thus accounting for socioeconomic neighborhood sorting.

There are other high-income occupations such as manager and locomotive engineers that also lead to increases in the neighbor child's adult income, but there are also low- or middle-income occupations such as teacher, clergyman, and clerical worker that lead to significant increases in neighbor children's adult income. Once again, since these estimates are relative to other children living the same neighborhood, this likely speaks to the occupation choice counterfactual. Teachers in general did not live in the most affluent neighborhoods, but living next door to a teacher exposed a child to opportunities that were relatively better than their peers.

There are also occupations that significantly *reduced* next door children's wage income in 1940 relative to other neighbor children. 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.

Girls also have a few professions whose influence increases income. Living next to a

[23] 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 occuaption.

milliner, bookkeeper, or stenographer during childhood increases annual earnings by $15-20, a 7-9 percent increase relative to average annual wage income ($204) and a 2-3 percent increase relative to average annual wage income among employed women. These results likely differ in magnitude from men because of differences between men and women in the extensive margin decision to participate in the labor force during this time period, which is influenced by neighbors occupation, as seen in Figure **??**. Similar to boys, living next to laborers or private laundresses actually led to lower annual income relative to other girls from the same neighborhood.

Although it is clear that living next to someone in a particular occupation as a child has an impact on future earnings, the channels through which this operates is not clear. It could all be driven by an occupation match effect if, for example, growing up next to a doctor increased a child's likelihood of becoming a doctor, but had no other effects. However, it is also possible that having a doctor for a neighbor increases exposure to information about high-paying jobs in general, or the human capital requirements necessary to qualify for a high-paying job. As noted above there are some low income, high education occupations, such as teacher or clergyman, that are not very transmittable to neighbor children (see Figure **??**) but do lead to income gains (see Figure 4a). For this reason, we next explore how childhood neighbors' occupations affect children's educational attainment.

6.2 Effects on Educational Attainment

In Figure 5, we find that living next door to someone of a particular occupation has a similarly heterogeneous impact on boy's educational attainment. Living next to a doctor or lawyer in 1910 increased average years of schooling by nearly 0.4 relative to other children on the census sheet. But even many of the low income, high education occupations that led to income gains lead to increases in educational attainment as well. 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, actually reduced children's educational attainment. Figure 5b shows similarly large effects for many professional occupations with educational requirements. Girls living next to nurses, bookkeepers, and teachers attain about 0.2 more years of education than the other girls on the street. Living next to some occupations decreases educational attainment, including laundresses, laborers, and kindred workers. These effects speak to the network and mentoring channel of occupation exposure.

Since neighborhood exposure to many occupations changes a child's ultimate educational attainment, it seems plausible that growing up next door to a doctor changes more than just the child's probability of becoming a doctor. In Figure **??** we plot the scaled next door coefficient from Equation 3 for boys, but change the outcome to be a binary measure for being in each of the top five highest paying, large occupations, including lawyer, manager/official/proprietor, foreman, or compositor/typesetter. We find that living next to a doctor in 1910 significantly increases the probability of being a lawyer (16 percent) or manager (8 percent) in 1940 relative to other boys on the 1910 census sheet, but it does not increase the probability of being in a high paying, low education occupation such as foreman or compositor/typesetter. Living next door to a doctor as a child significantly increased the probability that the child entered a high-paying, high-skilled occupation that requires educational training, but not high-paying manual or trade occupations. This is consistent with the effect on increased educational attainment. The benefit of growing up next to someone in a high-income occupation, like doctor, seems both general, directing children to occupations that require more education and training, but also specific, having the largest impact on children's decisions to become doctors. The implications are similar when looking at teachers, a low income, high education occupation. In Figure **??** we examine how growing up next to a teacher affects a boy's probability of being in one of the five occupations that have the highest educational attainment (doctor, lawyer, clergyman, insurance agent, and teacher). Consistent with these low wage, high education jobs inducing neighbors to attain more education, we see that growing up next to a teacher increases the probability of being a doctor as much or more than being a teacher, but also significantly increases the probability of being a lawyer or insurance agent.

7 Why Neighbors Matter: Social Ties and Homophily

Given the strong transmission of occupation, income, and education from childhood neighbors, we next explore what makes neighbors matter. To do this we build on the stacked regression analysis discussed above but explore heterogeneity by neighborhood characteristics and the characteristics of the actual neighbor to understand if measures correlated with network strength and the level of interaction matter.

7.1 Neighborhood Characteristics

The interconnectedness of neighborhoods might affect the magnitude of exposure effects. We examine three measures that potentially capture dimensions of neighborhood interaction and connectedness: county-level urbanicity, immigrant share, and the out-of-state share. Neighborhoods that are less transient and have more permanent residents are likely to have stronger social ties. For this reason we explore heterogeneity by the share of the population that was not born in that state, or the out-of-state share. If network strength matters we expect to see neighbor exposure effects to be stronger in counties with fewer out-of-state residents. For urbanicity the theoretical implications are more agnostic. In urban areas, residences are likely to be geographically close together, and so individuals may have more frequent interactions with their neighbors. At the same time, because of this proximity, children may be able to interact with more distant neighbors more easily in urban areas. It is also possible that urban life is more anonymous, and so interactions are more formative in rural areas (Dunkelman, 2017; Wirth, 1938). This is similar for immigrant shares. Households may interact more in immigrant enclaves (Damm, 2009), but immigrant communities could also provide alternative networks, making next door neighbors relatively less important for the transmission of information about occupations.[24]

We present the scaled next door neighbor coefficients split by county-level urbanicity, immigrant share, and the out-of-state share in Figure **??**. We find that exposure effects are significantly larger in neighborhoods that are more rural, have lower immigrant shares, and that have more residents who were born in the state (low out-of-state share). This patterns are consistent with more homogenous places with stronger connectedness facilitating occupational transmission.[25]

7.2 Individual Characteristics

Given the heterogeneity across neighborhoods, it is plausible that there heterogeneity across characteristics of the neighbors themselves. In particular, we test whether homophily among neighbors makes exposure effects stronger. To do this, we modify equation (4) but allow the effects to vary if the focal child and neighbor have the same characteristics, as follows

$$\begin{aligned}\text{Occ1940}_{ijs} =& \alpha + \beta_0 \text{OwnOcc1910}_{ij} + \beta_1 \text{NextDoorOcc1910}_{ij} \times \text{SameCharacteristic}_{ij} \\ & + \beta_2 \text{NextDoorOcc1910}_{ij} \times \text{DifferentCharacteristic}_{ij} + \text{Occ}_j \times \text{Age}_i \\ & + \text{Occ}_j \times \gamma_s + \epsilon_{ijs}, \end{aligned} \tag{5}$$

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. For example, when considering racial homophily, a Black boy living next to a White doctor would have $\text{SameCharacteristic}_{ij} = 0$ and

[24] Living in an area with a large share of immigrants might be particularly important if the focal individual themselves comes from an immigrant family or shares an ethnicity with their neighbors. We explore the importance of the similarity between two neighbors in the next section.

[25] In Appendix Table A3, 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. In all specifications, we find positive average transmission effects, suggesting that most types of neighborhoods in this time period were facilitating occupational spillovers.

$\text{DifferentCharacteristic}_{ij} = 1$.

In Table 10 Column 2 we test whether the presence of a child with the same age in the next door household affects the magnitude of the exposure 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. Recent studies indicate that the parents of children's peers are important in determining educational outcomes in modern settings (Chung, 2020; Fruehwirth and Gagete-Miranda, 2019). 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 (0.35 percentage points versus 0.31 percentage points).

In Column 3, we estimate exposure effects for next door neighbors who are and are not from the same birthplace, using the state of birth variable in the census. Boys are more likely to enter the occupation of their next door neighbor when the neighbor is born in the same state, 0.47 percentage points more likely for neighbors from the same state versus 0.23 percentage points more likely for neighbors from different states (or countries). We present F-test statistics showing that these two coefficients are strongly significantly different from one another. We see similar patterns in Column 4 when we examine differences by birth country (with all native-born grouped together).

In Column 5, 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 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 further away, but still on the same street. Racial segregation at the time is 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. [26] In Appendix Table **??** we also show differences by household income, household education, and last name homophily.

We also explore dimensions of gender homophily. To this point we have estimated the effect of a neighbor's occupation regardless of their gender. We now re-estimate a version of equation (4), but in place of the next door indicator include two mutually exclusive indicators that equal one if the neighbor is in the target occupation and either male or female. We provide these results for our sample of boys and girls in Table **??**. For both boys and girls we find significant exposure effects from both male and female neighbors. However, the exposure effects for boys are three times as large if the neighbor is male relative to female, while the exposure effects for girls are twice as large i the neighbor is female, relative to male. Consistent with recent evidence of gender homophily in friendships, we find that exposure effects are larger when gender's match. Taken together, the heterogeneity by neighborhood and neighbor characteristics are consistent with exposure effects being larger when social ties are potentially stronger.

8 Conclusion

Where an individual spends their childhood has important implications for their future economic success (Chetty et al., 2018). Using neighborhood microgeography reconstructed from historic, door-to-door census enumeration, we show that part of this can be explained by the composition of 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 percent, relative to other boys who were living on the same

[26] In unreported analysis, we restrict to sub-samples of children of the same race, and find that non-White children have statistically zero next-door effects 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.

street. Girls similarly had positive transmission of occupations, albeit concentrated among different occupations and at slightly lower rates.

This neighbor-to-neighbor transmission of occupation varies across occupation, with larger exposure effect from neighbors in high income, high education occupations. The occupations of childhood neighbors also have long-term economic impacts on children. 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 that grew up on the same street. Part of this comes through the direct effect 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 high-education occupation. At the same time, some occupations such as truck driver and laborer actually led to reductions in income and educational attainment for neighboring children. 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 by networking with older professional women in their neighborhood.

Overall, our results suggest that childhood neighbors matter, and who you grew up with can help explain some of the effect of place on children's long-run economic mobility. We find that neighbor exposure effects are larger in neighborhoods that are plausibly more 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 ties and 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. However, these effects could also be simply driven by exposure. Children might not know that a particular occupation exists in their choice set unless they know someone in that occupation. Although social connectedness between physical neighbors might be weaker now than in the past, these patterns suggest that the composition of socially connected adults in the neighborhood and community influence children's long-term outcomes.

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. Publisher: SAGE Publications Inc.

Akbar, Prottoy A., Sijie Li, Allison Shertzer, and Randall P. Walsh, "Racial Segregation in Housing Markets and the Erosion of Black Wealth," May 2019.

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

Andrews, Michael, "Bar talk: Informal social interactions, alcohol prohibition, and invention," *Alcohol Prohibition, and Invention (November 18, 2019)*, 2019.

Andrews, Michael J. and Chelsea Lensing, "Cup of Joe and Knowledge Flow: Coffee Shops and Invention," 2024.

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. Publisher: [Oxford University Press, Review of Economic Studies, Ltd.].

Atkin, David, M. Keith Chen, and Anton Popov, "The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley," June 2022.

Avdeev, Stanislav, Nadine Ketel, Hessel Oosterbeek, and Bas van der Klaauw, "Spillovers in Fields of Study: Siblings, Cousins, and Neighbors," 2023.

Barrios-Fernández, Andrés, "Neighbors' Effects on University Enrollment," *American Economic Journal: Applied Economics*, July 2022, *14* (3), 30–60.

Bayer, Patrick, Marcus D Casey, W. Ben McCartney, John Orellana-Li, and Calvin S Zhang, "Distinguishing Causes of Neighborhood Racial Change: A Nearest Neighbor Design," Working Paper 30487, National Bureau of Economic Research September 2022.

_ , 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, 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. Publisher: Oxford University Press.

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.*

Billings, Stephen B., Mark Hoekstra, and Gabriel Pons Rotger, "The Scale and Nature of Neighborhood Effects on Children: Evidence from a Danish Social Housing Experiment," December 2022.

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," September 2023.

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

Catalini, Christian, "Microgeography and the Direction of Inventive Activity," *Management Science*, September 2018, *64* (9), 4348–4364. Publisher: INFORMS.

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. Publisher: Oxford University Press.

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

_ **, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter**, "The opportunity atlas: Mapping the childhood roots of social mobility," Technical Report, National Bureau of Economic Research 2018.

_ **, 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. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

_ **,** _ **, 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.

Cook, Lisa D., Trevon D. Logan, and John M. Parman, "Racial Segregation and Southern Lynching," *Social Science History*, 2018, *42* (4), 635–675. Publisher: Cambridge University Press.

Corak, Miles and Patrizio Piraino, "The Intergenerational Transmission of Employers," *Journal of Labor Economics*, 2011, *29* (1), 37–68. Publisher: [The University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago].

Damm, Anna Piil, "Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence," *Journal of Labor Economics*, April 2009, *27* (2), 281–314. Publisher: The University of Chicago Press.

Dunkelman, Marc J., "Next-Door Strangers: The Crisis of Urban Anonymity," *The Hedgehog Review*, 2017, *19* (2).

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. Publisher: SAGE Publications Inc.

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," Technical Report 33009, National Bureau of Economic Research September, revised November 2024. NBER Working Paper Series.

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," November 2020.

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," 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. Publisher: MIT Press.

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," July 2024.

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

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. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System].

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

_ **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 (1982)*, February 2018, *199*, 77–86.

Matta, Juan and Alexis Orellana, "Neighbors' Spillovers on High School Choice," 2022.

Mertz, Mikkel, Maddalena Ronchi, and Viola Salvestrini, "Female representation and talent allocation in entrepreneurship: The role of early exposure to entrepreneurs," *Available at SSRN 4920176*, 2024.

Moretti, Enrico, "The Effect of High-Tech Clusters on the Productivity of Top Inventors," *American Economic Review*, October 2021, *111* (10), 3328–3375.

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

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. Publisher: British Medical Journal Publishing Group Section: Research.

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. Publisher: Elsevier.

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," April 2024, (32333).

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

Staiger, Matthew, "The Intergenerational Transmission of Employers and the Earnings of Young Workers," 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. Publisher: The University of Chicago Press.

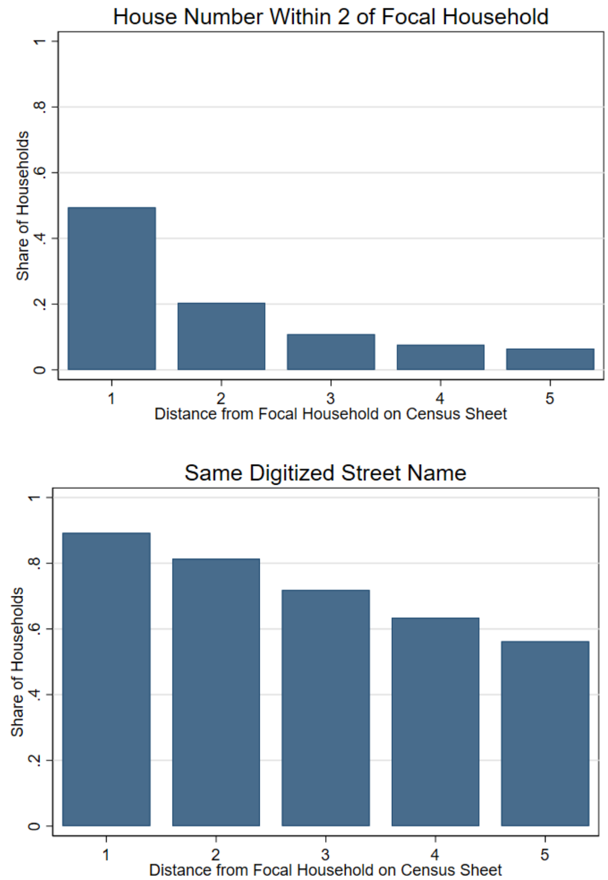
U.S. Bureau of the Census, Department of Commerce and Labor, "Thirteenth Census of the United States, April 15, 1910, Instructions to Enumerators," 1910.

Ventura, Maria, "Following in the family footsteps: Incidence and returns of occupational persistence."

Wirth, Louis, "Urbanism as a Way of Life," *American Journal of Sociology*, 1938, *44* (1), 1–24. Publisher: The University of Chicago Press.

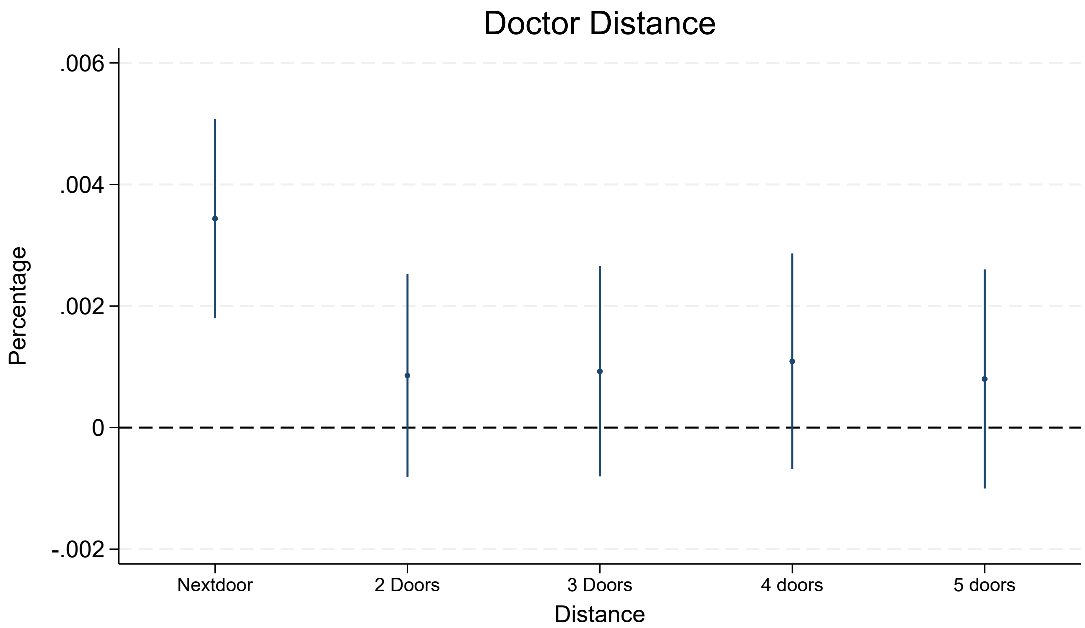
Tables and Figures

Figure 1: Sheet Proximity and Address Measures



Notes: Observation 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.

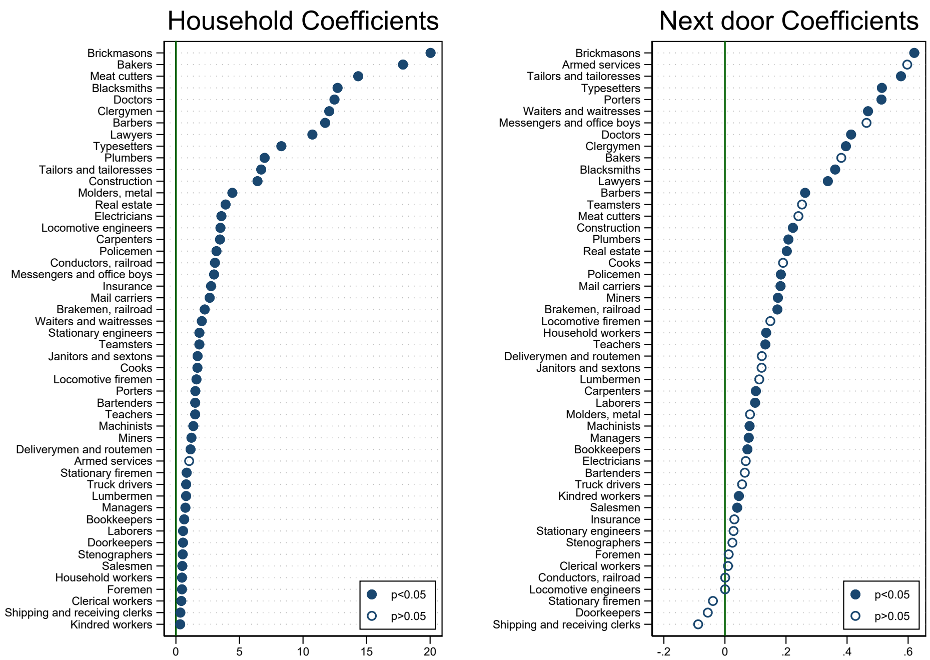
Figure 2: NeighborDoc1910 Coefficients



Notes: Sample restricted to boys between the ages of 5 and 18 in 1910 that can be linked to the 1940 census observation using the Census Tree links. The outcome is a binary measure that equals one if the boy is a doctor in 1940. This is regressed on binary measures that equal one if the boy's family lived next door, 2 doors away, three doors away, 4 doors away, or 5 doors away from a doctor in 1910, respectively, as well as a binary measure that equals one if the child had a doctor in their own household. These coefficients are plotted with 95 percent confidence intervals. The coefficient on having a doctor in their own household is not plotted. Census sheet fixed effects are also included, making this a comparison between boys who lived in the same local area that lived close to a doctor versus slightly further away. Standard errors are corrected for clustering at the household level.

Figure 3: Household and Next Door Coefficients, Individual Occupations

(a) 50 Largest Occupations for Men



(b) 25 Largest Occupations for Women

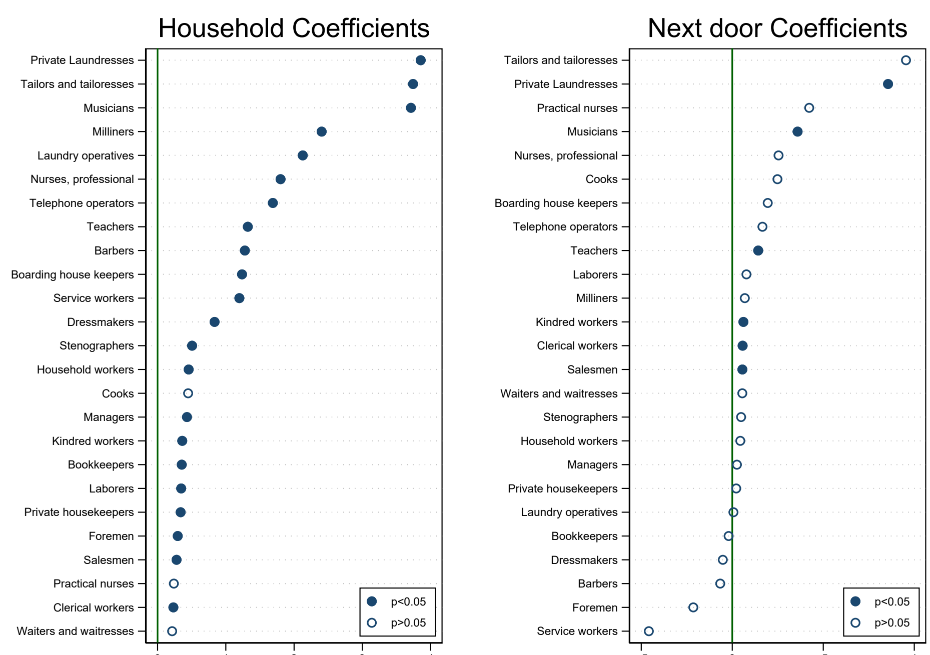
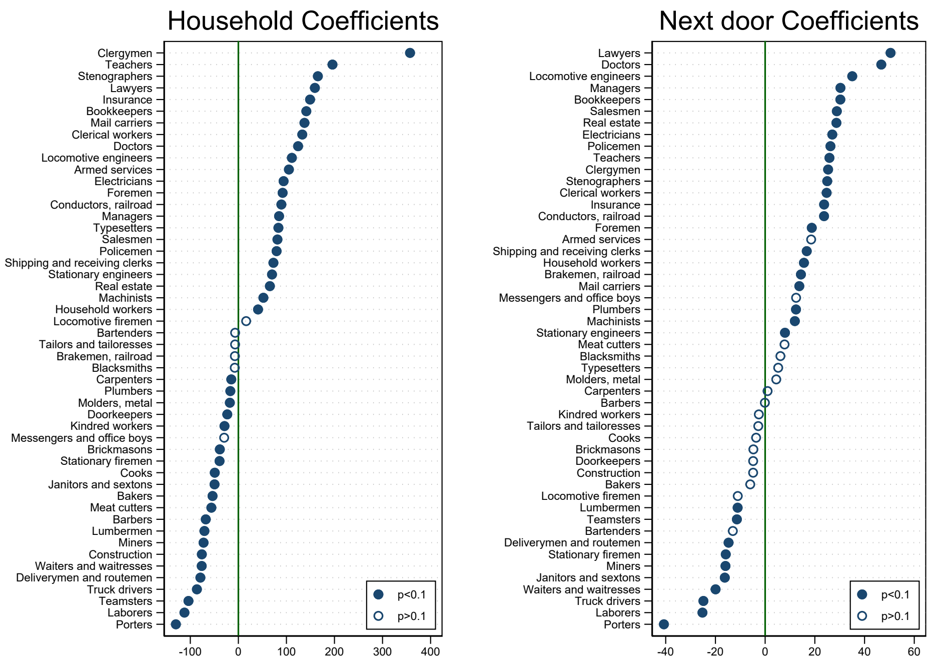


Figure 4: Household and Next Door Coefficients on Income

(a) 50 Largest Occupations for Men



(b) 25 Largest Occupations for Women

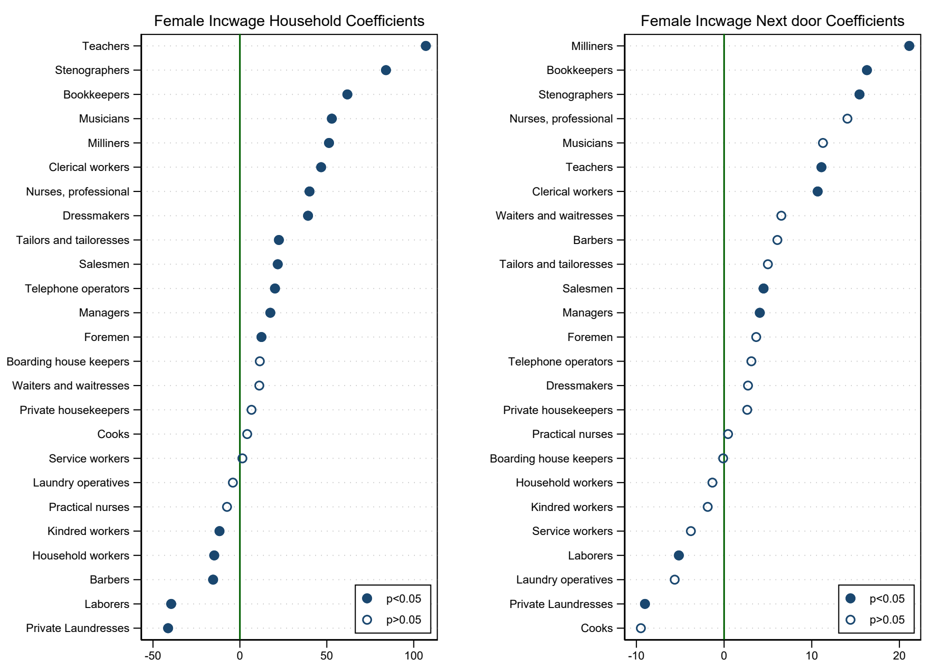
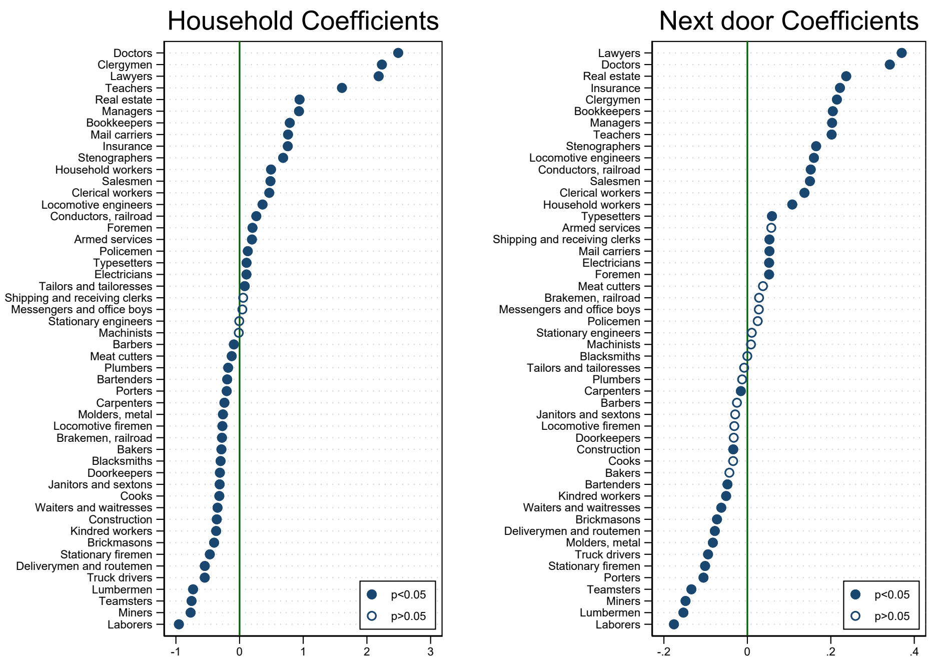


Figure 5: Household and Next Door Coefficients on Education

(a) 50 Largest Occupations for Men



(b) 25 Largest Occupations for Women

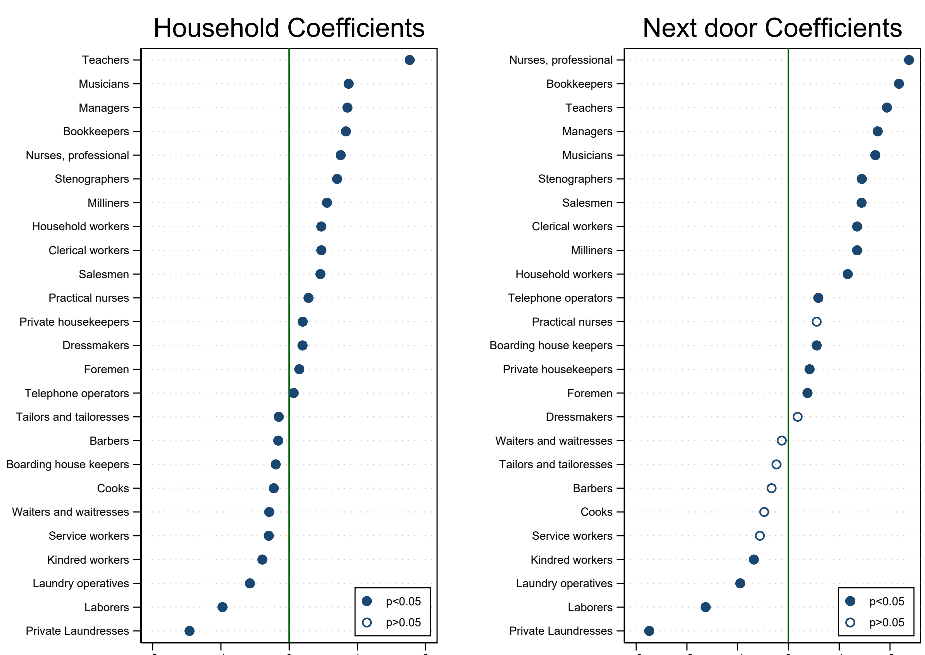


Table 1: Summary Statistics

	All 1910 children	1910 - 1940 Linked children	1910 Boys	1910 - 1940 Linked boys	1910 Girls	1910 - 1940 Linked girls
	(1)	(2)	(3)	(4)	(5)	(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
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 restricts the sample to boys only, while columns 5-6 restrict to girls only.

Table 2: 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	6,335,660	305,059	305,059	305,059

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

Table 3: Doctor Results Using Street Addresses

Dependent variable: Doctor in 1940	Sheet ordering: Baseline	Sheet ordering: Street sample	Street ordering	Street order IV
	(1)	(2)	(3)	(4)
Own household	0.0970***	0.1269***	0.1176***	0.1278***
	0.0018	0.0045	0.0040	0.0045
Next door neighbor	0.0032***	0.0036*	0.0028**	0.0087**
	0.0007	0.0019	0.0014	0.0043
R-squared	0.283	0.369	0.200	0.012
Untreated mean	0.0038	0.0061	0.0061	0.0061
Observations	6,044,153	1,053,150	1,053,150	1,053,150

Notes: Columns 1-2 use the ordering of census sheets to define neighborhood proximity. Columns 3 uses street ordering. Column 4 is an instrumental variables regression with street order variables instrumenting for sheet order variables.

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

Table 4: Doctor Results Linked to Various Censuses

Dependent variable: Doctor in…	1920	1930	1940
	(1)	(2)	(3)
Own household	0.0164***	0.0852***	0.0970***
	0.0007	0.0016	0.0018
Next door neighbor	0.0008***	0.0025***	0.0032***
	0.0003	0.0006	0.0007
R-squared	0.266	0.321	0.327
Untreated mean	0.0014	0.0065	0.0078
Observations	410,385	323,151	305,059

Notes: Columns 1-3 uses the outcome of being a doctor in 1920-1940. Sample excludes sheets with no doctors in 1910.

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

Table 5: Baseline Stacked Regression Results

(a) Men

Dependent variable: Target occupation in 1940	All sheets	At least one person in focal occupation per sheet: No geographic FE	City - enumeration district FE	Sheet FE
	(1)	(2)	(3)	(4)
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	316,783,008	26,649,908	26,649,908	26,649,908

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

(b) Women

Dependent variable: Target occupation in 1940	All sheets	At least one person in focal occupation per sheet: No geographic FE	City - enumeration district FE	Sheet FE
	(1)	(2)	(3)	(4)
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	104,389,272	11,135,893	11,135,893	11,135,893

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

Table 6: Stacked Regression Results Linked to Various Censuses - Boys

(a) All Links

Dependent variable: Target occupation in…	1920	1930	1940
	(1)	(2)	(3)
Own household	0.0365***	0.0450***	0.0381***
	0.0001	0.0002	0.0002
Next door neighbor	0.0032***	0.0037***	0.0034***
	0.0001	0.0001	0.0001
R-squared	0.318	0.324	0.323
Untreated mean	0.0262	0.0329	0.0331
Observations	37,437,872	28,531,044	26,649,908

Notes: Columns 1-3 uses the outcome of being the target occupation in 1920-1940. Sample excludes sheets with no one of the target occupation in 1910.

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

(b) Boys Found in All Censuses

Dependent variable: Target occupation in…	1920	1930	1940
	(1)	(2)	(3)
Own household	0.0375***	0.0471***	0.0410***
	0.0002	0.0002	0.0002
Next door neighbor	0.0032***	0.0036***	0.0034***
	0.0002	0.0002	0.0002
R-squared	0.405	0.385	0.377
Untreated mean	0.0257	0.0327	0.0330
Observations	15,473,431	15,473,431	15,473,431

Notes: Columns 1-3 uses the outcome of being the target occupation in 1920-1940. Sample excludes sheets with no one of the target occupation in 1910. Observations are children in 1910 that are linked to 1920, 1930, and 1940 censuses.
**p* <0.1, ***p* <0.05, ****p* <0.01.

Table 7: Stacked Regression Results Linked to Various Censuses - Girls

(a) All Links

Dependent variable: Target occupation in…	1920	1930	1940
	(1)	(2)	(3)
Own household	0.0168***	0.0107***	0.0085***
	0.0001	0.0002	0.0002
Next door neighbor	0.0016***	0.0011***	0.0010***
	0.0001	0.0001	0.0001
R-squared	0.338	0.351	0.356
Untreated mean	0.0255	0.0180	0.0157
Observations	23,530,340	13,458,635	11,165,884

Notes: Columns 1-3 uses the outcome of being the target occupation in 1920-1940. Sample excludes sheets with no one of the target occupation in 1910.

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

(b) Girls Found in All Censuses

Dependent variable: Target occupation in…	1920	1930	1940
	(1)	(2)	(3)
Own household	0.0140***	0.0090***	0.0080***
	0.0002	0.0002	0.0002
Next door neighbor	0.0011***	0.0011***	0.0009***
	0.0002	0.0002	0.0002
R-squared	0.426	0.395	0.393
Untreated mean	0.0200	0.0146	0.0149
Observations	6,885,595	6,885,595	6,885,595

Notes: Columns 1-3 uses the outcome of being the target occupation in 1920-1940. Sample excludes sheets with no one of the target occupation in 1910.

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

Table 8: Heterogeneity by Occupation Characteristics

Dependent variable: Target occupation in 1940	Baseline	High income	High education	High self-employment
	(1)	(2)	(3)	(4)
Own household	0.0381***	0.0537***	0.0439***	0.0571***
	0.0002	0.0003	0.0003	0.0003
Next door neighbor	0.0034***	0.0040***	0.0037***	0.0036***
	0.0001	0.0002	0.0002	0.0002
R-squared	0.323	0.337	0.323	0.339
Untreated mean	0.0331	0.0339	0.0421	0.0276
Observations	26,649,908	8,481,845	10,662,704	10,839,995

Notes: Column 1 includes all occupations. Columns 2-4 include occupations with income, education, and self-employment score above median respectively.

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

Table 9: Heterogeneity by Neighborhood Characteristics

Dependent variable: Target occupation in 1940	Baseline	Urban	Rural	High immigrant share	Low immigrant share	High out-of-state share	Low out-of-state share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Own household	0.0381***	0.0360***	0.0415***	0.0366***	0.0437***	0.0356***	0.0434***
	0.0002	0.0002	0.0003	0.0002	0.0004	0.0002	0.0003
Next door neighbor	0.0034***	0.0028***	0.0042***	0.0029***	0.0048***	0.0028***	0.0046***
	0.0001	0.0002	0.0002	0.0001	0.0003	0.0001	0.0002
R-squared	0.323	0.331	0.311	0.322	0.328	0.324	0.322
Untreated mean	0.0331	0.0317	0.0356	0.0322	0.0370	0.0319	0.0360
Observations	26,649,908	16,639,195	10,009,934	21,176,906	5,472,864	18,313,568	8,336,069

Notes: Each column reports results within different subsets of the census sample based on neighborhood characteristics.

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

Table 10: Heterogeneity by Neighbor Similarity

(a)

Dependent variable: Target occupation in 1940	Baseline	Birthplace	Birth country	Race	Last name	Same age	Income score	Education score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Own household	0.0381***	0.0381***	0.0381***	0.0381***	0.0381***	0.0382***	0.0376***	0.0375***
	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
Next door neighbor	0.0034***							
	0.0001							
Treatment 1: Same characteristic		0.0047***	0.0040***	0.0038***	0.0220***	0.0035***	0.0059***	0.0076***
		0.0002	0.0002	0.0001	0.0010	0.0003	0.0002	0.0002
Treatment 1 average		*0.0984*	*0.1595*	*0.2284*	*0.0041*	*0.0311*	*0.1300*	*0.1334*
Treatment 2: Different characteristic		0.0023***	0.0019***	-0.0083***	0.0030***	0.0031***	-0.0012***	-0.0038***
		0.0002	0.0002	0.0006	0.0001	0.0001	0.0002	0.0002
Treatment 2 average		*0.1453*	*0.0829*	*0.0086*	*0.2328*	*0.1639*	*0.0848*	*0.0814*
F-test		100.28	66.90	404.93	325.74	1.60	849.42	2223.73
Two-tailed p-value		0.0000	0.0000	0.0000	0.0000	0.2058	0.0000	0.0000
R-squared	0.323	0.323	0.323	0.323	0.323	0.323	0.323	0.323
Untreated mean	0.0331	0.0331	0.0331	0.0331	0.0331	0.0338	0.0338	0.0338
Observations	26,649,908	26,649,908	26,649,908	26,649,908	26,649,128	26,649,908	26,649,908	26,649,908

Notes: In columns 2-5, treatments compare head of household of the focal child to the target occupation holder next door. In column 6, same age is defined as any child in the target occupation house being within one year of the focal child. In columns 7-8 the treatments evaluate if both head of household of focal kid and target occupation next door are above/below median in occscore and edscore. The number of observations in column 5 is slightly less due to the fact that individuals whose last name is missing and whose neigbhors' last names are also missing are dropped from the regression.

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

(b)

Dependent variable: Target occupation in 1940	Baseline	White	Non-white	High income	Low income	High education	Low education
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Own household	0.0381***	0.0373***	0.0198***	0.0437***	0.0323***	0.0395***	0.0325***
	0.0002	0.0002	0.0015	0.0004	0.0003	0.0005	0.0002
Next door neighbor	0.0034***						
	0.0001						
Treatment 1: Same characteristic		0.0029***	0.0008	0.0018***	0.0032***	0.0018***	0.0028***
		0.0001	0.0013	0.0005	0.0002	0.0005	0.0002
Treatment 1 average		*0.2298*	*0.2166*	*0.0855*	*0.1806*	*0.1202*	*0.1671*
Treatment 2: Different characteristic		0.0029***	0.0010	0.0031***	0.0028***	0.0029***	0.0032***
		0.0006	0.0023	0.0003	0.0003	0.0003	0.0003
Treatment 2 average		*0.0064*	*0.0483*	*0.1494*	*0.0573*	*0.1155*	*0.0708*
F-test		0.00	0.01	4.44	1.12	3.90	1.64
Two-tailed p-value		0.9948	0.9277	0.0352	0.2902	0.0483	0.2006
R-squared	0.323	0.321	0.484	0.428	0.367	0.432	0.358
Untreated mean	0.0331	0.0328	0.0461	0.0303	0.0345	0.0303	0.0343
Observations	26,649,908	25,682,014	661,760	6,935,351	14,073,164	6,427,565	14,802,656

Notes: In columns 2-7, treatments compare head of household of the focal child to the target occupation holder next door.

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

Table 11: Effects of Doctors on Entering the Top 5 Occupations by Income

Dependent variable: Target occupation in 1940	Doctors	Lawyers	Managers, Officials, or Proprietors	Foremen	Compositors and Typesetters	Any top five occupation
	(1)	(2)	(3)	(4)	(5)	(6)
Own household	0.0970***	0.0105***	-0.0021	-0.0061***	-0.0022***	0.0970***
	0.0018	0.0011	0.0027	0.0009	0.0005	0.0032
Next door neighbor	0.0032***	0.0021***	0.0118***	-0.0008	0.0006	0.0169***
	0.0007	0.0007	0.0021	0.0008	0.0004	0.0024
R-squared	0.327	0.310	0.309	0.266	0.286	0.315
Untreated mean	0.0078	0.0132	0.1454	0.0201	0.0058	0.1922
Observations	305,059	305,059	305,059	305,059	305,059	305,059

Notes: Estimates for the effect of having a doctor living next door in 1910 on being in the target occupation in 1940.
*p<0.1, **p<0.05, ***p<0.01.

A Additional Tables and Figures (Online Appendix)

Figure A1: Household and Next Door Coefficients for 50 Largest Occupations for Men, Without Scaling by Mean of Untreated Kids

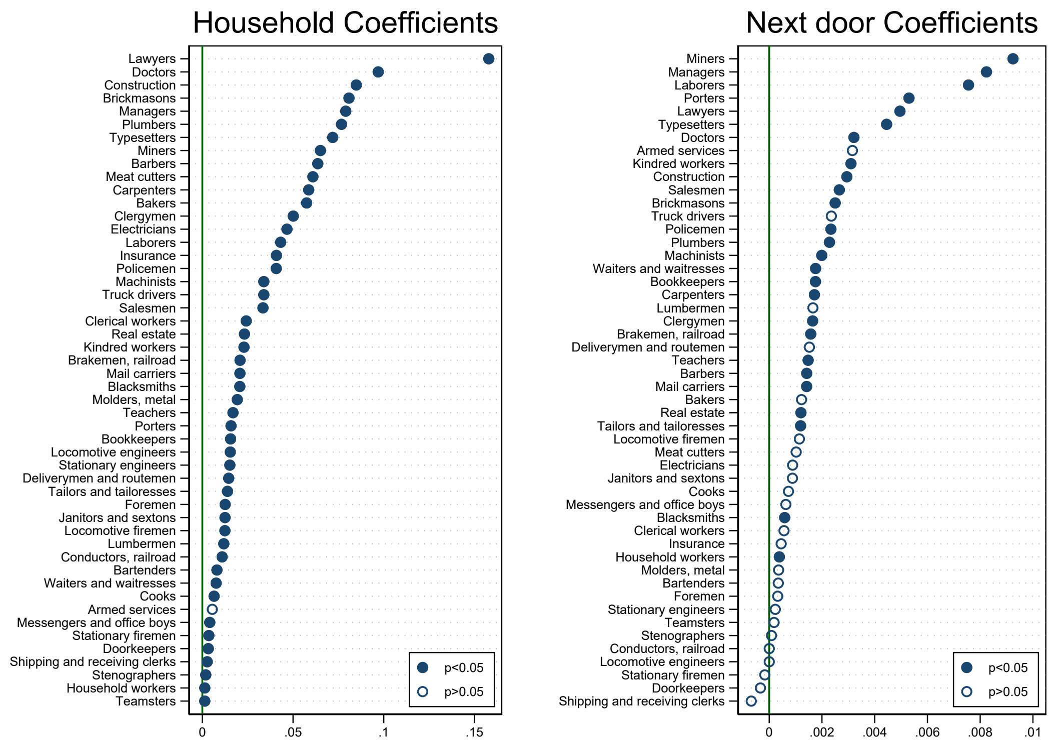


Table A1: Stacked Regression Results while Handling Farmers in Different Ways

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***	0.0042***	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	26,649,908	33,117,306	29,644,924	30,122,288	3,472,382	2,995,017	6,467,399

Notes: Column 1 excludes farm owners and laborers, column 3 excludes farm owners, column 4 excludes laborers, columns 5-6 uses only farm owner and/or farm laborers.
*p<0.1, **p<0.05, ***p<0.01.

Table A2: Heterogeneity by Income and Education

	High income		Low income	
Dependent variable: Target occupation in 1940	High education	Low education	High education	Low education
	(1)	(2)	(3)	(4)
Own household	0.0609***	0.0396***	0.0242***	0.0334***
	0.0004	0.0004	0.0004	0.0002
Next door neighbor	0.0053***	0.0014***	0.0017***	0.0036***
	0.0003	0.0002	0.0003	0.0002
R-squared	0.334	0.290	0.301	0.322
Untreated mean	0.0485	0.0103	0.0358	0.0314
Observations	5,356,209	3,125,636	5,306,495	12,861,567

Notes: Each column reports results within four groups of occupations based on whether the occupation is above or below the median income and education score.
*p<0.1, **p<0.05, ***p<0.01.

Table A3: 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	11,676,940	14,972,663	13,604,749	13,044,706	9,198,224	9,743,871	5,464,572	2,243,240

Notes: Each column reports results within different subsets of the census sample based on neighborhood or characteristics or region.

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

Table A4: Heterogeneity by Neighbor Similarity for Last Names that Are Unique to Each Sheet

Dependent variable: Target occupation in 1940	Baseline	Birthplace	Birth country	Race	Last name	Same age	Income score	Education score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Own household	0.0374***	0.0374***	0.0374***	0.0374***	0.0374***	0.0375***	0.0369***	0.0368***
	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
Next door neighbor	0.0030***							
	0.0001							
Treatment 1: Same characteristic		0.0042***	0.0036***	0.0034***	0.0187***	0.0033***	0.0056***	0.0072***
		0.0002	0.0002	0.0001	0.0048	0.0004	0.0002	0.0002
Treatment 1 average		*0.0908*	*0.1541*	*0.2335*	*0.0002*	*0.0320*	*0.1274*	*0.1335*
Treatment 2: Different characteristic		0.0023***	0.0018***	-0.0074***	0.0030***	0.0027***	-0.0013***	-0.0039***
		0.0002	0.0002	0.0007	0.0001	0.0002	0.0002	0.0002
Treatment 2 average		*0.1580*	*0.0937*	*0.0084*	*0.2410*	*0.1661*	*0.0906*	*0.0845*
F-test		43.01	42.79	230.11	10.66	1.99	571.95	1515.42
Two-tailed p-value		0.0000	0.0000	0.0000	0.0011	0.1583	0.0000	0.0000
R-squared	0.3260	0.3260	0.3260	0.3260	0.3260	0.3260	0.3260	0.3261
Untreated mean	0.0323	0.0323	0.0323	0.0323	0.0323	0.0329	0.0330	0.0330
Observations	18,558,482	18,558,482	18,558,482	18,558,482	18,558,404	18,558,482	18,558,482	18,558,482

Notes: This table restricts to individuals with unique last names. In columns 2-5, treatments compare head of household of the focal child to the target occupation holder next door. In column 6, same age is defined as any child in the target occupation house being within one year of the focal child. In columns 7-8 the treatments evaluate if both head of household of focal kid and target occupation next door are above/below median in occscore and edscore. The number of observations in column 5 is slightly less due to the fact that individuals whose last name is missing and whose neighbors' last names are also missing are dropped from the regression.

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