Transactions in GIS

Research Article

Transactions in GIS, 2014, 18(1): 25–45

The Co-evolution of Residential Segregation and the Built Environment at the Turn of the 20th Century: A Schelling Model

Seth Spielman* and Patrick Harrison†

**Department of Geography, University of Colorado*
†Information and Systems Engineering, University of Virgina

Abstract

To what degree does the built environment of cities shape the social environment? In this article we use a Schelling-like agent-based model to consider how changes to the built environment of cities relate to changes in residential segregation by income and ethnicity. To develop this model we exploit insights from a high resolution historical GIS which maps 100% of the population of Newark, NJ in 1880. Newark in 1880 had a complex social landscape characterized by areas of significant social and economic segregation and areas of relative integration. We develop a Schelling model capable of reproducing these residential patterns. We use this model to explore the decentralization of housing, a specific phenomenon associated with the demise of the walking city in the late 19th century. Holding agent preferences constant, but allowing the landscape of the Schelling model to evolve in ways that reflect historical changes to the built environment, produces changes to the social landscape that are also consistent with history. Our work suggests that changes in residential segregation do not necessarily imply changes to individual attitudes and preferences. Changes in residential segregation can be generated by changes to the built environment, specifically the geographic distribution of housing.

1 Introduction

Segregation – the residential separation of people based on race, ethnicity, and/or income – is an enduring characteristic of modern urban life. Dawkins (2004) offers five explanations for why residential segregation occurs, ranging from discrimination in housing markets and other forms of prejudice to residential preferences that differ by race. Charles (2001) notes that racial groups may reside in different areas because of objective differences in economic status, finding that, on average, whites spend more on housing than blacks. Clark and Ware (1997) report decreasing segregation as income increases. Fischer (2003, p. 685) finds that while overall racial segregation declined from 1970–2000 the relative importance of income/class to residential segregation increased. Segregation is a dynamic phenomenon; decade by decade, observed levels of segregation can change substantially (Logan et al. 2004). Segregation is also a complex phenomenon, shaped by economic factors, individual preferences, racial/ethnic discrimination, and institutions.

Segregation is fundamentally a spatial phenomenon having to do with the arrangement of people in space. However, the arrangement of people in space is constrained by the configuration of the built environment – particularly the location of housing. Changes to these constraints may affect the degree of residential segregation. Watson (2006) finds that the rate of

Address for correspondence: Seth Spielman, University of Colorado, Department of Geography, Guggenheim 110, Boulder, CO 80309-0260, USA. E-mail: seth.spielman@colorado.edu

 doi: 10.1111/tgis.12014

growth in housing markets affects both income and racial segregation because in faster growing metro areas housing markets are better able to adapt to evolving residential preferences. Watson's findings are interesting because they suggest that the built environment of cities may affect residential segregation.

In this article we consider how changes to the built environment of cities relate to changes in observed levels of residential segregation. We model the decentralization of housing and the demise of the walking city, a specific structural change in cities that began in the late 19th century and continued throughout the 20th century. Decentralization of housing is the decline of residential density in the city center and the rise of residential neighborhoods at increasing distances from the center (Bruegmann 2005, Hall 1996). We exploit insights from a high resolution historical GIS that maps 100% of the population of Newark, NJ in 1880 (Logan et al. 2010, Spielman and Logan 2012). We use these data to train a Schelling-like agent-based model that reproduces the residential patterns in 1880 Newark. We then use the trained model to explore how changes to the configuration of the built environment, mainly the decentralization of housing, influence ethnic and economic residential segregation.

2 Background and Literature Review

Space and the configuration of people are central themes in the literature on residential segregation (White 1983, Iceland et al. 2002). Methods for measuring segregation have been developed that are sensitive not just to the relative mix of people in an area, but also to their spatial arrangement (Wong 2005, Lee et al. 2008). These statistics quantify the arrangement and relative mixture of racial and ethnic groups in defined geographic areas and provide essential tools for describing observed residential patterns. These statistics are useful for descriptive purposes but they do not, in and of themselves, provide insight into the causes and consequences of segregation.

On the other hand, dynamic models of residential patterns inspired by Schelling (1969, 1971, 1978) explore the role of various social processes in the formation of segregation (Fossett 2006, Crooks 2010). A simple Shelling model (SSM) contains two types of agents, typically labeled with colors (such as black and white). The "landscape" of an SSM is a lattice of equally-sized squares – similar to a checkerboard. Each square or "cell" contains one agent and neighbors eight other squares, which also contain one agent each. In an SSM an agent (often a proxy for a household) has a preference for living in a cell where a minimum percentage of its neighboring cells are occupied by agents of the same color. Agents in cells without the minimum percentage of similar neighbors are "unsatisfied," and move to a new cell. Allowing agents to iteratively apply this process of observation and movement results in near complete segregation between the two agent types, even when agents have a significant tolerance for the other group (Zhang 2004). Schelling, in his original formulation of the model, adjusted agents' preferences for living near members of the other group. He found, as have many others, that even slight preferences for living near members of the same group at the individual level produce extreme residential separation at the group level. The impact of racial/ethnic preferences on residential outcomes seems clear and recent work has provided substantial insight into the evolution of preferences over time, however the origins and causes of racial preferences are enormously complex and difficult to explain (Clark and Fossett 2008).

The simple Schelling model has been dramatically extended to include real-estate markets, housing quality, realistic racial/ethnic preference schedules, income, status, and more than two racial/ethnic groups (Zhang 2004, Fossett 2006, Bruch and Mare 2006). For example, Zhang

2004 modifies a simple Schelling model to allow for a housing market, including a vacancy rate. Zhang's model not only reproduces the phenomenon of "white flight" from inner city neighborhoods but also mirrors changes in housing price differentials between white and black neighborhoods. Fossett's (2006) SIMSEG program has been used to simulate geographic variations in housing quality and affordability. Bruch and Mare (2006) and Clark (1991) have used survey data in Schelling-like models to gain insight into the drivers of segregation in real populations. As Crooks (2010, p. 663) notes, it is useful to think about Schelling-like models "on a continuum between abstract demonstrations to real-world applications". Clark and Fossett (2008, p. 4109) note that even abstract models can reproduce the "complex and sometimes subtle segregation patterns" seen in real cities, thus illustrating "how choices play out in the social fabric and lead to segregated residential outcomes". We do not believe that these types of models can explain the causes of segregation in the real world, but they can illustrate how micro-level behavior can produce (sometimes unexpected) macro-level patterns.

The patterns of segregation produced by Schelling models are part of their appeal. Varying degrees of segregation by race, ethnicity, and income have been the norm in industrial and post-industrial cities in the U.S. However, this was not always the case. Early in the industrial revolution, before the advent of mass transportation, cities were very dense and more diverse. Charles Booth's maps of poverty in London are a striking example – one notices that there are often very poor streets mixed in among better-off neighborhoods (Dyos et al. 1982). Greenberg (1981), using data from the Philadelphia Social History Project, argues that residential segregation in Philadelphia in the late 19th century was shaped by access to employment. Greenberg (1981) claims that structural constraints, such as the need to walk to work, played a larger role in shaping the urban residential landscape than cultural constraints enforcing homophiliy. The ethnic composition of industries and the spatial distribution of employment within the city played a larger role than preference, according to Greenberg. Areas with industries that tended to be dominated by a single ethnic group appeared to be ethnically segregated. However, areas near industries in which employees were ethnically diverse, and areas where multiple types of employment were concentrated, tended to be diverse. An important exception was for blacks – even though they were a relatively small portion of the population, they appear to be spatially isolated from others in ways that cannot be explained by employment. The basic Schelling model is ill suited to capture the complex residential landscape of the late 19th century.

Spielman and Logan (2012) study four cities in 1880 (Newark, NJ; Buffalo, NY; Cincinnati, OH; Albany, NY) using address-level maps of the entire population from the Urban Transitions Historical GIS and find that found that the central (core) areas of these cities tends to be more ethnically diverse, but wealthier and more economically homogeneous than other parts of the city. Further from the city centers, neighborhoods tend to be increasingly mono-ethnic. However, when looking at fully mapped household level census enumeration records, Spielman and Logan find a complex relationship between ethnicity and income. Their results show evidence of ethnic mixing – but they find that residential separation by ethnicity in some areas co-existed with ethnically integrated neighborhoods in others (Figure 1).

The late 19th century was a period of great change for many cities. Zunz (1982) argues that in Detroit, ethnic and economic segregation dramatically increased between 1880 and 1920. The patterns observed by Spielman and Logan (2012) may only represent one phase in that transition. This transition was closely, if not causally, linked to advances in mass transportation (Hall 1996, Warner 1962). In the 19th century urban transportation options were limited, and those that existed were expensive. In 1850, a round trip transit ticket in Philadelphia cost about 26% of the daily wage for an unskilled worker. By 1880, the same trip cost

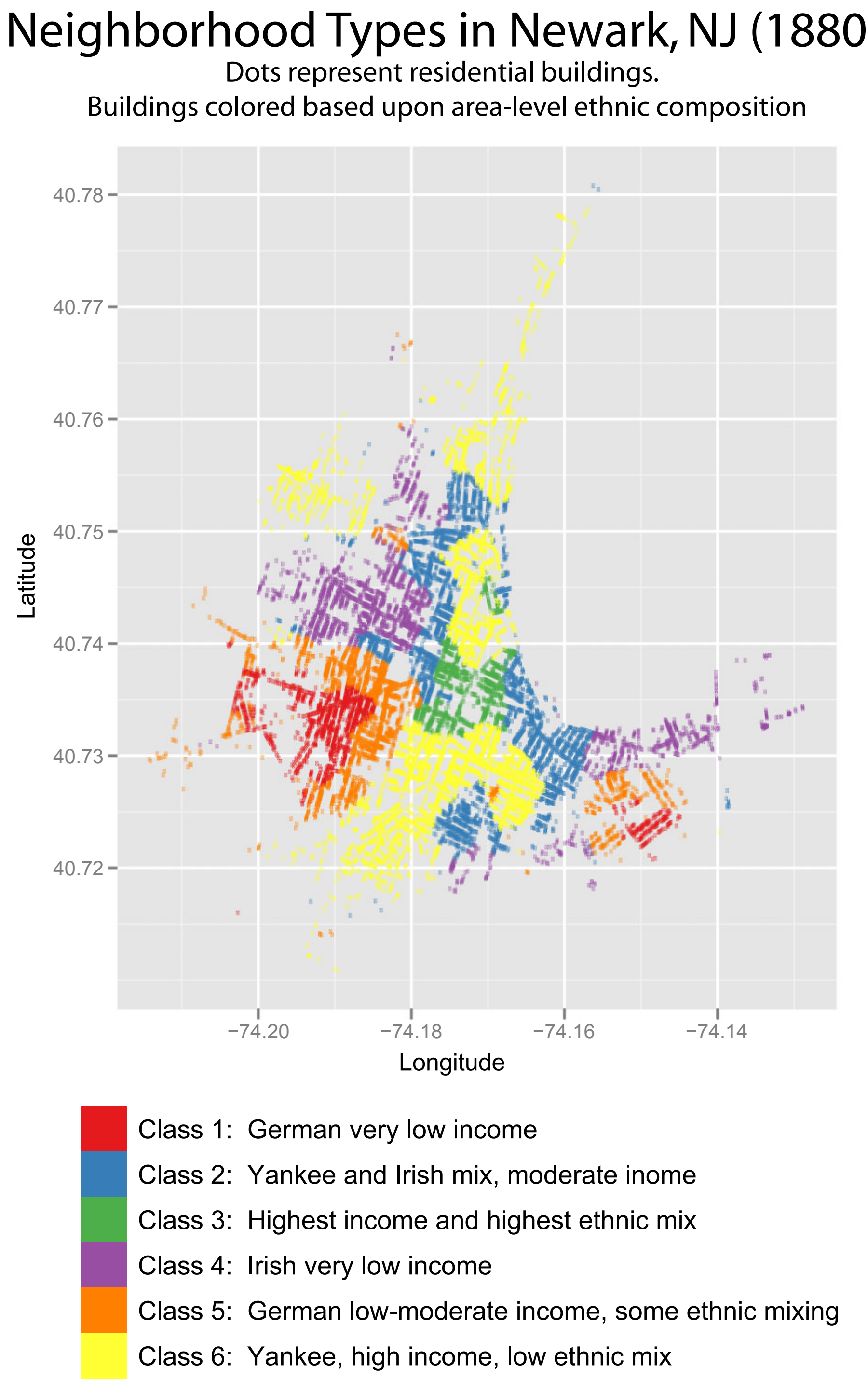

Figure 1 Residential neighborhood types in Newark, NJ (1880)

14% of daily wages. However, within 10 years of 1880, mass transportation systems expanded greatly. Hershberg (1981) finds that per capita mass transit ridership in Philadelphia increased around 70% every decade from 1870 to 1890. After 1890 the increases accelerated – by 1910 the average Philadelphian made 200 trips on mass transit per year. The relative cost of urban transportation has continued to fall steadily: by 1980 a person earning the federally mandated minimum wage could buy a round trip ticket on mass transit in Philadelphia for just 4.4% of their daily earnings (Hershberg 1981, p. 153).

The changes in Philadelphia were characteristic of other cities at the time. Due to the advent of affordable mass transportation, people no longer had to live within walking distance of work, and the hyper-dense urban cores of the late 19th century became the "two-part city" described by Warner (1962). A two-part city is characterized by separate spaces for living and working; the separation of home and work implied by the concept of a two-part city requires widespread commuting. The rise of commuting corresponded with a dramatic shift in the physical structure of cities. Bruegmann (2005, p. 19) plots the population density of London as a function of distance from 1800 to 1950. These plots show a "flattening out" of the population distribution within the city; the central city became less dense (residentially) and the city became more expansive geographically. Bruegmann (2005) shows that in 1840, once one was eight miles from the center of London population densities were essentially rural, while by 1920 those same regions had nearly 15,000 people per square mile.

This shift in the built environment within cities may also have important implications for the development of ethnic and income segregation. We model this transition from a dense compact city to a more expansive, lower density form within the framework of a Schelling-like model. The dynamic framework of a Schelling model lends itself well to questions about the evolution of cities, but until recently data limitations have hampered historical inquiry. We use recent advances in Historical GIS to provide an empirical foundation for a Schelling-like agent-based model.

3 Research Questions

Explaining the co-existence of ethnically segregated and integrated neighborhoods is a challenge for Schelling models (Clark 1991). It has been found that Schelling models have a tendency to produce either segregated or integrated outcomes, but not outcomes where only parts of a city are segregated (Laurie and Jaggi 2003). In this article we seek first to develop an extended, Schelling-like model that is capable of reproducing the type of patterns observed in the late 19th century – diversity in the urban core and segregation in peripheral neighborhoods (Figure 1). We then use the model to explore how the physical structure of cities can affect levels of segregation along ethnic and economic lines. Specifically, we seek to examine how the shift from a compact, dense urban structure characteristic of cities in the late 19th century to the more expansive, lower-density structure of cities in the 20th century affects the development of the social environment. To our knowledge, these are both novel extensions of the Schelling framework. We believe these questions are both of historical interest and contemporary relevance, as some American cities in the past decade have reversed the trend and experienced significant re-densification through infill development in the urban core.

4 Data

The data used to create our agents and validate our model were compiled by the North Atlantic Population Project (NAPP) at the Minnesota Population Center and mapped by the Urban

Transition Historical GIS Project (UTP) at Brown University (Logan et al. 2011, Minnesota Population Center 2008). Historical Geographic Information Systems like the UTP are becoming increasingly common. Gregory and Healy (2007) note that using GIS in historical inquiry allows one to ask questions about pattern and distribution in ways that non-GIS archival sources do not. Despite the potential for historical GIS to provide novel insights into the geographic organization of society and its evolution, Gregory and Healy (2007) also observe that analytical studies of historical data using GIS are rare. While we generally agree with this observation, there are some notable exceptions (such as Dorling et al. 2000, Hershberg 1981, and Gregory 2008). To our knowledge, this is the first use of a historical GIS in the context of a Schelling-like modeling effort.

The NAPP/UTP use the 100% digital transcription of the handwritten records compiled by the individual enumerators who conducted the 1880 decennial census. These digital transcriptions were organized by the Church of Latter Day Saints and are made available by the Minnesota Population Center. For 39 major cities, UTP has augmented the NAPP data through the addition of street addresses for all residents, and has mapped those addresses using historical sources. This study is based on residential patterns observed in a subset of four cities described by Spielman and Logan (2012): Albany, NY, Buffalo, NY, Cincinnati, OH, and Newark, NJ. All four cities were among the top 25 largest cities in the U.S. However, we train our model using the observed ethnic segregation patterns in Newark, NJ. Newark's selection as the focal city is largely arbitrary – each of the cities exhibited similar types of settlement patterns (Spielman and Logan 2012).

These four cities had similar overall ethnic composition. The primary population groups were German immigrants, Irish immigrants, and native-born whites (Yankees); no other single group represented more than 10% of the population. Germans and Irish are defined here as persons who were born in Germany or Ireland, or had at least one parent who was born in Germany or Ireland. Yankees are whites born in the U.S. with U.S.-born parents. Together, these groups comprised about 80% of the population in each city.

The other variable that we draw on is socioeconomic status, as determined by the NAPP using the occupation of the highest status member in a household. Occupations were ranked on a scale from 0–100 based on the average education and earnings of persons in each occupation as measured in 1950. The resulting scale, the Duncan Socio-Economic Index (SEI), is commonly used as a measure of socioeconomic status. Sobek (1996) compared the average income of men in each of 140 occupations in 1890 with the income of men in those same occupations in 1950. Sobek finds a correlation between the two of 0.93, concluding that the SEI scale is valid measure of socioeconomic status for the late 19th century.

Using these definitions, in 1880 there were 8,783 German households in Newark, NJ, 6,247 Irish households, and 7,517 Yankee households ($n = 22{,}547$). The highest SEI in our dataset was 96 and the lowest was zero. On average Irish had the lowest SEI (25.3), followed by Germans (31.1), and Yankees (39.1); analysis of variance shows the observed differences in income between ethnic groups to be statistically significant. Figure 2 shows the distribution of income (SEI) within each ethnic group. Using these data, we find that residential segregation by ethnicity was moderate globally, but extremely high in some regions of the city. At the street block level, the Wong (2005) spatial dissimilarity index is 0.53 for Irish and German, 0.39 for Irish and Yankee, and 0.54 for German and Yankee. Non-stationary patterns such as the ones we observe (in Figure 1) can be difficult to quantify. The segregation statistics reported above are somewhat misleading, because they represent an overall metropolitan average and ignore significant local variation.

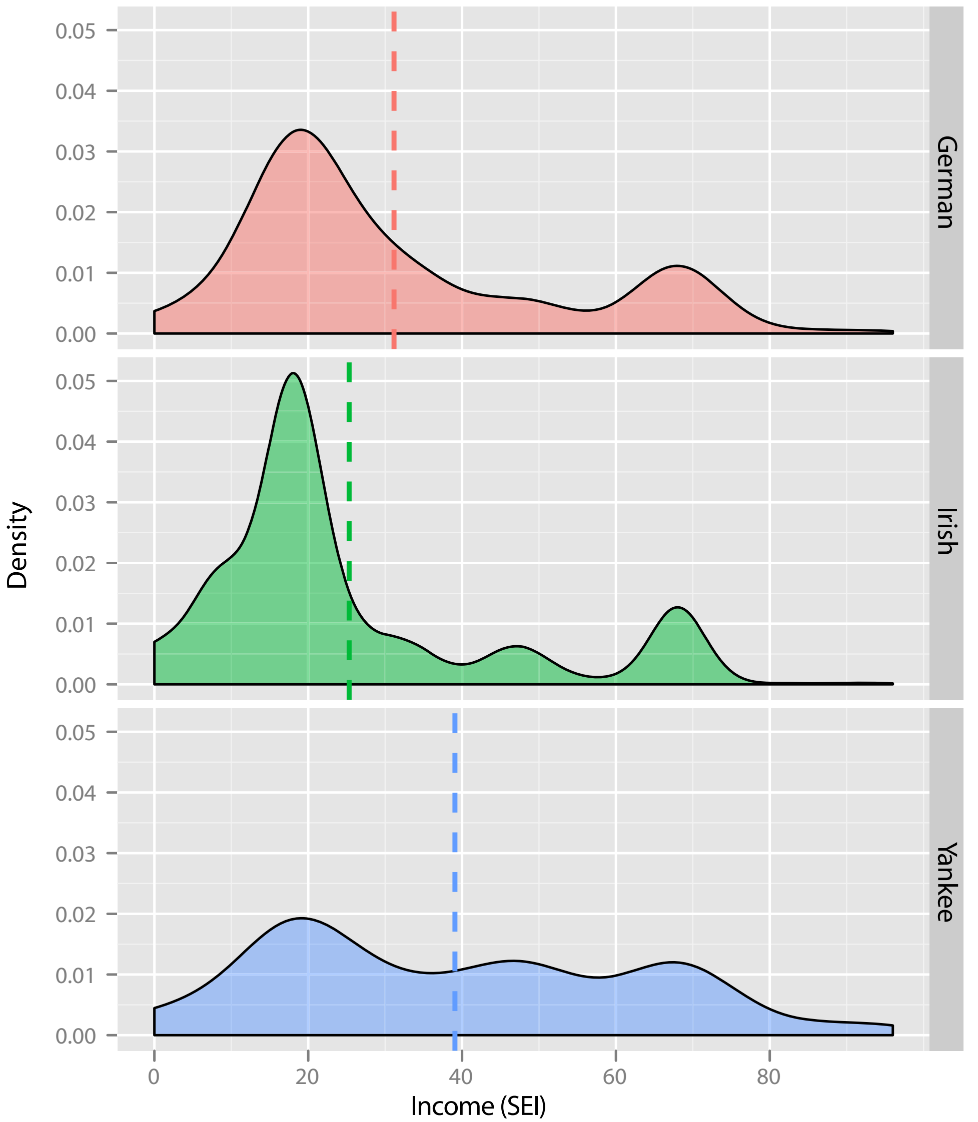

Figure 2 Income by ethnic group, Newark, NJ 1880. Mean denoted by dashed line

5 Methods

Our model, developed in NetLogo 4.1.3, is an extension of the simple Schelling model (Willensky 1999). We have made three major modifications to the simple Schelling model:

1. We generate a population of agents that mirrors the population of households in Newark, NJ, in 1880. Agents belong to one of three ethnic groups (German, Irish, or Yankee). The agents are assigned an income (SEI) based upon the observed distribution of incomes in 1880.

2. We parameterize the landscape upon which agents interact. This parameterization distributes housing units across the landscape to create housing density profiles like those observed in late 19th and early 20th centuries. While we have experimented with a number of different shapes of density profiles, here we focus our attention on the simplest case, in which we distribute housing units using various Gaussian distributions.
3. We develop a utility function for the agents that allows us to reproduce the observed geographic relationships between ethnic composition and incomes.

Our work falls somewhere in the middle of the continuum between pedagogic and realistic Schelling models. The characteristics of our agents are empirically derived, but the mechanisms they use to select residential locations are too simple and stochastic to reflect reality. Fossett (2006) and Zhang (2004) have developed sophisticated residential choice algorithms for Schelling models that account for housing quality, price, and demand. Our model was reliably able to reproduce the patterns similar to those we observed in 19th century Newark while considering only the ethnic and economic preferences of the agents. The model emphasizes parsimony, rather than trying to capture all variables that may have an effect on observed residential patterns (Table 1). The only variables included in the model are directly measured by the 1880 census. The model does not include a real estate markets or complex housing search procedures, as such mechanisms are beyond the scope of the current investigation.

We used this model to conduct two distinct experiments. The first is designed to examine how the model responded to variations in parameters, and to determine combinations of parameters that produce residential patterns in the model similar to those observed in the empirical 19th century data. The second experiment varies the landscape on which the agents interact and is designed to mirror changes to urban environments that occurred in the early 20th century. It seeks to understand if the evolution of the built environment, holding preferences for ethnicity and income constant, results in changes to observed patterns of segregation. This second experiment is a novel extension to the Schelling framework that explores the role of physical environmental constraints on segregation. Previous work by Fossett (2006) has explored the role of spatial variations in housing quality and price on segregation, but not the role of urban form or housing unit density.

5.1 Creating Agents

In generating agents for our model, we seek to preserve both the relative proportion of each ethnic group in the population and the income distribution within each group. Mirroring the relative proportion of households in Newark, the population of agents in our model is 39% German, 28% Irish, and 33% Yankee. Figure 2 shows the SEI distribution used for each ethnic group, generated empirically from the Newark data. While we use SEI as a proxy for income, it is important to realize that SEI is a quantification of a qualitative variable. The census recorded people's professions, so common generic answers such as "laborer" result in spikes in Figure 2 (such as the spike around 18 in all three plots). It was not possible to find statistical descriptors of these irregular distributions. Instead, for each ethnic group we divide the empirical income distribution into sixteen discrete "bins."

Each agent has two characteristics that remain fixed throughout an entire run of the model:

1. *ETH*: Ethnicity which can take the value "German," "Irish," or "Yankee"
2. *SEI*: Income, assigned based upon ethnicity-specific income distribution

where ETH_i and SEI_i refer to the ethnicity and income of the i^{th} agent

5.2 Creating Landscapes

Crooks (2010) notes that the landscapes used in Schelling models tend to be flat planes on which a single agent inhabits each cell. Our work partially overcomes these limitations; we generate landscapes such that housing units are distributed among the cells according to a bivariate, discretized Gaussian distribution with mean of zero and a user-specified variance. This means that cells can house a variable number of agents, and cells near the "city center" can house more agents than cells on the periphery. We use a single variance parameter, which we refer to as the housing density distribution parameter, resulting in an isotropic distance decay of housing units from the center. For example, when the housing density distribution parameter is fixed at 7.5 roughly 95% of all housing units are within 15 cells of the center of the landscape, when it is fixed at 4.5, 95% of all units are within nine cells of the center.

These landscapes roughly reflect population density gradients of industrial cities during the late 19th to early 20th centuries, a period during which the rise of mass transportation fundamentally restructured cities (Hall 1996). Bruegmann (2005) shows that London's density profile evolved from what, in our model, would be a very small setting of the housing density distribution parameter in 1841 to a what would be a large housing density distribution parameter in 1921. During this period the population density in London was essentially log linear with respect to distance from the city center. This relationship is approximated by the housing density distribution parameter. By 1950 London's population distribution had the characteristics of a log-normal density profile. We have explored more complex functional forms, including "donut" cities with an uninhabited core and anisotropic surfaces. We limit our discussion to the symmetrical Gaussian case, because it provides a reasonable and parsimonious description of cities during our study period.

The landscape is generated such that it has a vacancy rate of approximately 3–5%, that is, there are always 3–5% more housing units than there are agents. Changes to the housing density distribution parameter alter the spatial distribution of housing units but the number of agents, and therefore the number of housing units, remains fixed. We run our model on a landscape of 33x33 cells. Depending on the housing density distribution parameter selected, some number of these cells have a housing capacity of zero, meaning that they cannot accommodate agents and are not active in the model.

5.3 Utility Function

Agent behavior is controlled by utility, which is determined by a specialized utility function and measured along a 0–1 scale. The utility function contains an ethnic term (left term in Equation 1 below) and an economic term (right term in Equation 1 below). Each term is weighted by an agent-specific weight w_i, calculated as $w_i = \left(\frac{SEI_i}{SEI_{max}}\right)^{ME}$, where SEI_{max} is the maximum SEI in the city of Newark (96) and ME is a model parameter called Money Exponent (ME). For any agent i, w_i will be bounded between 0 and 1. Agents with higher SEI values will have values of w_i closer to 1, while agents with lower SEI values will have values of w_i closer to 0. If the Money Exponent (ME) is set to 1 the weight (w) increases as a linear function of SEI. This weight term is described in Figure 3, which plots the weight as a function of SEI for several values of ME.

An agent's utility at a given location in the landscape is based upon both its own characteristics and the characteristics of the other agents within its field of view (Equation 1). The field of view is determined by a model parameter called Agent Vision and denoted r (Table 1).

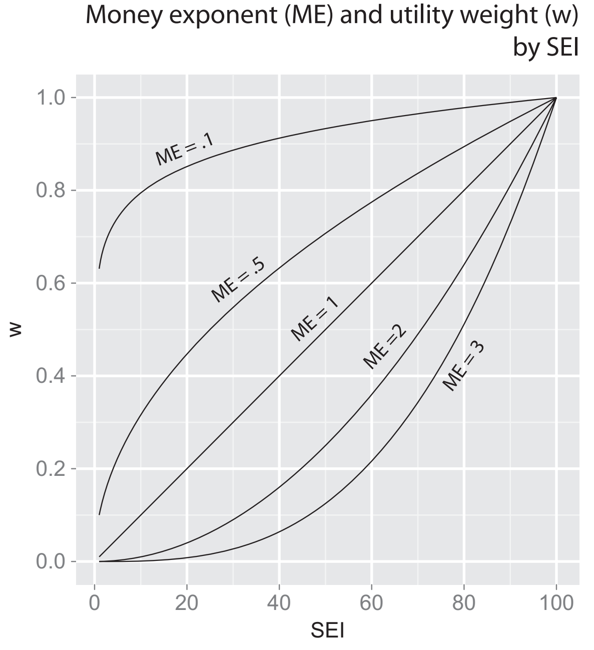

Figure 3 Money exponent and the utility weight

Table 1 Model parameter descriptions

Model parameter Name	Description
Agent vision (r)	This parameter determines how far agents will "look" (when compiling the set of other nearby agents) while calculating their utility.
Money exponent (ME)	Determines how an agent's own income affects the relative weights that agent places on the ethnicity and income level of other nearby agents.
Utility threshold	Agents are "satisfied" with their current location if their utility (see eq. 1) is equal to or greater than the utility threshold.
Housing density distribution	Determines the housing density profile of the city. Higher values lead to a concentration of housing in the center. Lower values lead to a more even distribution of housing units.

This vision parameter encodes the amount of information about the city available to agents in the calculation of utility. Vision has been viewed as a way to represent a host of factors that limit (or enhance) an agent's knowledge of its environs, such as real estate agents (Laurie and Jaggi 2003). To calculate the utility of its current location, an agent identifies all agents within r units of itself. We denote this set $N_i(r)$. For each agent j in $N_i(r)$ we evaluate it's ethnicity relative to the *ith* agent – δ_{ETH_i,ETH_j} takes a value of 1 if agent i and j have the same ethnicity otherwise 0. Hence, the ethnic term of the utility function is simply the proportion of all agents near agent i that are of the same ethnicity. The right term in the utility function measures the relative affluence of the agents within r units of the focal agent (i). We denote the median *SEI* within r units of the i^{th} agent using $\widetilde{SEI_i(r)}$.

$$utility_i = \left((1-w)\frac{\sum_{j=1}^{j\in N_i(r)} \delta_{ETH_i,ETH_j}}{|N_i(r)|} \right) + \left(w\,\frac{SEI_i - \widetilde{SEI_i(r)}}{\widetilde{SEI_i(r)}} \right) \tag{1}$$

The ethnic term (left) is multiplied by 1-w, and the economic term (right) is multiplied by w. This causes lower-income agents (w closer to 0) to place relatively more weight on the ethnic term than the economic term, and higher-income agents (w closer to 1) to place relatively more weight on the economic term than the ethnic term. In both the historic and modern contexts, it is common for ethnic segregation to decline as a function of income (Clark and Ware 1997, Spielman and Logan 2012). This aspect of our utility function captures this phenomenon.

The Money Exponent is central to the calculation of utility, as it determines how much weight most agents place on the ethnic term and the economic term. An agent near the upper end of the income distribution will always have a w near 1, while an agent with an *SEI* of 0 will always have a w of 0. The large majority of agents that fall somewhere in between on the *SEI* scale can place quite different weights on the ethnic term and economic term, depending on the value of *ME* (Figure 3).

5.4 Model Execution

At the initialization of each model run, all agents are placed onto the landscape into a random cell with an unoccupied housing unit. This random initial condition creates an integrated city. The model then proceeds iteratively, with every agent evaluating its utility during each iteration. Agents calculate their utility using the utility function described above, and compare the result with the utility threshold, a user-defined parameter on a 0–1 scale (Table 1). If an agent's utility is below the threshold, it moves to a new random cell with an unoccupied housing unit. This process continues until one of three stopping criteria is met. The first is based on the agent utilities: if 98% of the agents are satisfied (i.e. their utility exceeds the utility threshold), the model run terminates. The second stopping criterion is met if the number of unsatisfied agents has not changed by at least 10% in the last 20 iterations. This condition implies that the model has approached a steady state with fewer than 98% of the agents satisfied. The final stopping criterion is based on time: if the model has not met one of the other stopping criteria after 200 iterations, the run terminates.

At model termination, we output a series of statistics describing the aggregate urban pattern. We calculate pair-wise spatial and non-spatial segregation indices for all ethnic groups, using both the non-spatial dissimilarity index (D) and Wong's (2005) spatial measure of dissimilarity. We also output the correlation between population density and income, the average utility of the agents, the percent of agents that have a utility exceeding the utility threshold, and the number of iterations required for the run.

Figure 4 shows output from three model runs and some of the associated statistical descriptors. Each model is run with the same set of parameters but using different "landscapes," or spatial distributions of housing units. Cells are colored according to the relative ethnic mix of the agents located there, with red for Germans, green for Irish, and blue for

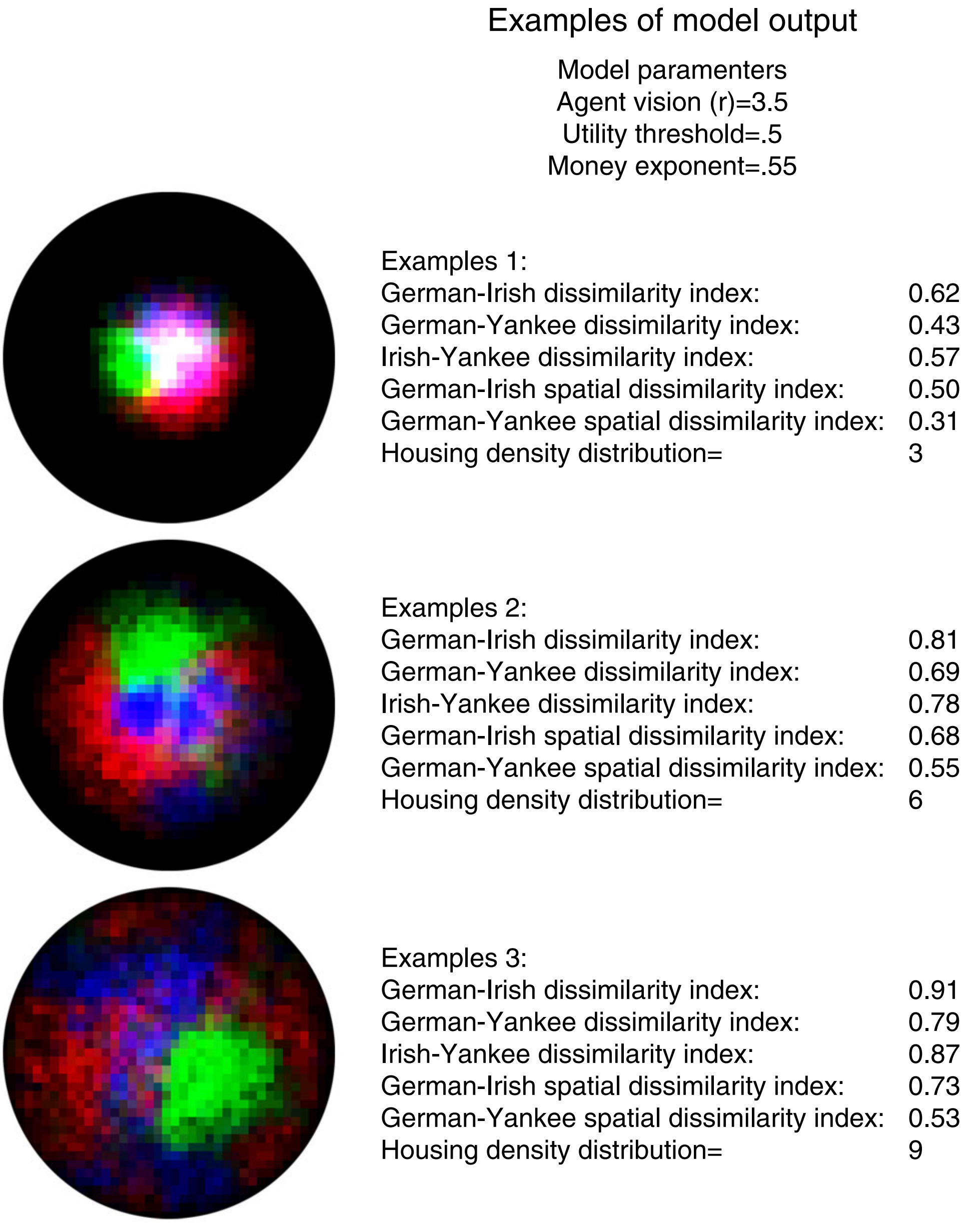

Figure 4 Examples of model output

Yankees. Cells that are solid red, green, or blue represent ethnically homogenous areas, while cells of intermediate colors represent significant ethnic mixing. The brightness of each cell corresponds to the population density. The white center region in Example 1 is a hyper-dense urban core. In the first example, 95% of all housing units are within six cells of the city center. The segregation indices observed tend to increase as the landscape becomes less dense and more spread out. In all three examples, some degree of ethnic mixing can be observed in the city center, while ethnically homogeneous neighborhoods tend to form at the periphery. The examples below are just three of many thousands of model runs conducted.

5.5 Experimental Design

5.5.1 Experiment 1: Full parameter sweep

We use a simple experimental design to identify combinations of model parameters that produced similar residential patterns to those observed in the empirical 1880 data. We specify a range of values for the money exponent, utility threshold, housing density distribution, and agent vision parameters and take their cross product, creating a four-dimensional parameter space. The Money Exponent parameter takes values between 0.2 and 1.95 incremented by 0.35. The utility threshold takes values between 0.2 and 0.6 incremented by 0.1, and the agent vision parameter takes values between 1 and 5 incremented by 1. Housing density distribution parameter takes values from 4.5 to 10.5 incrementing by 1. For each combination of parameters we conducted eight independent model runs, resulting in 10,080 independent model runs. We describe the results of these runs in the following section.

5.5.2 Experiment 2: Varying the built environment

We conduct a second, smaller experimental run in which we use a set of model parameters capable of reproducing the type of residential segregation patterns observed in Newark, NJ. The purpose of this experiment is to explore how the evolution of urban form that occurred between the late 19th and early 20th century could affect residential segregation patterns, holding preferences constant. We hold the entire set of parameters constant, except the housing density distribution parameter, which we vary from 2 to 15 units ("varying the built environment"). When the housing density distribution parameter is small, housing is concentrated in the core (like a walking city); when it is large, the core is less dense and housing is spread further from the center (as in a transit-oriented city).

5.6 Model Validation

Our model places an empirically-based population of agents onto an abstract surface, which makes it difficult to validate. This population matches the ethnic and economic composition of Newark, NJ in 1880. Our population includes only the three main ethnic groups, which together constitute roughly 80% of the city's population. However, these agents interact on an abstract space that does not resemble Newark's geography. The model is validated based upon its ability to reproduce residential settlement patterns with statistical similarities to the NAPP data and the patterns observed by Spielman and Logan (2012).

The parameter regime that produced residential segregation statistics closest to those observed in 1880 was Money Exponent = 0.55, the Utility Threshold of 0.4 or 0.6 and agent vision between 3 and 5. This parameter regime produced segregation statistics for each of the pairwise groups that were within 0.01 of the observed value. We refer to this set of parameters as the "fitted model" and used them to conduct the second experiment.

We also examined the relationship between SEI and proximity to the urban core. In 1880, as one moved out from the city center, SEI tended to decline. At the end of each model run, we recorded the correlation between the population density of each cell and the SEI of agents residing in that cell. When using landscapes with a dense core (i.e. a low value of housing the density distribution parameter), we consistently found positive correlations, suggesting that the wealthier agents tended to cluster together in the city center. However, there was some variability in this correlation – it ranged from a high of 0.8 to a low of −0.2 with a mean of 0.4. The correlation between population density and SEI is sensitive to the landscape, a phenomenon we hope to explore in the future.

6 Results

6.1 Experiment 1

Across the 10,000+ model runs in the first experiment, there is high degree of correlation among the pairwise spatial and non-spatial dissimilarity statistics between the different groups (Figures 4 and 5). In our data, we saw the lowest segregation among Irish and Yankee groups and higher levels of segregation between Germans and Irish, and Germans and Yankees. In Newark in 1880 we notice especially strong segregation between low income Irish and low income Germans. In the real world, this phenomenon is probably due to a complex set of social relations, including the linguistic isolation of German-speaking immigrants. German immigrant communities in the U.S. in the late 19th century formed self-contained neighborhoods in which the everyday language was German (Zunz 1982). Our model does not include language, however it does include group-specific income distributions. The Irish and German populations have relatively large segments in the lower end of the SEI distribution. These population segments will place more emphasis on ethnicity, resulting in ethnic segregation. This may create the lower correlations between the Irish-Yankee segregation statistics and the German-Irish segregation statistics observed in Table 2.

Given the high degrees of correlation we observe between the different measures of segregation, we focus our attention on a single outcome – the German-Irish spatial dissimilarity statistic.

Figure 5 shows how German-Irish segregation (vertical axis) responds to changes in the various model parameters (horizontal axes). They include the data from the 10,000+ model runs in the first experiment. These plots should be interpreted with caution; for example, the plot showing the utility threshold (lower right panel of Figure 5) shows the level of segregation when the threshold is set to 0.2, 0.3, 0.4, 0.5, and 0.6. For each of those five levels the other parameters in the model vary – the plot does not statistically isolate the effect of the parameter (i.e. it is not an effect plot). The lines in each of the four panels in Figure 5 are locally weighted scatterplot smoothing fits (LOWESS). In general, these plots show that there is substantial heterogeneity in the level of segregation produced when one parameter is held fixed and the others are allowed to vary within the experimental parameter space.

One notices that increases in the housing density distribution, utility threshold, and the money exponent parameters all, on average, increase the level of segregation in the resulting model. However, there seems to be a tipping point in the utility threshold parameter, beyond which further increases result in decreases in segregation. Through further experimentation we found that in a wide variety of parameter regimes including a utility threshold value above 0.55, the agents in the model failed to settle into stable patterns. Under these circumstances, many agents were unable to find any location that provided satisfactory levels of utility; in these instances the resulting segregation statistics are unusually low (below 0.2).

Increases in the agent vision parameter, on average, resulted in decreases in the level of segregation. However, few of the data points are near the LOWESS line – most are clustered in

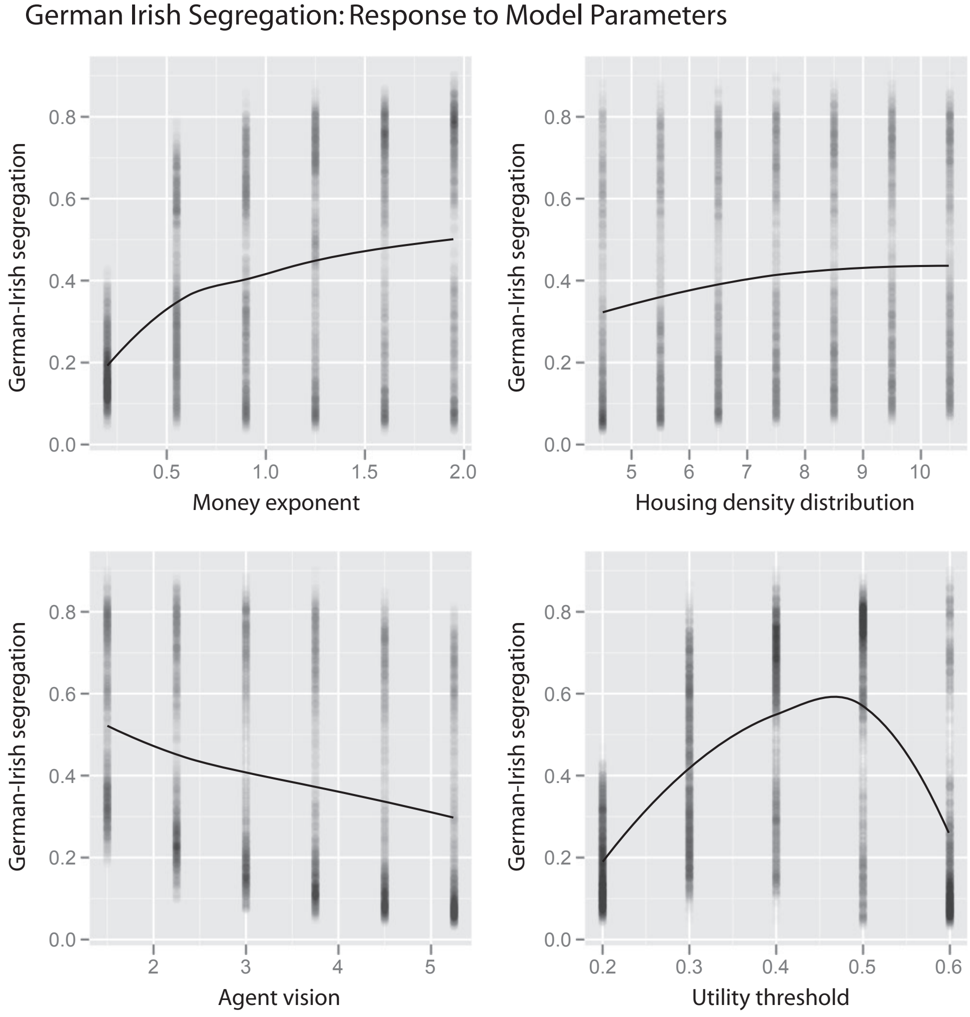

Figure 5 Experiment 1: Response to model parameters

outcomes with a high or a low level of segregation. These clusters of high and low values can be somewhat explained through the interaction plots in Figure 6.

These interaction plots show the average level of German-Irish segregation for every pairwise combination of model parameters. Horizontal and vertical axes take their minima and maxima from the parameter ranges used in Experiment 1 (see experimental design section). In this graphic, the color values of each pixel are determined by the average level of segregation for all model runs containing that combination of parameters. These plots provide a visual summary of how the model responds to variations in parameters. They are not to be confused with output from the model.

The right-most plot on the bottom row of Figure 6 shows the interaction of agent vision and housing density distribution parameters. The interaction between the two parameters is nonlinear: very high and very low values of agent vision are, for most values of the housing density parameter, associated with low values of segregation. However, when in the middle of

Table 2 Pearson correlation among segregation statistics in experiment 1

	German Irish spatial dissimilarity	German Yankee spatial dissimilarity	Irish Yankee spatial dissimilarity	German Irish dissimilarity	German Yankee dissimilarity	Irish Yankee Dissimilarity
German Irish spatial dissimilarity	1.00	0.89	0.88	0.83	0.77	0.66
German Yankee spatial dissimilarity	0.89	1.00	0.87	0.77	0.84	0.66
Irish Yankee spatial dissimilarity	0.88	0.87	1.00	0.80	0.78	0.79
German Irish dissimilarity	0.83	0.77	0.80	1.00	0.92	0.90
German Yankee dissimilarity	0.77	0.84	0.78	0.92	1.00	0.88
Irish Yankee dissimilarity	0.66	0.66	0.79	0.90	0.88	1.00

Figure 6 Experiment 1: Parameter interactions and German-Irish segregation

the values range for the agent vision parameter of the housing density parameter segregation levels tend to be high.

Generally, these interaction plots suggest complex and non-linear interactions among parameters – for example, the interaction between money exponent and utility threshold is difficult to explain. Because of the complexity and variability in our analysis of the full parameter space, we focus on a narrow set of parameters in Experiment 2. These parameters, the "fitted model," were selected from the results of the first experiment because they closely resemble the patterns we observe in the late 19th century – a relatively ethnically diverse and high-income central city surrounded by ethnically segregated, lower-income neighborhoods.

6.2 Experiment 2

In the second experiment, we held all parameters constant at their "fitted model" values, except for the housing density distribution, which we allowed to vary. We find that changes to the physical structure of the city (housing density distribution), holding all other model parameters constant, produces large changes in the average level of segregation (Figure 7). Compact, dense city forms tend to result in lower levels of ethnic and income segregation than less dense, more expansive forms. We notice this result in both the first and second experiments, though the effect is more pronounced in the second experiment with the fitted model.

For higher values of the housing density distribution parameter, the difference in housing capacity between the city center and the periphery is less pronounced, and housing units are more evenly distributed across the landscape. This causes the configuration of the model to become more like the traditional Schelling framework, where housing units are distributed

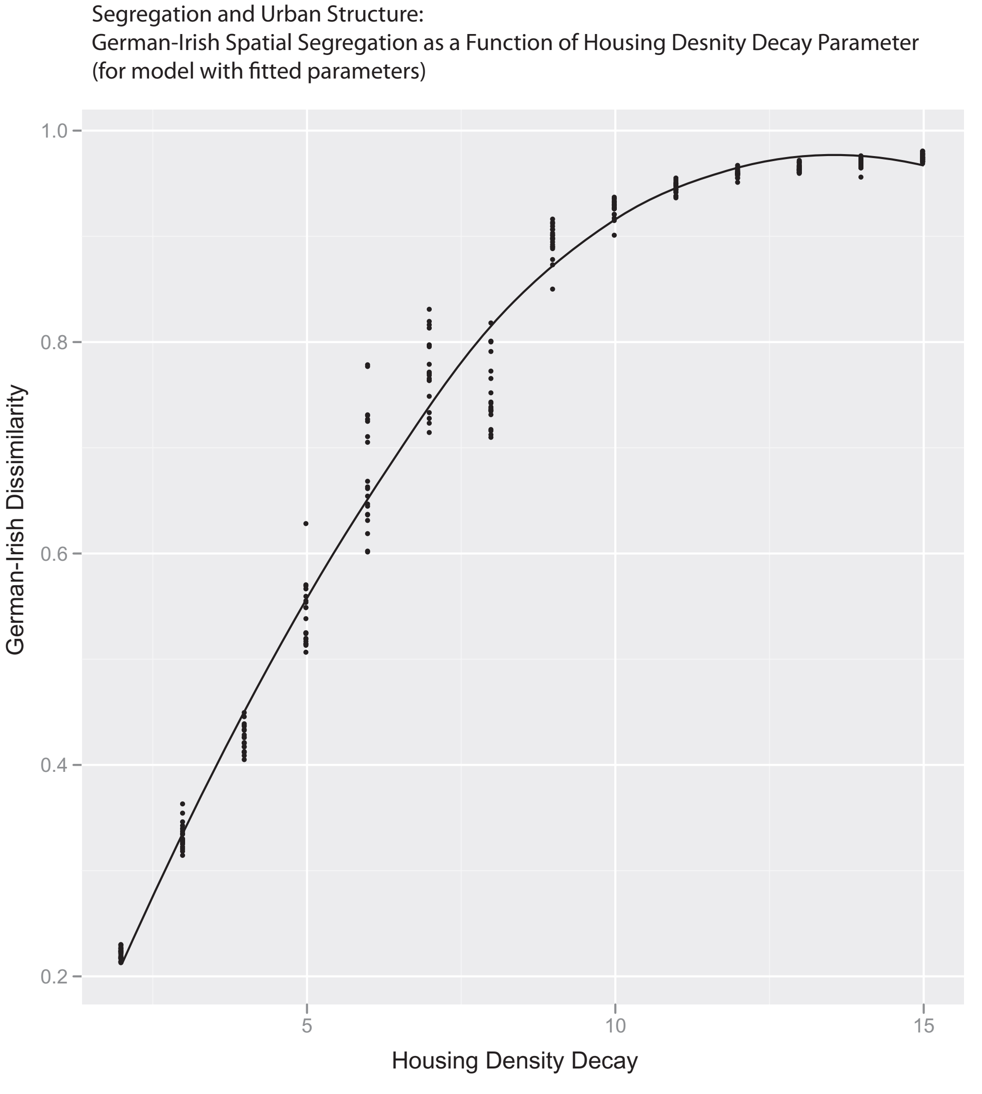

Figure 7 Experiment 2: Urban form and ethnic segregation

uniformly on a plane. Under these conditions, the model output moves closer to the traditional Schelling result of strong ethnic segregation, despite the fact that the higher income agents place little utility on ethnic homogeneity. The more expansive space of the model allows agents to sort themselves by both ethnicity and income, resulting in higher segregation along both ethnic and economic lines (Figures 7 and 8).

7 Conclusions

While not entirely surprising, the second experiment provides an important and intriguing result. In the social sciences it has long been held that the physical arrangement of people in

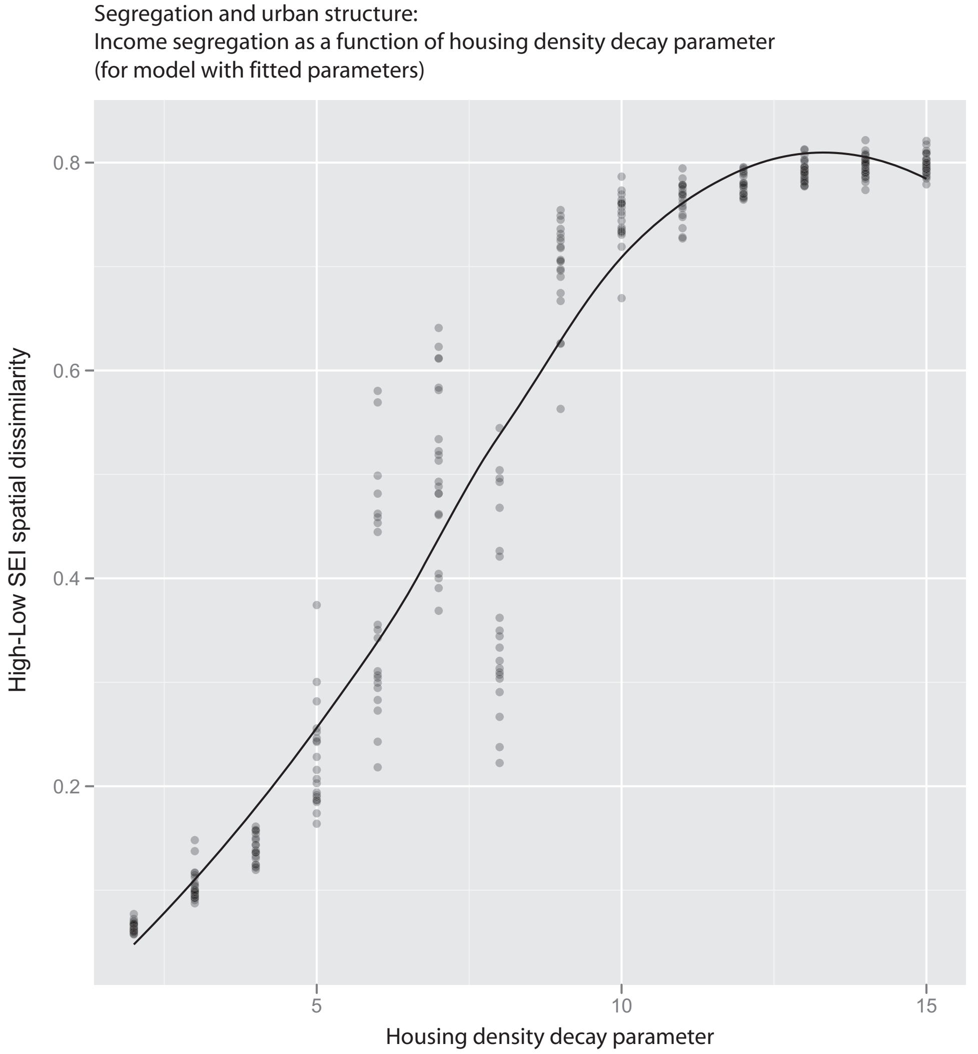

Figure 8 Experiment 2: Urban form and income segregation

space both reflects and shapes attitudes and social processes (Abbott 1997). In the literature on residential segregation these attitudes are often understood to be some type of preference – a preference for a neighborhood with a certain ethnic composition, housing stock, and/or economic characteristics (Dawkins 2004). The social processes relate to the economics of housing markets, discrimination, and differential access to information. However, we find that by altering the built environment of a city, while holding all social factors constant, we are able to produce substantial changes in segregation outcomes. A city's physical structure plays an important role in the manifestation of social attitudes and preferences. We develop this insight by allowing a fixed set of attitudes and a fixed process to play out on a varying landscape. Our changes to the built environment mirror historical changes that happened to many cities in the late 19th and early 20th centuries.

During this period, segregation by ethnicity, race, and income increased substantially in some cities (Zunz 1982). Our results mirror these increases. However, we are not suggesting that the changes to the social landscape of cities around the turn of the last century were necessarily due to changes in the built environment. It is a fact that the two occurred contemporaneously, but a Schelling-like agent-based model cannot establish causation. Our model establishes urban form as a plausible and novel contributor to observed metropolitan segregation patterns. We find that the built and social environments of cities co-evolve. In discussions of segregation this co-evolution does not receive sufficient attention. How significant is the effect of the built environment relative to other factors such as preferences, economic constraints, and discrimination? Do preferences evolve more or less rapidly than the built environment? We are unable to answer these questions, but clearly, they affect how one interprets changes in residential segregation.

The changes in residential segregation produced by our model are quite pronounced. We believe that these changes raise epistemological questions about the inferences one can reliably draw from changes in socio-spatial patterns. Our work suggests that changes in the observed level of segregation do not necessarily require changes to attitudes, preferences, or institutions. Over the past 20 years, segregation has generally declined in the U.S. Concurrently, there have been significant increases in residential density in many inner cities (resulting in changes to the spatial distribution of housing units). Our model raises a question – might some portion of the recent declines in residential segregation be due to changes in the physical structure of cities? We believe this question warrants further investigation.

References

Abbott A 1997 Of time and space: The contemporary relevance of the Chicago school. *Social Forces* 75: 1149–82

Bruch E and Mare R 2006 Neighborhood choice and neighborhood change. *American Journal of Sociology* 112: 667–709

Bruegmann R 2005 *Sprawl: A Compact History*. Chicago, IL, University of Chicago Press

Charles C Z 2001 Processes of residential segregation. In O'Connor A, Tilly C, and Bobo L (eds) *Urban Inequality: Evidence From Four Cities*. New York, Russell Sage Foundation: 217–71

Clark W A V 1991 Residential preferences and neighborhood racial segregation: A test of the Schelling segregation model. *Demography* 28: 1–19

Clark W A V and Fossett M 2008 Understanding the social context of the Schelling segregation model. *Proceedings of the National Academy of Sciences, USA* 105: 4109–14

Clark W A V and Ware J 1997 Trends in residential integration by socioeconomic status in southern California. *Urban Affairs Review* 32: 825–43

Crooks A T 2010 Constructing and implementing an agent-based model of residential segregation through vector GIS. *International Journal of Geographical Information Science* 24: 661–75

Dawkins C J 2004 Recent evidence on the continuing causes of black-white residential segregation. *Journal of Urban Affairs* 26: 379–400

Dorling D, Mitchell R, Shaw M, Orford S, and Smith G D 2000 The Ghost of Christmas Past: Health effects of poverty in London in 1896 and 1991. *British Medical Journal* 321: 1547–51

Dyos H, Cannadine D, and Reeder D 1982 *Exploring the Urban Past: Essays in Urban History*. New York, Cambridge University Press

Fischer M J 2003 The relative importance of income and race in determining residential outcomes in U.S. urban areas, 1970–2000. *Urban Affairs Review* 38: 669–96

Fossett M 2006 Ethnic preferences, social distance dynamics, and residential segregation: Theoretical explorations using simulation analysis. *Journal of Mathematical Sociology* 30: 185–274

Greenberg S 1981 Industrial location and ethnic residential patterns in an industrializing city: Philadelphia, 1880–1975. In Cuttler W W and Gillette H (eds) *Philadelphia: Work, Space, Family, and Group Experience in the Nineteenth Century*. New York, Oxford University Press: 204–32

Gregory I N and Healey R G 2007 Historical GIS: Structuring, mapping and analyzing geographies of the past. *Progress in Human Geography* 31: 638–53

Gregory I N 2008 Different places, different stories: Infant mortality decline in England and Wales, 1851–1911. *Annals of the Association of American Geographers* 98: 773–94
Hall P 1996 *Cities of Tomorrow*. London, Blackwell
Hershberg T (ed) 1981 *Philadelphia: Work, Space, Family, and Group Experience in the Nineteenth Century: Essays Toward an Interdisciplinary History of the City*. New York, Oxford University Press
Iceland J, Weinberg D H, and Steinmetz E 2002 *Racial and Ethnic Residential Segregation in the United States, 1980–2000*. Washington, D.C., U.S. Census Bureau, Special Report Series
Laurie A J and Jaggi N 2003 Role of 'vision' in neighbourhood racial segregation: A variant of the Schelling segregation model. *Urban Studies* 40: 2687–704
Lee B A, Reardon S F, Firebaugh G, Farrell C R, Matthews S A, and O'Sullivan D 2008 Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review* 73: 766–91
Logan J R, Jindrich J, Shin H, and Zhang W 2010 Mapping America in 1880: The urban transition historical GIS project. *Historical Methods* 44: 49–60
Logan J R, Spielman S, Xu H, and Klein P N 2011 Identifying and bounding ethnic neighborhoods. *Urban Geography* 32: 334–59
Logan J R, Stults B J, and Farley R 2004 Segregation of minorities in the metropolis: Two decades of change. *Demography* 41: 1–22
Minnesota Population Center 2008 *North Atlantic Population Project: Complete Count Microdata* (Version 2.0, Machine-readable database). Minneapolis, MN, Minnesota Population Center
Schelling T C 1969 Models of segregation. *American Economic Review* 9: 488–93
Schelling T C 1971 Dynamic models of segregation. *Journal of Mathematical Sociology* 1: 143–86
Schelling T C 1978 *Micromotives and Macrobehavior*. New York, Norton
Sobek M 1996 Work, status, and income: Men in the American occupational structure since the late nineteenth century. *Social Science History* 22: 169–207
Spielman S E and Logan J R 2012 Using high resolution population data to identify neighborhoods. *Annals of the Association of American Geographers* 102: in press
Warner S B 1962 *Streetcar Suburbs: The Process of Growth in Boston 1870–1900*. Cambridge, MA, Harvard University Press
Watson T 2006 Inequality, metropolitan growth, and residential segregation by income. In Burtless G and Pack J R (eds) *Brookings-Wharton Papers on Urban Affairs*. Washington, D.C., Brookings Institution: 1–52
White M J 1983 The measurement of spatial segregation. *American Journal of Sociology* 88: 1008–18
Wilensky U 1999 *NetLogo*. Evanston, IL, Northwestern University, Center for Connected Learning and Computer-Based Modeling (available at http://ccl.northwestern.edu/netlogo/)
Wong D W S 2005 Formulating a general spatial segregation measure. *Professional Geographer* 57: 285–94
Zhang J 2004 A dynamic model of residential segregation. *Journal of Mathematical Sociology* 28: 147–70
Zunz O 1982 *The Changing Face of Inequality: Urbanization, Industrial Development, and Immigrants in Detroit, 1880–1920*. Chicago, IL, University of Chicago Press

