The effect of residential segregation on social segregation: evidence from Flickr

Kirsten Cornelson*

Abstract

In this paper, I simulate the effect of desegregating U.S. cities on inter-racial interactions. For this exercise, I make use of a newly created dataset measuring inter-racial interactions through social media photographs. In my simulation, I focus on a single causal effect of neighborhoods on interactions: the effect operating through the disutility of travel. While the disutility of travel is quantitatively important (in that it predicts significant deviations from perfectly integrated social behavior, even in the absence of preferences over the race of interaction partners), the estimated effect of desegregating cities on inter-racial interactions is negative. This is because desegregation eliminates the positive effect of residential sorting on inter-racial interactions.

*University of Notre Dame Economics Department. 3051 Jenkins-Nanovic Halls, Notre Dame IN, 46556. Email: kcornels@nd.edu.

1 Introduction

There is a high degree of physical and social isolation between blacks and whites in the United States. In 2010, the average black American lived in a Census block that was 54.1% black, despite the fact that blacks made up only 12.2% of the population as whole.[1] Even more striking, Echenique and Fryer (2007) report that, as of 2000, over 60% of Census blocks in most states contained residents of only one race. Blacks and whites rarely marry each other: of all married couples with either a black or white spouse in 2014, just 1.2% were inter-racial. There is significant segregation in the workplace (Hellerstein and Neumark, 2008), in high school friendship networks within the same school (Echenique and Fryer, 2007), and between university classmates (Marmaros and Sacerdote, 2006).

To what extent are the physical and social dimensions of racial isolation causally related to each other? Sociologists have long argued that residential and social segregation reinforce each other in a self-perpetuating cycle (e.g., Wilson, 1987; Massey and Denton, 1993; Krysan and Crowder, 2017). For this to be true, it must be the case that neighborhoods have a causal effect on our probability of interacting with different people. We currently understand very little, however, about the role of neighborhoods in shaping people's social lives. This is due in part to the difficulty in estimating the causal impact of neighborhoods in the presence of residential sorting. Because individuals with strong preferences for same-race interactions have an incentive to live in more segregated areas, it is difficult to disentangle the effect of neighborhoods from the effect of the preferences that help create residential segregation. Furthermore, even with good estimates of the causal effect of changing neighborhoods on individuals, it is difficult to understand the potential general equilibrium effects of policies that significantly alter the distribution of races across different neighborhoods.[2]

In this paper, I make progress on this important question by simulating the effect of a policy that completely desegregates American cities on social segregation by race. In this simulation, I focus on a single causal channel through which neighborhoods influence social interactions, one on which I argue that we already have good information: the effect of physical distance. While neighborhoods may influence social interactions in a variety of ways, the most fundamental of these channels is proximity

[1] Calculation by author, using 2010 Census data.

[2] The Fair Housing Act of 1968 requires government agencies to actively pursue the goal of residential desegregation. Today, this goal is most prominantly pursued by HUD, which promotes socioeconomic and racial integration both directly (for example, through Section 8 housing vouchers) and indirectly through funding rules and/or sanctions against other government agencies (for example, in a recent conflict with the City of Houston (Morris, 2018).)

itself. Social interactions require travel, and this travel acts as a price; moving two individuals into neighborhoods that are further apart will reduce the frequency of their interactions by increasing the cost of interactions between them. Focusing narrowly on the effect of physical distance is helpful, because it allows me to sidestep the identification problem that plagues most estimates of neighborhood effects. As I argue later in this paper, the causal effect of physical distance on social interactions is captured by a parameter that has already been estimated many times: the disutility of travel. Using estimates of this parameter from previous research by Thomadsen (2005), Davis (2006), and McManus (2007), among others, in a structural model of social interaction decisions, I simulate how inter-racial interactions would change if we eliminated the excess physical distance between members of different races. Because this estimate ignores other influences of neighborhoods on social interactions (for example, through the effects of schools or other local public goods), it can be thought of as a lower bound on the effect of residential desegregation on social segregation.

The first step in this procedure is to measure the existing degree of social segregation in American cities. As it turns out, this is a non-trivial task. There currently exist no large-scale data that ask individuals about the race of their interaction partners. To measure social segregation, I create a new source of data that contains information on the inter-racial interaction behavior of a large sample of Americans. These data are derived from a large sample of photographs posted on the social media website Flickr from 2006-2015. I run these photographs through face detection and race classification software to measure the racial breakdown of people appearing in Flickr users' photos. Using a survey, I show that the distribution of race in social media photographs is a highly accurate measure of the racial breakdown of real-life social contacts. The photographs I use are also geotagged (they contain latitude and longitude coordinates appended by the camera at the time the photo is taken), which allows me to link Flickr users to cities and neighborhoods.

With this measure of inter-racial interaction in hand, I start by decomposing individuals' inter-racial interaction behavior into three pieces. The first is a piece that is explained by the causal effect of physical distance. Because residential segregation implies a higher time cost for inter-racial interactions relative to same-race interactions, it presents a barrier to racial integration. I calculate the magnitude of this effect by combining existing estimates of the distaste for travel with geographic information on the distance between neighborhoods within cities, and on the existing population distribution by race. The parameter I calculate tells me how often individuals would interact with people of different races, if physical distance were the only factor influencing their

interaction decisions. The second and third pieces are residual terms, which in the context of my model can be interpreted as reflecting preferences for cross-racial interactions, and the effect of residential sorting, respectively. The preference piece captures the average person's tastes for interacting with a randomly selected different-race individual in their city. The sorting piece captures the additional social interactions that are generated because individuals tend to live near particular different-race people that they like better than the average other-race individual in their cities.

Next, I use these parameters to simulate the effect of completely desegregating each city by randomly assigning individuals to neighborhoods, such that each neighborhood (Census tract) contains a representative sample of individuals in the city. This procedure eliminates the excess physical distance between members of different races, which is the only causal effect of neighborhoods that I consider in my model. In particular, I assume that the "preference" terms I estimate remain fixed. If there are other important causal effects of neighborhoods, it may be the case that these residual terms change with desegregation: in addition to removing the effect of physical distance, desegregation may create other changes that cause blacks and whites to interact more frequently with each other. Because I ignore these other potential channels, the change in inter-racial interaction behavior should be treated as a lower bound on the effect of desegregation on social interactions.

This exercise leads to a surprising result. Despite the fact that I have assumed a positive, causal effect of neighborhoods on interactions, desegregation causes the inter-racial interaction rate to *fall*. This is because desegregation does not just affect physical distance: it also eliminates the piece of inter-racial interactions caused by residential sorting. Individuals are no longer close to the other-race neighbors that they enjoy interacting with, and rather than replacing them with less-preferred individuals, they simply stop interacting as frequently. Both the rate of overall interactions and the proportion of cross-racial interactions fall, on average, leading to a large decline in the number of cross-race interactions. To the best of my knowledge, this potentially positive role of residential sorting for cross-racial interactions has not previously been acknowledged.

As I show in a supplementary exercise, the negative effect of desegregation on social interactions does not arise because the causal effect of physical distance on social interactions is small. In particular, I examine how the inter-racial would change in a different simulation exercise: one where individuals remain in their existing residential locations, but all factors influencing social interactions *except* physical distance are eliminated. I show that physical distance alone can account for a 15-20% reduction in inter-racial interactions, compared to a world in which neither physical distance nor

other factors are present. This is line with previous research, which shows that individuals highly value their time and avoid excess travel. Therefore, while my model does not capture all causal effects of neighborhoods, it does account for an important factor influencing interaction decisions. This causal effect is outweighed, however, by the negative effect of eliminating residential sorting.

In addition to fitting into the large literature on the causes and consequences of residential segregation, this paper is also related to a smaller literature that directly examines the relationship between physical distance and the probability of interacting. Empirically, Marmaros and Sacerdote (2006) causally identify the effect of proximity on social interactions in college dorms using random assignment to rooms. They show that individuals who live closer together in dorms are more likely to interact. It is difficult, however, to assess the quantitative relevance of their results in more general settings. Patacchini, Picard and Zenou (2015) is the only other paper that presents a theoretical model in which the probability of interaction is causally affected by distance. The authors model the returns to interaction as increasing in the partners' interaction rate (which they interpret as social capital), and show that more centrally located agents will interact more often. Using the Add Health dataset (a survey of teenagers that asks individuals to nominate up to 5 friends within the same survey), they show that agents are more likely to be friends when they live closer together, and that more physically central agents appear to be more central in the network as well. While the model I present below is similar in spirit to the model presented in Patacchini, Picard and Zenou (2015) (except for the latter's emphasis on increasing returns to social capital, which I do not consider), my model is better-suited to the kind of quantitative exercise I propose, because it produces equilibrium equations based on a very small number of parameters. Our papers also differ in the outcomes they consider: Patacchini, Picard and Zenou (2015) examine the implications of their model for the level of social capital in a city, and its relationship to transportation costs, while I examine the consequences of residential segregation in predicting social segregation between racial groups. Finally, the data I use to examine the predictions of my model cover a wider segment of the population, and are much larger - the Add Health dataset contains only about 1500 individuals with information on both interactions and residential location, while my Flickr data contains information on 90,000 individuals over the course of approximately 10 years.

The paper that is closest to mine is Davis et al. (forthcoming), which examines the role of spatial frictions in generating racial segregation in restaurant visits, using data from Yelp. The authors estimate the distaste for travel using the frequency of restaurant visits based on distance from the

user's home, work or commute path. They find that spatial frictions account for about half of the observed segregation in restaurant choices. Compared to this earlier paper, mine differs in using an external measure of spatial frictions, as well as by measuring segregation in social interactions directly rather than inferring it from restaurant visits. Additionally, I am able to simulate the effect of desegregating residential location on interactions. Our results on the importance of spatial frictions in driving interaction behavior are similar, however.

Finally, there is a large literature that attempts to assess the importance of physical distance in the market for spatially differentiated goods, such as gas stations, movie theatres, coffee shops and liquor stores (e.g., Thomadsen, 2005; Davis, 2006; McManus, 2007; Manuszak and Moul, 2009; Houde, 2012; Seim and Waldfogel, 2013). The model of social interactions I present below is analagous to the structural models of supply and demand presented in this literature; the key difference is that both the "supplier" and the "customer" travel in my case. I provide a detailed overview of this literature and its results with respect to the disutility of travel in the data section below.

In the next part of the paper, I outline the structural model that underlies both of my simulation exercises. This is adapted from the marriage matching model of Choo and Siow (2006). In the following section, I show how the model can be used for each of the exercises I propose. Section 4 describes the data that I require to implement the simulation exercises, including a description of the new Flickr dataset that I have developed for this purpose. Section 5 contains my results, while the last section concludes.

2 Theory

In this section, I present a discrete choice, transferable utility model of social interactions adapted from the marriage matching model of Choo and Siow (2006). The equilibrium conditions from this model will provide the basis for my simulation exercise. In this model, each agent resides at a location in a city, and makes a decision each period about whether to interact, and with whom. The partner characteristics that they choose are race and neighborhood of residence, which may affect the utility of an interaction either directly or indirectly (for example, through the correlation between neighborhood and educational attainment.) The transfer that clears the market for interactions in the model is the choice of meeting point, which affects utility because agents are assumed to dislike

travel.[3]

The transferable utility assumption implies that the equilibrium frequency of interactions between any two groups of agents will depend only on the joint surplus that is created by interactions between them, and on population supplies. All else equal, the total surplus created by interactions will be declining in the physical distance between two agents, because they must jointly travel this distance in order to meet.[4] The causal effect of distance on social interactions can therefore be captured by agents' disutility of travel.

While the language I use to explain the model describes individuals' residential locations, the model can equally be interpreted as describing individuals' decisions when they start from work. Empirically, the only difference between the decision to interact when starting from work or home is that the distribution of individuals' starting points (neighborhoods) will differ: an agent's "neighborhood" will describe their location of work rather than their location of residence.[5] While I will apply the model to the case where individuals are likely to start from home (i.e. weekend interactions), it would be straightforward to analyze the impact of integrating workplaces in a similar manner.[6]

2.1 Model setup

Agents live in neighborhoods in a city, which are spatially arranged on a line. Agents must decide whether to interact with anyone, and if so, with whom to interact. I assume that each person chooses a single interaction partner.[7] The partner characteristics that the agent chooses are race and location.[8] I refer to agent types as (i, r), where i is indexes the individual's location on the line (her neighborhood) and r indexes their race. I assume that there are a finite number of racial

[3]The results of the model will be similar so long as there is any way for agents to transfer utility between them (through choice of activity, who pays, etc.)

[4]In the model, I assume that agents always meet somewhere on the line in between them. In real life, agents may choose to visit locations elsewhere in the city. This can be incorporated into the model by allowing agents to have direct preferences over particular locations, which makes this an imperfectly transferable utility model in the manner of Galichon, Kominers and Weber (2016). This extension will not fundamentally change the predictions of the model, however, because the joint surplus of an interaction will still decline in the distance between any two agents' homes; this is the minimum distance that they must jointly travel.

[5]The overall level of interactions may also differ when starting from work because of "time of week" effects. For example, the individual may get less utility from socializing after work because she is tired. So long as this applies equally to all interactions, it will not affect the results of my model.

[6]In what follows, I treat interactions starting from home as being determined separately from interactions from work. In doing this, I am making use of an implicit assumption in the model: that choices over each interaction are made independently. This rules out, for example, a declining marginal utility of interactions with a given race.

[7]The model should extend to the case where people meet in groups, with a minor extension: the city must now be 2-dimensional. This is so that it is possible to find a location that perfectly compensates every group member.

[8]In practice, agents may care about a wide variety of partner characteristics, such as age or education. As I describe below, preferences over the race and location of interaction partners can be interpreted as reflecting average values of these other characteristics across different groups, as well as any race- or location- specific preferences.

groups R and neighborhoods N. If two individuals decide to interact, they meet at a third location in the city m, which is somewhere on the line in between them.[9]

If an individual agent g of type (i, r) interacts with an individual h of type (j, s) at a meeting point m, his utility is:

$$U_{ghirjs} = \alpha_{js}^{ir} - 2\delta d(i, m) + \epsilon_{ghirjs}$$

where α_{js}^{ir} is the average utility generated for agents of type (i, r) from socializing with an agents of type (j, s); $d(i, m)$ is the physical distance between the agent's home and the meeting point; δ is the disutility of distance[10]; and ϵ_{ghij} is an I.I.D. shock with a type I extreme value distribution. The term α_{js}^{ir} will capture average levels of education, age, or other characteristics that affect the match value between a typical member of type (i, r) and a typical member of type (j, s). The term ϵ_{ghirjs} can be interpreted as reflecting individual partners' deviations from the average level of characteristics for members of their types; for example, ϵ_{ghirjs} may be positive if both partners have higher-than-average education for their tract and race. This term also reflects more idiosyncratic shocks, such as personality characteristics, that affect the valuation of a match.

The agent may also choose not to socialize at all, which I denote as choosing partner type "0". I normalize the intrinsic utility from spending time alone to zero.

2.2 Equilibrium

Following Choo and Siow (2006), the model described above will lead to a quasi-demand function describing the number of (j, s) type interactions demanded by all individuals of type (i, r), which depends on the meeting point m (interpreted as the "price" in this model). The quasi-demand function takes the form

$$ln(\mu_{irjs}^{q}) = ln(\mu_{ir0}) + \alpha_{js}^{ir} - 2\delta d(i, m) \tag{1}$$

[9] As I describe below, the particular location m that is chosen will adjust to ensure that the market clears.

[10] I model the effect of distance as entering linearly into the individual's production function. This implies that each kilometer traveled has the same effect on utility. It is possible to extend the model to capture a non-linear effect of distance, which has been found to be empirically relevant by Davis (2006). In this case, utility is not perfectly transferable through the choice of meeting point, so agents must have some other way to compensate each other (through choice of activity, who pays, etc) in order to clear the market. In an earlier version of this paper (Cornelson, 2017), I show that incorporating non-linear utility produces results similar to increasing δ by approximately an order of 2. I show below how the results change when I double δ.

where μ^q_{irjs} is the total number of type (j,s) interactions demanded by agents of type (i,r) and μ_{ir0} is the equilibrium number of (i,r) agents who choose to spend time alone. In equilibrium, demand for these interactions by agents of type (i,r) must equal the "supply" of these interactions by agents of type (j,s):

$$ln(\mu^s_{irjs}) = ln(\mu_{0js}) + \alpha^{js}_{ir} - 2\delta d(j,m) \tag{2}$$

The meeting point m will adjust to ensure that this is the case.[11] Setting demand for interactions equal to supply, solving for m, and plugging this back into the quasi-demand equation gives:

$$\ln\left(\frac{\mu_{ijrs}}{\sqrt{\mu_{i0}\mu_{0j}}}\right) = \frac{1}{2}[\alpha^{ir}_{js} + \alpha^{js}_{ir}] - \delta d(i,j)$$

To close the model, note that there is an adding-up constraint. If we denote the population of race r living at i as $f^r(i)$, then $\mu_{ir0} + \Sigma_{t\in R}\Sigma_{k\in N}\mu_{irkt} = f^r(i)$. This lets us rewrite the equilibrium equation entirely in terms of the endogenous terms μ_{irjs} and the parameters of the model:

$$ln\left(\frac{\mu_{irjs}}{\sqrt{(f^r(i) - \Sigma_{t\in R}\Sigma_{k\in N}\mu_{irkt})(f^s(j) - \Sigma_{t\in R}\Sigma_{k\in N}\mu_{ktjs})}}\right) = \frac{1}{2}[\alpha^{ir}_{js} + \alpha^{js}_{ir}] - \delta d(i,j) \tag{3}$$

This equation says that the frequency of interactions between types (i,r) and (j,s) (scaled by a function of the total number of partners of each type) is equal to the per-partner surplus created by interactions between these types of individuals.[12] Note that it is only the total surplus that matters here; any difference between α^{ir}_{js} and α^{js}_{ir} is irrelevant, since the agent who gets more utility from the interaction can always compensate the other partner through the choice of meeting point m. For the purposes of the simulation exercise, I will rewrite the term $\frac{1}{2}[\alpha^{ir}_{js} + \alpha^{js}_{ir}] = \overline{\alpha}_{irjs}$, the average utility generated by the interaction.

Equation 3 highlights the role of physical distance in generating social interaction behavior. Because individuals dislike travel, we can increase or destroy the surplus from an interaction by moving agents geographically further apart or closer together, as captured by the term $-\delta d(i,j)$.

[11] Specifically, setting Equation 1 equal to Equation 2 and solving for the distance travelled by individual i gives $d^*(i,m) = \frac{1}{2}d(i,j) + \frac{1}{4\delta}ln\left(\frac{\mu_{ir0}}{\mu_{0js}}\right) + \frac{1}{4\delta}[\alpha^{ir}_{js} - \alpha^{js}_{ir}]$.

[12] Note that this identical to the standard result from discrete choice problems that forms the basis for logistic regression models. The only difference is that we have extended the problem to a case where agents must coordinate their decisions. The assumption of transferable utility permits this coordination to take place.

This is one reason why we should expect interaction rates to change if we alter the geographic distribution of individuals within cities. In what follows, I will explore the consequences of changing this distribution, holding other components of the model - in particular, the average utility terms $\overline{\alpha}_{irjs}$ - constant.

This equilibrium equation holds for every $(i,r),(j,s)$ pair in the city. If we let N be the number of types, this condition gives us a system of $\frac{(1+N)N}{2}$ equations. Choo and Siow (2006) show that, given values for the right-hand side of the equilibrium equation and a vector of population supplies $f^r(i), f^s(j)$ for each neighborhood, there is a unique vector of social interactions μ_{irjs} (of size $\frac{(1+N)N}{2}$) that will solve this system of equations. I use this fact in my simulation exercises, described in the following subsections of the paper.

3 Simulation exercise

In my main simulation exercise, I explore the potential for policy makers to influence inter-racial interactions by reallocating individuals' residential locations within cities. Specifically, I consider the extreme case where a policy maker is able to completely randomize individuals to neighborhoods. Using the equilibrium condition from my model, I simulate the both the overall interaction rate and the inter-racial interaction rate that would occur after this randomization. The steps of the simulation exercise, which I describe in more detail below, are as follows:

1. Decompose interaction rates into a piece explained by preferences, a piece explained by distance, and a piece explained by residential sorting

2. Randomly reassign residential locations, so that there is an even distribution of each race across tracts.

3. Calculate the frequency of inter-racial interactions that would take place with this new distribution of individuals, holding the preference terms estimated in Step 1 constant.

There are two important caveats to this exercise First, the "preference" terms I estimate in step 1 are actually residuals which capture all influences on social interactions other than physical distance. If there are important neighborhood effects other than physical distance (for example, through the effect of schools on interaction behavior), it may not be appropriate to hold these terms constant when we reallocate individuals to new neighborhoods in cities. My estimates should therefore be

thought of as lower bounds on the impact of residential desegregation on social segregation. Secondly, the exercise I consider is quite extreme: while it may be possible for policy makers to decrease residential segregation using tools such as housing vouchers, it is not likely that they will achieve (or want to achieve) a completely random distribution of people across neighborhoods. Despite these caveats, we will see that the exercise provides important insights on the potential unintended consequences of policies altering the spatial distribution of individuals within cities.

A second important feature of this exercise is that it can be used to analyze the impact of desegregation of either residential locations and workplaces together, or residential locations alone. In particular, if we assume that the reallocation of individuals holds at all points in time, we can interpret the model as reflecting a change that integrates workplaces as well as residences. If we instead prefer to think of an experiment that integrates residential locations only, we can assume that the level of interactions only changes during times when individuals are likely to be at home (i.e., weekends). While I will focus on the latter interpretation, it is straightforward to calculate the total effect on interactions by appropriately weighting the (unchanged) interaction behavior starting from work and the altered interaction behavior starting from home.

3.1 Step 1

The first step of my simulation procedure is to estimate the average utility created by different types of interactions. Using the equilibrium condition given in Equation 3 in the last section, and rearranging, we can write the equilibrium number of interactions between types (i, r) and (j, s) as:

$$\mu_{irjs} = e^{\overline{\alpha}_{irjs} - \delta d(i,j)} \sqrt{\mu_{ir0}\mu_{0js}} \tag{4}$$

Recall here that $\overline{\alpha}_{irjs}$ is the average utility generated by an interaction between type (i, r) and (j, s), and that μ_{ir0} is the number of partners of type (i, r) that remain unmatched in equilibrium. All of the terms on the right hand side of this equation will have observable counterparts in my data, with the exception of the preference parameters $\overline{\alpha}_{irjs}$. Specifically, I can calculate the distance term $-\delta d(i, j)$ using existing estimates of the parameter δ along with geographic information about the distance between pairs of neighborhoods within cities. The terms μ_{ir0} and μ_{0js} can be calculated using information on the population of each neighborhood, combined with information on the average interaction rate. I describe where I get each of these pieces of information in the next section.

Collecting this group of observable terms into a single parameter $\hat{\mu}_{irjs}$, I can rewrite the equilibrium equation as:

$$\mu_{irsj} = e^{\overline{\alpha}_{irjs}} \hat{\mu}_{irjs} \tag{5}$$

Note that the observable term $\hat{\mu}_{irjs} = e^{-\delta d(i,j)} \sqrt{\mu_{ir0}\mu_{0js}}$ can be interpreted as the interaction frequency that would take place if we eliminated preferences over interaction partner characteristics ($\overline{\alpha}_{irjs} = 0,\ \forall ir, js$).[13] The term $e^{\overline{\alpha}_{irjs}}$ represents the difference between the actual number of interactions and those that occur in this imaginary, preference-less situation.

If I observed μ_{irjs}, the actual interaction frequency for every neighborhood-race pair, I would be able to use Equation 5 to calculate $\overline{\alpha}_{irjs}$ directly. Unfortunately, however, my data do not allow me to observe the interaction rate between every neighborhood/race pair. Instead, I observe the total frequency with which an individual living in neighborhood i, of race r, interacts with individuals of race s. This is equal to:

$$\mu_{irs} = \Sigma_j \mu_{irjs} = \Sigma_j e^{\overline{\alpha}_{irjs}} \hat{\mu}_{irjs} \tag{6}$$

Write the average utility generated by an interaction in the following way:

$$\overline{\alpha}_{irjs} = \alpha_{rs} + \varepsilon_{irs} + \eta_{irjs} \tag{7}$$

The term α_{rs} represents the average utility created by an interaction between someone of race r and someone of race s within the city. The term ε_{irs} represents the deviation from this average for individuals in neighborhood i (for all j). This term would be negative, for example, if race-r individuals in neighborhood i were unusually racist against individuals of race s. The terms η_{irjs} represents the additional deviation when someone of race r from neighborhood i interacts with someone of race s from neighborhood j. This will reflect any specific compatibility between these two groups. Note that both the terms ε_{irs} and η_{irjs} arise because of residential sorting, a process that creates systematic differences in observable and unobservable characteristics between individuals living in different neighborhoods. To the extent that these characteristics affect the utility from a

[13] Note that the terms $\hat{\mu}_{irjs}$ are *not* equilibrium outcomes. If we eliminated preferences over partner characteristics, the overall interaction rate (and therefore the terms μ_{ir0}) would change, which is not captured in my construction of $\hat{\mu}_{irjs}$. We can instead think of $\hat{\mu}_{irjs}$ as the initial non-equilibrium result of eliminating preferences.

match, we expect that the term $\overline{\alpha}_{irjs}$ will vary depending on the location of the two interaction partners (i.e. that both ε_{irs} and η_{irjs} will typically not be zero.) I assume that both ε_{irs} and η_{irjs} are normally distributed, although not in a way that is independent of residential location.

Using this equation in the equilibrium condition and taking logs gives:

$$ln(\mu_{irs}) = \alpha_{rs} + \varepsilon_{irs} + ln(\Sigma_j e^{\eta_{irjs}} \hat{\mu}_{irjs}) \tag{8}$$

Without loss of generality, rewrite the summation in the logarithm term on the right hand side, so that the equation becomes:

$$ln(\mu_{irs}) = \alpha_{rs} + \varepsilon_{irs} + ln(\Sigma_j \hat{\mu}_{irjs} + D) \tag{9}$$

The term D here is simply the difference between $\Sigma_j e^{\eta_{irjs}} \hat{\mu}_{irjs}$ and $\Sigma_j \hat{\mu}_{irjs}$; it captures additional interactions between members of (i, r) that take place with members of race s because of preferences and the associated residential sorting.[14]

Equation 9 guides my estimation of the average racial preference parameters α^{rs}. For this estimation, I start with data on the residential location and inter-racial interaction frequency for a set of individuals. For each individual h in my dataset, I construct the terms μ_{hirs} and $\Sigma_j \hat{\mu}_{irjs}$ and run the following non-linear regression across individuals within a city c, separately for each partner race s:

$$ln(\mu_{hirs}) = \alpha^c_{rs} + ln(\Sigma_j \hat{\mu}_{irjs} + D^c_{rs}) + \varepsilon_{hirs} \tag{10}$$

This regression relates the interaction frequency for individual h, of race r, living in neighborhood i, with members of race s, to the predicted interaction frequency if we eliminated preferences over interaction partners ($\Sigma_j \hat{\mu}_{irjs}$). The term α_{rs} captures preferences over the average person of race s in the city. The term $\Sigma_j \hat{\mu}_{irjs}$, as discussed earlier, is the predicted frequency of s-interactions that would take place based on the effect of physical distance and population supplies alone. The term D reflects the deviation from this level of inter-racial interactions that occurs because individuals have preferences over non-race characteristics that may vary by neighborhood. In other words,

[14] To see this, note that in expectation, D will be equal to $\overline{\hat{\mu}_{irjs}}(e^{\frac{1}{2}\theta^2} - 1) + cov(\hat{\mu}_{irsj}, \eta_{irsj})$, where θ is the variance of η_{irsj}. θ captures the degree to which preferences for partners vary strongly across neighborhoods, while $cov(\hat{\mu}_{irsj}, \eta_{irsj})$ captures the extent to which individuals prefer their near neighbors. Both of these terms will tend to be more positive in the presence of residential sorting, causing $\Sigma_j e^{\eta_{irjs}} \hat{\mu}_{irjs}$ to exceed $\Sigma_j \hat{\mu}_{irjs}$.

even without direct preferences over race ($\alpha_{rs} = 0$), we would expect μ_{irs} to differ from $\Sigma\hat{\mu}_{irjs}$. This is because individuals can sort into neighborhoods where they like the particular partners of race s relatively better; this will tend to push the inter-racial interaction frequency above the level predicted by $\hat{\mu}_{irjs}$ alone. The entire term $\Sigma_j\hat{\mu}_{irjs} + D$ captures the inter-racial interaction frequency we would predict based on this process. The non-linear estimation process chooses α_{rs} and D to minimize the sum of squared error terms ε_{hirs}.

The constant term α_{rs} from this regression, which I interpret as the average utility created by an r, s interaction, captures the extent to which the interaction frequency between r and s in the city deviates from the level predicted by the term within the logarithm, while ε_{hirs} (the error term in the regression) captures the individual's deviation from this average. With sufficient data, I could break ε_{hirs} into a neighborhood by race fixed effect (which I would interpret as ε_{irs} from the model) and an individual error term. In practice, I will often only have one observation per neighborhood, which precludes the use of neighborhood-specific fixed effects. Note, however, that the neighborhood-specific variation will not be important when I move to the next steps of my simulation exercise.

It is important to emphasize that the goal of this regression is not to estimate any causal relationship. The causal effect of physical distance in my model is already incorporated in the term $\delta d(i, j)$, which appears in $\hat{\mu}_{irjs}$. Instead, this regression is intended to decompose the frequency of inter-racial interactions into a piece that can be explained by this causal effect and into a residual that I interpret as reflecting racial preferences. Of course, the term $\hat{\mu}_{irjs}$ captures only one particular channel through which neighborhoods influence interactions. If this is true, deviations from the predicted level of inter-racial interactions may result from some of these other causal channels. This implies that my "preference" terms $\hat{\alpha}_{rs}$ may actually overstate the extent of racial preferences. This means that my estimates should be interpreted as lower bounds on the impact of residential desegregation.

3.2 Step 2

The goal of this simulation exercise is simulate the frequency of inter-racial interactions that would take place if we eliminated residential segregation completely. The specific counter-factual I use is one where individuals of different races are evenly distributed throughout the city, with each individual being randomly assigned to a specific neighborhood.

To achieve this distribution of individuals, I assume that every neighborhood has the same total population as it does currently, but that the fraction of individuals of race r is equal to the proportion of these individuals in the *city* population for every tract. Random assignment of specific individuals implies that each neighborhood will, on average, contain a representative sample of individuals of each race.[15] I use this fact in the derivation of my simulation results below.

3.3 Step 3

The third step of my exercise is to calculate the inter-racial interaction frequencies that would take place if individuals had the same preferences estimated in the first step of the exercise, but had the geographic distribution simulated in the second step. To do this, I use the equilibrium equation from my model. Letting N denote the set of neighborhoods in a city, this is:

$$ln\left(\frac{\mu_{irjs}}{\sqrt{(f^{*r}(i)-\Sigma_{t\in R}\Sigma_{k\in N}\mu_{irkt})(f^{*s}(j)-\Sigma_{t\in R}\Sigma_{k\in N}\mu_{ktjs})}}\right)=\hat{\alpha}_{rs}-\delta d(i,j)$$

In this equation, $f^{*r}(i)$ is the new number of individuals living in neighborhood i of race r, after the redistribution. The preference term on the right hand side is the average match value for an r,s pair, as estimated in step 1 of the simulation. Importantly, the match values do not vary systematically across neighborhood pairs the way they do in real life. That is, the terms ε_{irs} and η_{irjs} will both be equal to zero. This is because we have eliminated any sorting into neighborhoods on the basis of observable or unobservable characteristics. The average value of a match between a person of race r in neighborhood i and a person of race s in neighborhood j will be the same, no matter which i and j we choose.[16] Using my estimates $\hat{\alpha}_{rs}$, along with calculations of $\delta d(i,j)$ and the simulated population supplies $f^*()$, I can solve this system of equations for μ_{irjs}. These can then be aggregated into predictions about both the overall interaction rate, and the inter-racial interaction rate.

[15] In practice, random assignment will not result in an exactly even distribution of interaction-relevant characteristics across neighborhoods. It is possible for my model to accommodate this random variation. To do this, I can allow my estimated α_{rs} to vary with the observable characteristics of each individual (e.g. education, age) and the average level of these characteristics among individuals of race s in the city; this provides an approximation of how interaction values are affected by these characteristics. Then, I could randomly re-assign individuals and construct a predicted α_{ijrs} for each neighborhood-race pair, based on the characteristics of each neighborhood. This procedure should not affect my results on average, however; as a result, I do not implement it here.

[16] Of course, the match value of specific individuals in these neighborhoods may be higher or lower than this; this is captured by the error terms ϵ_{hgirjs} in each individual's utility function.

3.4 Supplementary exercise

In a secondary exercise, I attempt to provide an intuitive sense of how important my estimated disutility of travel terms δ are likely to be in influencing individuals' interaction decisions. To do this, I use a version of the model that assumes i) that individuals remain at their existing residential locations, but ii) that physical distance is the *only* factor affecting decisions about whether to interact and with whom. Specifically, the equilibrium condition I use is:

$$ln\left(\frac{\mu_{ij}}{\sqrt{(f(i)-\Sigma_k\mu_{ik})(f(j)-\Sigma_l\mu_{lj})}}\right) = -\delta d(i,j) \tag{11}$$

A key difference between this equation and Equation 3 is that the equation no longer has the subscripts r and s indicating the race of the interaction partners. If individuals care only about physical distance, then they should draw randomly from the population living in each neighborhood. Nonetheless, there will be some social segregation induced by individuals' behavior in this model. This is because black and non-black people are not randomly distributed across cities; non-black people will tend to have more non-black people in nearby neighborhoods, and conversely for black people.

This equation says that the log number of interactions between neighborhoods i and j, once scaled to reflect population supplies, should be fall with the distance between the two neighborhoods, at a rate equal to the disutility of travel. As noted above, given a value for δ and $d(i,j)$ for every neighborhood pair as well as a vector of population supplies $f()$, I can solve this equation for the unique vector μ_{ij} that satisfies this set of equations.

To see how the equilibrium terms μ_{ij} are aggregated into inter-racial interactions, consider interaction behavior for individuals living in neighborhood i (of any race). The number of interactions between people in i and people of race s in neighborhood j will be equal to:

$$\mu_{ijs} = \frac{f^s(j)}{\Sigma_z f^z(j)}\mu_{ij}$$

That is, the proportion of individuals of race s make up a proportion p of the total population in a neighborhood, then p will also represent the proportion of μ_{ij} interactions that involve with a partner of race s in neighborhood j. By adding this up over all neighborhoods j, we can calculate the total number of interactions that occur between individuals in neighborhood i and individuals

of race s. If we divide this by the total number of interactions for individuals in neighborhood i, we have the proportion of interactions for neighborhood i that occur with race s. Note that this will be the same for individuals of *any* race who live in i; individuals who live in the same neighborhood will behave the same way according to this model. However, because of residential segregation, individuals of race r will tend to be heavily represented in neighborhoods that are physically close to large numbers of race r partners. Individuals living in these neighborhoods will therefore have a disproportionately high number of r partners, compared to the city population. This is what will cause social segregation in the model.

We can get a sense of how important the distaste for travel is by comparing the the inter-racial interaction rate generated by this model to two objects. The first is the random interaction rate. This is the rate that would occur if individuals matched randomly within cities. For example, for a typical non-black person in a U.S. city, the random black interaction rate would be around 13% - the population frequency of black people in an average U.S. city. If physical distance alone is able to generate a significant reductions in inter-racial interactions relative to the random rate, then we can infer that the distaste for travel is "large" in an absolute sense.

The second object with which we may wish to compare the results of the simulation is an index of existing social segregation. The index I use is simply the absolute difference between the random inter-racial interaction rate and the actual inter-racial interaction rate. As I show below, the typical non-black person in my data appears to have about 3% of their interactions with black people. With random matching, this would be around 13%, meaning that there is approximately a 10 percentage point gap between how individuals would behave in a perfectly integrated world (i.e., no racial preferences or physical segregation) and how they actually behave. In a similar way, I can construct a measure of how much social segregation there would be in the simulation exercise I describe above. By comparing simulated to actual social segregation, we can get a sense of how large the distaste for travel is, *relative* to other factors driving social segregation.

4 Data

To perform my simulation, I need four pieces of information. First, I need to know the existing population of each neighborhood, $f^r(l_i)$, separately by race. Secondly, I need to know the geographic distance between neighborhoods within a city; this corresponds to $d(l_i, l_j)$ in my model. Third, I

need an estimate of the disutility of travel δ. Finally, I need information on the social interaction behavior of Americans: both how often individuals in different neighborhoods socialize (which, combined with information on population will allow me to estimate the terms μ_{ir0} from my model), and how frequently they socialize with members of different races (which corresponds to the term μ_{irs} in my model). In this section, I describe where I get each of these piece of information.

4.1 Population distribution

Information on the population distribution by neighborhood and geographic distances are available from the U.S. Census Bureau. Throughout the analysis, I will define a neighborhood as a Census tract. The average population in a Census tract in the U.S. is around 4,000 individuals. I restrict the analysis to pairs of Census tracts within the same Core-Based Statistical Area (CBSA.)[17] There are 933 CBSAs in the United States (excluding Puerto Rico), of which I use 202 in my main analysis.[18] These CBSAs account for just over 80% of the U.S. population. The mean number of Census tracts one of these CBSAs is 271, ranging from 31 to 4,701.

Different measures have been used to capture the degree of racial segregation within cities. One popular measure is the Duncan index, which measures the fraction of black or non-black residents within a city that would have to move to produce an even distribution of racial groups over Census tracts.[19] The first column of Table 1 shows that the mean Duncan index in my sample is 0.521, indicating that about half of all residents in a typical city would have to move to achieve perfect integration. The next rows show how the Duncan index varies across the four Census regions. According to this measure, segregation is highest in the Midwest, with an average Duncan index of 0.596, and lowest in the Pacific, with an average Duncan index of 0.457.

Echenique and Fryer (2007) construct an alternative measure of residential segregation that is intended to closely capture the probability of social interactions across different Census blocks. This measure is calculated based on information on the black population of each Census block, the black

[17] CBSAs consist of "one or more counties and includes the counties containing the core urban area, as well as any adjacent counties that have a high degree of social and economic integration (as measured by commuting to work) with the urban core." (Census Bureau, 2016.)

[18] The restrictions that lead to the reduced sample of CBSAs are that I must have an estimate of the disutility of travel, which requires information on both travel speeds and hourly wages; and that I must have at Flickr users represented in at least 25 tracts. The latter restriction results in more eliminations from my pool of CBSAs, but is required in order to implement the regressions in Step 1 of Simulation B.

[19] The Duncan is calculated using the formula $D_c = \Sigma_t |\frac{\text{Black}_{tc}}{\text{Black}_c} - \frac{\text{Nonblack}_{tc}}{\text{Nonblack}_c}|$ where Black_{tc} is the number of black individuals living in tract t in city c, and Black_c is the total number of black individuals in the city (and similarly for non-black individuals).

population of neighboring Census blocks, and those neighbors' degree of segregation, implying that the entire distribution of the city population is incorporated into the final index. The authors show that this index satisfies several desirable properties, including that it is invariant to how the boundaries of neighborhoods are drawn (unlike, for example, the Duncan index.) Their measure of black segregation at the MSA level is provided on the authors' websites.[20] Column (2) of Table 1 shows the mean level of this index (which also varies between 0 and 1) for the CBSAs in my sample, and how it varies across Census regions. Throughout the paper, I refer to this index as the SSI. While the mean level of the SSI is similar to the Duncan at 0.577, this index tells a markedly different story across regions. It shows that there is a much more substantial degree of black segregation in the South and Midwest (with both indices around 0.7) than in the Northeast (around 0.5), and a much lower degree in the Pacific region (around 0.25). As I describe in the results section, the results from my secondary simulation exercise coincide much more closely with the SSI than with the pattern shown in the Duncan index.

4.2 Distance

To measure the geographic distance between pairs of Census tracts, I use shapefiles provided by the U.S. Census Bureau. I calculate the great-circle distance between the central latitude and longitude of each pair of tracts within a CBSA. The third column of Table 1 shows that, on average, two randomly selected tracts within the same CBSA are 41.1 km apart. This varies from 37.5 km in the Pacific region to 44.9 km in the Northeast.

Table 2 shows the mean distance to an average black and non-black person within the same CBSA, for black and non-black individuals separately.[21] This is somewhat lower than the average distance between tracts (not population weighted), with estimates ranging between 23 and 28 km. Somewhat surprisingly, the average non-black person lives *closer* to the average black person than they do to other non-black people. This can occur when the black population is concentrated in the city center, which is a pattern that holds in many U.S. cities. Note, however, that this does not imply that distance doesn't play a role in generating social segregation. While I model the disutility of distance as being linear, my model can produce highly non-linear relationships between distance and the probability of interactions (even without any interaction-specific preferences.) In this case,

[20] Specifically, the data were available from from http://www.its.caltech.edu/ fede/segregation/ as of February 2019.

[21] Throughout the paper, I use the racial categories "black" and "non-black". This is driven my Flickr photographs data, in which I can distinguish black from non-black people but not different categories of non-black people.

it will not be the distance to the average black person that matters for interaction behavior, but the frequency of black people in a person's immediate neighborhood.

4.3 Disutility of travel

A number of papers have estimated individuals' disutility of travel in the context of estimating demand for movie theatres (Davis, 2006; Thomadsen, 2005), liquor stores (Seim and Waldfogel, 2013), coffee shops (McManus, 2007), and gas stations ((Manuszak and Moul, 2009; Houde, 2012). The typical strategy of these papers is to examine how much consumers are willing to pay, in terms of price, to avoid extra travel to a location that is further away. A key assumption for identifying the distaste for travel in this way is that consumers otherwise value the competing locations similarly; that is, that there is no correlation between a location's distance from the consumer and its unobservable characteristics. In some cases, such as for gas stations near a consumer's commute path, this seems reasonable. In other cases where the assumption is more tenuous, a variety of instruments have been used to try and causally identify the effect of distance.[22]

Table 3 summarizes the findings of these papers. The estimated willingness to pay to avoid a minute of travel varies quite substantially in this literature, both in absolute magnitude (ranging from about \$0.10-\$0.57 per minute in 2002 dollars) and in relation to average hourly wages (with the hourly valuation ranging from 0.5-2.5 times the average hourly wage.) Broadly speaking, however, the results can be grouped into two sets: one set implying a valuation of time at about the average hourly wage (Davis, 2006; McManus, 2007; Manuszak and Moul, 2009), and another set implying a time valuation of about twice the average hourly wage (Thomadsen, 2005; Houde, 2012; Seim and Waldfogel, 2013). I show estimates using both of these values in my simulation results.

To construct an estimate of the disutility of travel for each city in my sample, I start by estimating the hourly wage for each city. Information on hourly wages is available for some metropolitan areas from the Bureau of Labor Statistics. In order to preserve the majority of CBSAs in my sample, however, I instead impute average hourly wages by using information on state-level hourly wages and the ratio of median income in the CBSA to median income in the state. For each city, I construct a "per minute" disutility of travel equal to either one or two times the wage per minute in that city.

Next, I convert the dollar valuation of time to utility terms using the estimates from Houde

[22]For example, Manuszak and Moul (2009) use a tax hike near Cook County to estimate consumer's willingness to pay to travel across county lines to purchase gasoline.

(2012), which are available in both units of measure. This gives me an estimate of disutility per minute, which I then convert into "per kilometer" format by using information on travel speeds from the Google Maps API.[23] I have sufficient information on income and travel speeds to calculate the disutility of travel for 862 CBSAs, which include all 202 CBSAs in my main analysis sample. As shown in the fourth column of Table 1, the mean disutility of travel across cities is around 0.458 per kilometer, which corresponds to a dollar valuation of around $0.46 per kilometer in 2010 dollars. The table also shows how the disutility of travel varies by region. The cost of traveling 1 km is highest in the Northeast and Pacific, and lowest in the South.

A limitation of this procedure is that variation across cities is imposed by assumption, not by revealed behavior. We can get some sense of whether the implied distaste for travel actually corresponds to individuals' travel behavior by using the travel patterns in my Flickr data. As I explain in the next section, the main purpose of my Flickr data is to measure individuals' cross-racial interactions. Because the photos are geotagged, however, they also provide some information about how individuals move throughout their home cities. Table 4 shows the relationship between my predicted disutility of travel at the city level and the fraction of photographs that are taken within 1, 3, 5 and 10 km of a Flickr user's estimated home location.[24] The table shows that cities with a higher estimated distaste for travel have a higher proportion of photos taken close to home. For example, the coefficient on δ for photos taken within 1 km of home is 0.158 and is significant at the 10% level. This implies that 1-standard deviation increase in δ (approximately 0.105) is associated with a 1.7 percentage point increase in the proportion of photos taken within 1 km of home, a 10% increase over the baseline of approximately 15%. The coefficients for photos taken within 3, 5, or 10 km of home are also positive, and are significant at the 1-5% level. [25]

[23]Specifically, I choose 10 randomly selected pairs of Census blocks within a CBSA and query the API for a driving time between them on a Saturday afternoon at 3 pm. In my main results, I use Flickr photos taken on weekends to measure social interactions, in order to capture time periods when individuals are likely to be leaving from home.

[24]I describe how I infer users' home locations, and provide evidence that I am correctly identifying these locations, in the next section.

[25]In principle, disutility of travel may vary at the tract level. This may occur because individuals value their time differently, or because travel speeds differ across tracts. In an earlier version of this paper (Cornelson, 2017), I attempt to estimate tract-level disutilities of travel using information on travel patterns from Flickr. While these parameters did a better job of predicting Flickr users' travel behavior, they made essentially no difference to the interaction results. For this reason, I maintain the assumption of a city-level disutility of travel here.

4.4 Social interactions

My simulation exercise requires that I have two pieces of information about social interactions: the overall interaction rate (which, along with population estimates, allows me to calculate the terms μ_{ir0} in my model), and the frequency of inter-racial interactions (the term μ_{irs}), by race and by neighborhood. Unfortunately, this information is not available in any large, publicly available dataset. The data used in earlier research on social interactions includes the Add Health dataset (a survey of teenagers; e.g., Echenique and Fryer, 2007), the Social Capital Community Benchmark Survey (a survey of individuals living in cities that asks respondents how often they participate in different social activities; e.g., Brueckner and Largey, 2008) and the DDB Needham Lifestyle Survey (a survey that asks similar questions as the SCCBS; e.g., Glaeser and Gottlieb, 2006). Of these, only the Add Health has information on either residential location or cross-racial interactions; however, both pieces of information are available only for a relatively small subsample of respondents.[26] Additionally, this data source only allows us to measure interaction behavior for a very specific group (teenagers), nearly two decades ago.

To measure interaction behavior, I instead rely on a combination of the American Time Use Survey (ATUS) and a novel dataset I have constructed using Flickr photographs. The ATUS is an annual survey of a representative sample of Americans that asks respondents to keep a diary recording what they are doing and who they are with for every 15 minute segment of the day. This can be used to construct an estimate of interaction rates. Unfortunately, the ATUS does not contain geographic information below the state level. To arrive at ATUS estimates for tracts, I use the demographic information present in the ATUS to predict the interaction rate based on tract demographics. I show that this measure predicts the interaction rate well in an independent survey.

While the ATUS is informative about overall interaction behavior, it contains no information about the inter-racial interaction rate. To measure inter-racial interaction behavior, I turn to my Flickr dataset. Flickr is a popular photo-sharing website. As of 2013, the site had around 87 million users uploading approximately 3.5 million photos per day (Jeffries, 2013). I downloaded a large sample of public photographs on Flickr, along with their metadata, and ran them through face detection and race classification algorithms. The racial breakdown of faces in the photographs will provide me with information on individuals' inter-racial interaction rates. The metadata, which

[26] Patacchini, Picard and Zenou (2015) examine the relationship between physical distance and the probability of friendship in the Add Health data, using a sample of about 1500 respondents that have sufficient information on both residential location and social interactions.

contain information such as the make of the camera and the time of the photograph, sometimes include "geotags", which are latitude and longitude coordinates appended by cameras that have access to the internet (smart phones, for example, and higher-end digital cameras.) These geotags allow me to link Flickr users to cities and neighborhoods.

To validate the use of Flickr data to measure inter-racial interaction behavior, I also ran a survey of MTurk workers about their use of social media and their interaction behavior. I show that the racial breakdown of faces in an individual's photographs is an excellent indicator of the racial breakdown of that person's friends. In fact, the faces in a *single* photograph explains approximately 39% of the variation in actual interaction behavior, with the fraction of black faces demonstrating an almost one-for-one relationship with the fraction of the individual's friends that are black.

In the remainder of this section, I provide more information on how I construct measures of interaction behavior from these sources of data, including the construction of the Flickr dataset and the MTurk survey.

4.4.1 Interaction rates

I use the American Time Use Survey (ATUS) to arrive at my estimates of the overall interaction rate for each tract/race pair. My measure of interaction rates in the ATUS will be the probability that a respondent spends any time with friends on his or her diary day. This measure corresponds closely to the decision margin in my theoretical model. The mean of this variable is 22.7% for both races, indicating that about one-fifth of the population spends time with friends on a randomly selected day. This varies somewhat across different times of the week, with an average of 20.6% on weekdays and 24.5% on weekends. Because I will be focusing on the impact of desegregating residential locations, which are more likely to impact social interactions on weekends, I will focus on weekend socializing in both the ATUS and Flickr results.

Table 5 shows how this probability of weekend socializing varies with age, education, and geographic region, separately for black and non-black respondents. For both racial groups, the interaction rate shows a U-shape in age; the coefficients on age and age squared indicate that interactions decline with age until approximately age 58-61, before starting to rise again. For the non-black sample, the interaction rate shows an approximately linear and increasing pattern in education. For the black sample, interaction rates are similar for all educational groups except for those with post-graduate degrees, who have higher interaction rates. There are no regional differences in interaction

behavior for blacks, and only one marginally significant different for whites (with slightly higher interaction rates in the Midwest, compared to other regions of the country.)

I use the results from the model shown in Table 5 to predict the interaction rate for each Census tract based on tract demographics, separately for the black and non-black samples. My predicted interaction rate varies from 18.5%-50.1% for blacks (with a mean of 25.3%), and from 13.7% to 65.2% for whites (with a mean of 24.5%). Because the interaction rate is based on demographic characteristics available in the American Community Survey, I am able to predict it for nearly all of the approximately 70,000 tracts in the U.S.

To validate the use of my predicted interaction rates, I compare the ATUS predictions of interaction rates to actual interaction rates in a survey of MTurk workers. The survey was administered to approximately 1,600 MTurk workers in the summer of 2018. To be included in the survey, a worker had to be living in the United States, and must have posted at least 1 photograph to social media over the past year. This latter restriction was imposed because I will also use the survey to validate the use of my Flickr measure of inter-racial interactions. The survey contained modules asking the workers about basic demographic information; their time use over the previous day, including information on who was present at each moment and the race of those individuals; and their social media use. The demographic module was always presented first, while the other three modules were presented in random order.

Table 6 shows demographic information on the MTurk sample, compared to the U.S. adult population. As might be expected, MTurk workers are not representative of the U.S. population at large: they are significantly younger, more highly educated, and more likely to be white or Asian than other U.S. adults. This is in line with previous work that examines the characteristics of MTurk workers (Berinsky, Huber and Lenz, 2012; Huff and Tingley, 2015). To the extent that these characteristics are reflected in the MTurk workers' residential locations, however, the predicted interaction rates should still be valid for this sample.

I construct a measure of the social interaction rate in my sample using the time use portion of the survey. This module asked respondents what their primary activity was during each 3 hour period of the previous day, and who was with them during that time. I constructed the questions in the time use module to be as similar as possible to the American Time Use Survey questions, using similar question wording, and the same breakdown of activities present in the ATUS lexicon. As in the ATUS, I measure the interaction rate by constructing an indicator for whether a respondent

spent any time with friends on the day in question. 25.8% of the workers in my survey spent time with friends the previous day, which is quite similar to the ATUS mean.[27]

MTurk surveys automatically include information on the respondents' latitude and longitude while taking the survey. Assuming that this location will typically be the user's home, I use these locations to connect respondents to a predicted tract-level interaction rate constructed from the ATUS data. Then, I examine whether the ATUS prediction corresponds well with their actual interaction behavior. Unfortunately, I have too few black respondents (113) to be able to examine the relationship between these variables in the black sample; I therefore restrict myself to the non-black sample in this exercise.

A regression of the MTurk worker's interaction indicator on his or her predicted interaction rate produces a coefficient of 0.787, which is significant at the 1% level. This coefficient is not statistically distinguishable from 1, which is the coefficient I would expect in this regression. The constant is 0.046 and is not significantly different from 0. When combined with the fact that there is a great deal of error in my "home" assignment processes (which should bias the coefficient in this regression downward), this regression suggests that the estimated interaction rates do a good job of predicting actual interaction behavior.

4.4.2 Inter-racial interactions

To measure the inter-racial interaction rate, I rely on a new dataset I have created using Flickr photographs. In this dataset, I observe the racial breakdown of the individuals in a Flickr user's photos; I use this as my measure of the relative frequency of black interactions. I show below that this measure is quite strongly correlated with actual inter-racial interaction behavior in my MTurk survey. The Flickr photographs also contain geotags, which allows me to infer a home location for each user. While there will doubtless be a great deal of error in the assignment process, I provide evidence below that I am identifying the home neighborhood correctly on average.

To build this dataset, I began by identifying a set of around 170 million geotagged Flickr photographs, all taken within the U.S. between 2006-2015. To do this, I started by pulling a random sample of about 10% of all geotagged photographs taken in the U.S. over this period. Then, I pulled every photograph ever taken by the approximately 365,000 users in this initial sample. In order to

[27]The survey was run on a Sunday-Tuesday, with the majority of respondents answering on a Monday. Note that the questions referred to the previous day, which means I am capturing interaction behavior primarily on Sundays.

remain in the sample, a Flickr user had to i) take the majority of their photographs in the U.S., ii) post photos taken on at least 3 separate days within a single year, for at least one year, and iii) have at least one face in the sample of photos that I use. The second restriction is required in order to infer a home location for each user. The third restriction is required in order to infer something about the user's inter-racial interaction behavior.

I link users to home locations by assigning them to the modal CBSA in which they take pictures, and to their modal Census tract within the CBSA. I do this separately for every year in which a Flickr user posts photographs. In order to abstract from potential moves by Flickr users (some of which may be induced by error in the home assignment process), I assign Flickr users consistently to the home location from the year on which the user posted on the maximal number of days. I also keep only photographs that are taken in a user's home city. I additionally restrict analysis to the 202 CBSAs that contain at least 1 Flickr user in at least 25 separate tracts. This restriction is required in order to run the regression in Step 1 of Simulation B, which identifies the preference parameters α using cross-tract variation. My final sample of Flickr users is comprised of around 87,000 users who post around 18 million photographs in their home cities.

Appendix Tables A1, A2 and A3 provide evidence that I have correctly identified users' home locations. Table A1 shows that the home tracts are visited far more often than any other tract. The table shows the number of unique "visits" (day by tract level observations) to the home location and to other Census tracts the user visits. The average user is observed in her assigned home Census tract on 14 separate days; for any other Census tract that the user visits at least once, the mean number of visits is around 3. For a typical Flickr user, about 43% of all visits are in the home tract.

In Table A2, I show that the surroundings in the home tract are observably different from other tracts the user visits. The table shows the types of venues that appear in the home location and in other visited tracts, using information from the Foursquare database. Foursquare is a service that allows individuals to "check-in" at different locations, providing information to friends and family about where they are. Foursquare maintains a database of venues, which is searchable by latitude and longitude. For a sample of around 20,000 owners, I randomly select one photograph taken in their home tract and one photograph taken outside of their home tract, and search the Foursquare database for venues within a 25 meter radius around the location where the photograph was taken.[28] I divide venues into five categories: food and drink (e.g., restaurants, bars, coffee

[28]The Foursquare API maintains rate limits which limit the number of searches to their database each day. This is

shops), entertainment (e.g., parks, movie theaters, art galleries), stores, offices, and "other". The latter category is mainly comprised of other commercial buildings that are not designated specifically as office buildings; among the most common types of venues in this category are banks, doctor's offices, and barbers/salons. I compare the number of venues I find of each type when the user is in his or her assigned home tract and when her or she is elsewhere. There are fewer venues of all type near the user when she is in her home tract, and the difference is highly significant for four of the five venue types. When combined with the results for visits, these results show that Flickr users spend substantially more time in their "home" tracts, even though there are fewer commercial venues to visit in these areas.

Finally, in Table A3, I use the one piece of information I have on Flickr users - their names - to examine how city and tract demographics correlate with the user's demographics. For users with last names on their profiles, I use information on the racial distribution of the 1000 most common last names in 2010 (Census Bureau, 2010) to construct a probability that the user is white, black, Hispanic or Asian. I then examine how this probability predicts the proportion of people that are of the same race in a user's CBSA; in the other tracts she visits (aside from the home tract); and in her home tract. Table A3 shows the results for each race in separate panels. The table shows that home tract demographics are more strongly correlated with a user's race than the demographics of other tracts she visits, or than the demographics of the city as a whole. Increasing the probability that a user is white by 1 percentage point increases the proportion white in her CBSA by 0.073 percentage points; the proportion white in visited tracts by 0.099 percentage points; and the proportion white in her assigned home tract by 0.107 percentage points. The results are similar for other races. This provides further evidence that user's assigned home tracts are likely to be strongly correlated with their actual neighborhood of residence.

The information in Table A3 suggests that we can use the characteristics of a Flickr user's assigned home tract as a proxy for their own characteristics. This provides a way to examine the demographic characteristics of the Flickr sample and how these differ from those of the typical American. Table 7 shows the average characteristics of Flickr users' CBSAs/tracts, and compares these to those of an average population living in one of the CBSAs in my sample, and to the population of the entire U.S. The Flickr users are more concentrated in larger cities, with an average city size of around 5.3 million, compared to 4.7 million for a typical resident of the same cities. This implies that Flickr

why I use a smaller sample of owners and photographs.

users are disproportionately concentrated within the larger cities in my sample. My sample cities are much larger than those inhabited by an average American, which have an average size of 3.8 million. City-level segregation is quite similar for Flickr users and other Americans. However, the two groups live in different areas within cities: Flickr users are concentrated in wealthier and more educated tracts, compared to both other residents of the same cities and the U.S. population in general. Flickr users have a slightly lower proportion of whites, Blacks and Hispanics in their tracts than the typical American (particularly for the latter two ethnic groups) but a higher proportion of Asians. While Flickr users are not representative of the U.S. population, I will attempt to use information from the Flickr sample to predict interaction rates outside of the sample. This process is described in detail below, after I explain my measures of social interactions.

Once I have linked users to home locations, I measure their inter-racial interactions by running their photographs through face detection and race classification software. The face detection algorithm was provided by MIT Information Extraction. Kazemi and Sullivan (2014) report that it has a 95% accuracy rate, with most of the error accounted for by false negatives. However, the rate of false negatives appears to be substantially higher in the Flickr data. In a sample of around 17,500 photographs which I hand-coded, I found 13,560 faces in total, while the algorithm found just 6,861. The lower accuracy rate may be due to the fact that the Flickr photographs are often of relatively low quality, and include many faces that are partially turned away from the camera. The rate of false positives also appears to be elevated relative to that reported in Kazemi and Sullivan (2014), at around 8%; this is due primarily to the fact that Flickr users often take pictures of statues, art with faces, and animal faces. Other non-face objects are rarely identified as faces. 18.6% of the photos in my database are found by the algorithm to have any faces in them. Based on the error rates cited above, I estimate that the actual frequency is about twice as large, in the 35-40% range.

I next run all photographs with faces in them through a race classification algorithm, which classified each face as either black or non-black. The algorithm itself was provided as part of the Scikit Learn machine learning module for Python. I trained the classifier using the faces in a random sample of 20,000 Flickr photographs. Table 8 shows the "confusion matrix" from the testing process, which shows the fraction of non-black/black faces that are categorized as non-black or black. Non-black faces are correctly categorized 87% of the time, while black faces are correctly categorized 75% of the time. These error rates are in line with the standards in the literature for this type of

classification task (Han and Jain, 2014).)[29] While the classifier is correct in the large majority of cases, the error in the assignment process does create a problem for estimating the fraction of black faces in the sample. This is because most of the faces in Flickr photographs are non-black: in a sample of hand-coded photographs, I estimated the fraction of black faces to be around 5%. As a result, the 13% of non-black faces that are classified as black are numerically a far larger set than the 25% of black faces that are misclassified. This results in a much higher estimated frequency of black faces - around 21% - than is actually the case.

To correct for this error, I fit a model linking the proportion of black faces found by the classifier to an actual frequency of black faces according to a set of hand-coded photographs. I first divided Flickr users into 100 groups based on the fraction of black faces found by the classifier (0-1%, 1-2%, etc.) I sampled up to 5 users from each percentile range (not all ranges contained 5 users), then downloaded a random sample of 50 photographs with faces for each of these users. I hand-coded the photographs to arrive at an actual frequency of black faces for each user.

The rate of black faces in this sample is very low. The mean number of black faces is around 5%, and just 14 users have a majority of black faces in their photographs. This is in spite of the fact that users with a high proportion of black faces found by the algorithm were oversampled when I selected photographs to hand-code. It appears then that black individuals are highly under-represented among Flickr users. For this reason, I will assume going forward that my Flickr sample is entirely non-black. As I discuss further in the next section, this does not interfere with my simulation procedure. Every cross-racial interaction for a non-black person represents a cross-racial interaction for a black person as well. I assume that these interactions are generated by the preferences of both interaction partners, and attempt to estimate a parameter capturing this joint surplus.

Figure 1 shows the plot of actual black faces in the hand-coded sample against the number of black faces found by the classifier, with users grouped into 2.5-percent ranges (0-2.5%, 2.5-5%, etc) based on the algorithm's classification. There is an upward sloping and non-linear relationship between the proportion of black faces found by the detector and the user's actual proportion of black faces. Even for users with a very high proportion of black faces found by the algorithm, the actual fraction of black faces is relatively low, with a maximum of just over 30% for users who are found to have 95% of black faces by the algorithm. This supports my assumption that the vast majority

[29] Accuracy rates are much higher in "constrained" classification tasks, where pose and illumination are constant across subjects.

of users in my sample are non-black. The line on the graph shows the fitted relationship, which I estimate using a linear and quadratic term. I use this relationship to predict the proportion of black faces for each Flickr user, based on the algorithm's results. The mean fraction of black faces in the broader sample is around 3.7%.

Figure 2 shows how the fraction of black faces found in a user's photographs is related to the percentage of black people in the user's assigned home tract. The relationship between these two variables (shown by the red line) is positive and significant, indicating that individuals living in tracts with more black people have a higher frequency of black interaction. Note, however, that the relationship is quantitatively quite small. For individuals assigned to black-majority tracts, black faces make up an average of 4.2% of all faces in photographs, compared to 3.4% for individuals assigned to black minority tracts. Individuals in all tracts are predicted to have a proportion of black faces under 6%. As indicated by the size of the circles in the graph, the number of individuals living in highly black tracts is quite small; the vast majority of Flickr users are assigned to tracts with fewer than 15% black residents. This raises the possibility that measurement error in either the home assignment or face detection process could be biasing the relationship between tract characteristics for highly black tracts. As these tracts make up a small portion of my overall sample, however, I do not expect this error to affect my estimation procedure.

It is possible that the rate of black faces in a Flickr user's photographs is not representative of the user's actual rate of interaction with black friends. This could arise for two reasons. First, it is likely that users take pictures of family, which means that the racial representation of people in the photographs will tend to overstate the degree of social segregation among friends. Second, it is possible that users "curate" their photographs to show either a higher or lower frequency of inter-racial interactions. To examine the relationship between social media photographs and actual social behavior, I turn to my survey of MTurk workers. In this survey, I ask detailed questions about respondents' social contacts and behavior; I also ask for information about the racial breakdown of faces in a recent social media photograph. I then compare the user's actual social behavior to the behavior depicted in the photo.

Recall from Table 6 that the sample of MTurk workers is disproportionately young and educated, and more likely to be white or Asian than the U.S. population at large. For education and race, this approximately mimics the demographics for the Flickr sample that I estimated in Table 7. The relationship between social photos and social interactions in the MTurk sample is therefore likely to

be similar to that found in the Flickr sample.

For each MTurk respondent, I construct a measure of the inter-racial interaction rate based on the time use portion of the survey. Recall that in this module, I ask users to account for their activities for each 3 hour portion of the day. I also ask who was present during each activity, and for the racial breakdown of the people involved. I use this information to construct the average proportion of black people present when the user spends time with friends alone.[30] This measure corresponds closely to the inter-racial interaction rate in my model, because it captures the relative amount of time spent with black friends, compared to the total amount of time spent with friends. A drawback of this measure, however, is that it can be constructed for a small number of people in my sample: only 461 of the approximately 1500 respondents spent any time with friends on the sample day, and only 217 of these spent time with friends alone (which is required for me to use the fraction of black people present as a measure of the proportion of black friends.)

To connect inter-racial interactions to social media photos, I ask respondents to choose the last photograph of a social event that they posted to a social media site, and to report the racial breakdown of people in the photograph. This exercise is randomly determined to occur before or after the reporting of social behavior. For non-black respondents, the mean reported fraction of black faces is 4%, which is very similar to my Flickr data.

Figure 3 shows a scatter plot of the inter-racial interaction rate for the non-black sample against the fraction of black faces in their social media photographs. The figure shows the fitted line from a regression of the friends measure on the photos measure (with no constant), along with the 45-degree line. The two lines are indistinguishable from one another. The regression coefficient is 1.002 and is significant at the 1% level. The fraction of black faces in this *single* photograph can explain approximately 39% of the variation in the proportion of black friends. This suggests that the fraction of black individuals in a user's photographs is an excellent measure of the user's inter-racial interaction rate.

My simulation strategy requires that I know the actual black interaction rate for non-black individuals in my 202 CBSAs. To correct for observable differences between my Flickr sample and other individuals in these cities, I regress the black interaction rate on CBSA and tract characteristics, and use these relationships (shown in Appendix Table A4) to predict the black interaction rate for each

[30] The user may indicate that multiple categories of people (friends, spouse/partner, children, other family, coworkers, etc) were present for a given activity.

city. This exercise suggests that the black interaction rate is slightly higher among Flickr users than among non-black Americans more generally: my estimated black interaction rate for all non-black Americans is 3.5%, compared to 3.7% in the Flickr sample.

5 Results

5.1 Main results: the effect of desegregating cities

In this exercise, I randomly reallocate individuals to new locations in their home city, and calculate how inter-racial interaction behavior would change. Step 1 of my exercise is estimating the residual parameters α_{ww} and α_{wb} from my model. Recall from Section 3 that the estimating equation I use for this is:

$$ln(\mu_{hirs}) = \alpha^c_{rs} + ln(\Sigma_j \hat{\mu}_{irjs} + D^c_{rs}) + \varepsilon_{hirs} \tag{12}$$

In this regression, h indexes individuals, i neighborhoods, r and s are racial categories, and c is a city. The dependent variable is the estimated fraction of time each Flickr user spends interacting with non-black and black people respectively. This is calculated as the tract-level interaction rate multiplied by the fraction of time the user spends interacting with non-black/black people, as estimated from their photographs. For a typical user, the non-black interaction rate would be around 0.241 (a 25% interaction rate, of which 96.5% is with non-black people) and the black interaction rate would be around 0.01. The key variable on the right-hand side of the equation is $\hat{\mu}_{irjs}$, which I construct based on the disutility of travel, distance, and population supplies for every tract and link to Flickr users based on their home tract locations. I run this regression separately using non-linear least squares for white-white interactions and white-black interactions, and separately for each city. I constrain the coefficient on $\Sigma_j \hat{\mu}_{irjs}$ to be equal to 1. The constant term outside of the logarithm is my estimate of α^c_{rs}, while a constant term within the logarithm term is interpreted as reflecting D^c_{rs} - a measure of the strength of residential sorting based on interaction-relevant characteristics.

This procedure produces a separate estimate of α^c_{ww} and α^c_{wb} for each city c, and, when combined with the error term ε_{hirs}, for each Flickr user. In a next step, I adjust the estimates of α^c_{ww} and α^c_{wb} to take account of the observable differences between Flickr users and other residents of their cities. Specifically, I run a regression of ε_{hirs} on the same set of observable characteristics shown in

Table A4 and use the coefficients to predict the average level of ε_{hirs} for non-black individuals in the city; I then add this to the term $\hat{\alpha}^c_{sr}$ to arrive at my final estimates.

The fifth column of Table 1 shows the average of the parameters α^c_{ww}, as well as its variation by region. The mean estimate across cities is -3.996. The negative value indicates that individuals prefer not to interact at a randomly selected moment in time; this occurs because individuals choose not to socialize about 75% of the time. Of course, the realized value of interactions *when the individual chooses to interact* are positive. This occurs when the random shock in the individual's utility function is sufficiently positive. The table also shows that individuals in the South appear to enjoy same-race interactions more than individuals in other regions, with individuals in the Northeast enjoying interactions the least.

The sixth column of Table 1 shows the estimates of α^c_{wb}, while the seventh column shows the ratio of α^c_{wb} to α^c_{ww}. The latter measure captures the relative distaste for black interactions, compared to non-black interactions. The average value of α_{wb} is -8.384, indicating that the typical non-black American dislikes interacting with black people about twice as much as she dislikes interacting with non-black people. This ratio ranges from about 1.7 in the Northeast, to 2.4 in the South.

Table 9 shows how this relative distaste for black interactions varies by city characteristics. The first column confirms the regional differences, and shows that relative distaste for black interactions is significantly higher in the South than in the other three regions. The second column shows that the relative distaste for black interactions is higher in more segregated cities, while column (3) shows that it is higher in cities where blacks make up a large proportion of the population. Column (4) shows that there is less distaste for black interactions in larger cities. The fifth column combines these covariates, and shows that the regional differences disappear once you control for the proportion black in the population, but that all other relationships remain similar.

Although I have estimated the terms α_{ww} and α_{wb} for non-black individuals only due to lack of data on black interaction rates, these parameters can be interpreted as capturing the preference for same-race and different-race interactions respectively. When I implement my simulation exercise, I will impose this assumption, and set $\alpha_{bb} = \alpha_{ww}$.

In the next step of my simulation exercise, I reallocate individuals to new locations in the city. In doing so, I eliminate the term D from the interaction decision. In other words, I eliminate all residential sorting based on observable or unobservable characteristics. Each resident therefore views

individuals from every other neighborhood as being (on average) inter-changeable.[31] This allows me to calculate the equilibrium pattern of interactions between every pair of neighborhoods in this world, using the equilibrium condition from the model.

While I have interpreted the parameters α as reflecting preferences for interactions, the more accurate interpretation is that they capture all factors other than physical distance that shape interaction behavior. Importantly, this could include other causal effects of neighborhoods. If this is the case, the terms α_{ww} and α_{wb} will not remain constant when I reallocate individuals to new neighborhoods. This means that my estimates place a bound on the impact of desegregating cities. In other words, if interaction behavior is strongly shaped by neighborhood channels other than the effect of physical distance, then redistributing individuals in a less segregated way may have a more positive effect on inter-racial interaction behavior than that which I estimate.

Table 10 shows the results of this exercise. In the first two rows of the table, I highlight an important and unintended potential consequence of efforts to make cities less segregated: a decline in the overall interaction rate. By removing individuals from the immediate vicinity of interaction partners that they like, the relative cost of socializing rises. My results suggest that this channel is quite strong. Depending on the value of δ that I use and the group I consider (blacks or non-blacks), the simulated interaction rate falls from around 25% to between 5-15%. In Table 11, which breaks the results for non-blacks down by region, I show that this effect is relatively strong in the Northeast, where interaction rates fall to 2.5-5%. The Pacific region also shows a large decline, with interaction rates falling to 4-10%. This is because, as shown in Table 1, these regions have relatively high distaste for travel. The effect is somewhat more muted in the Midwest and South, where individuals are more willing to travel.

In the next rows of Table 10 and Table 11, I show how the proportion of interactions that take place with black people changes when we desegregate cities. The results for the non-blacks in Table 10 show that, for the country as a whole, the relative black interaction rate *falls* when we desegregate cities. Again, this must be related to the fact that individuals are no longer sorted into neighborhoods where they like the black population better. In real life, individuals interact with their black neighbors, both because they are close by and because these neighbors are more closely matched on observable and unobservable characteristics than the typical black person within

[31] Of course, the actual individuals they choose to interact with are not likely to be random; this is captured in the random shock in the utility function.

a city. When we move these desirable black interaction partners further away, the black interaction rate falls. This effect is not mechanical. As shown by the regional breakdown in Table 11, the decline in the rate of black interactions is driven by the South and the Pacific regions, with the black interaction rate actually rising slightly in the Northeast and Midwest. Nonetheless, the potential for black interactions to fall when cities become less segregated is, to my knowledge, a new result in this literature.

The final rows of Table 10 and Table 11 combine the results on the total interaction rate and the fraction of black interactions to calculate the amount of time that non-blacks spend interacting with blacks. Unsurprisingly, the net effect of a reduction in social interactions and a decline in the relative black interaction rate is to reduce the overall black interaction rate quite substantially. I estimate that non-blacks in the U.S. currently spend about 1% of their non-work time interacting with black people. Under the desegregated regime, this would fall to 0.1-0.3%. Table 11 shows that this is true in every region, even those where the relative black interaction rate rises. This is because the decline in the number of social interactions is much larger in scale than the slight increase in the proportion of black interactions in these regions.

5.2 Supplementary results: is the disutility of travel "big"?

It is possible that the results from my main simulation exercise occur because the disutility of travel is not an important factor in determining individuals' interaction behavior. In this case, the positive causal effect of desegregating neighborhoods in my model is quite limited, and could be outweighed by a relatively small "sorting" effect. In this exercise, I attempt to rule out this possibility by showing that the distaste for travel can explain important deviations in interaction behavior.

Recall that in this simulation, I model a world in which individuals maintain their existing residential locations but make interaction decisions purely based on the distaste for travel. Despite the lack of racial preferences in this world, there will be some social segregation induced by racial segregation in geographic location. By calculating the magnitude of social segregation in this world and comparing it to the perfectly integrated ideal, we can get a sense of whether the parameter δ (the disutility of travel) is quantitatively large; that is, whether it alone can explain significant deviations in interaction behavior from the case of random matching.

The results of this exercise are presented in Table 12. In the first row of the table, I show the actual rate of black interaction for non-blacks. As described in the previous section, I believe

this number to be around 3.5%, indicating that non-black individuals have about 3.5% of their interactions with black people. Unfortunately, the lack of black Flickr users in my data preclude me from measuring the fraction of black interactions for black individuals in similar way.

In the second row of the table, I show what fraction of interactions would be with black people under random matching within cities. This is equal to the proportion of black people in an individual's CBSA. For the average non-black person, random matching would imply that 13.1% of their interactions were with black people. This number is higher for black people (19.4%), because there is some segregation present even across cities. The difference between the random and actual proportion of black interactions is a measure of social segregation. This measure is shown in row 3 of the table. This measure of social segregation is 10.4 percentage points for non-blacks, and cannot be calculated for the black population.

I present the results of my simulation exercise in row 4 of the table. In columns (1) and (2), I show the proportion of black interactions that would occur for non-blacks and blacks, respectively, in the world where interaction decisions are made based purely on the distaste for travel and the parameter δ is pegged to be equivalent to the average hourly wage in each city. The simulation shows that, in this world, non-blacks would see blacks for about 11.1% of their interactions, while blacks would see other black partners for about 37.1% of their interactions. Columns (3) and (4) show the results of the same exercise, with δ increased to be equivalent to 2 times the average hourly wage. In this case, the proportion of black interactions for non-blacks would be around 10.3% and for blacks would be around 42.0%.

Row 5 of columns (1) and (3) of the table shows the percentage decrease in the proportion of black interactions for non-blacks, compared to the random matching ideal. Based on the existing degree of residential segregation, avoidance of travel alone can account for a 15-20% decline in the relative frequency of cross-racial interactions for non-blacks. As shown in columns (2) and (4) of the same row, the same exercise suggests an increase in the relative frequency of black interactions for black people, on the order of 90-120%. The desire to avoid excess travel, combined with the existing degree of residential segregation, seems to be sufficiently large to induce significant deviations from integrated behavior, even in the absence of racial preferences.

Another way of scaling the effect of the distaste for travel is to compare the social segregation generated by distance alone to the degree of social segregation that exists in real life. This tells us about the *relative* importance of δ compared to other factors (captured in the parameters α in my

model) in explaining interaction behavior. As shown in row 6 (columns (1) and (3) only), distance alone can explain 20-30% of the existing level of social segregation. Again, this suggests that the avoidance of travel is quite a strong motivation when it comes to explaining interaction behavior.

Table 13 shows how the results of this exercise vary across Census regions, for non-blacks only. This table shows that both the absolute and relative impact of physical distance are highest in the Midwest and South (explaining a 20-30% reduction in the cross-racial interaction rate from the random ideal, and 20-40% of the total amount of social segregation), moderately sized in the Northeast, and almost non-existent in the Pacific. Examining the regional values of segregation and δ in Table 1 suggests that this pattern is driven primarily by the relatively high degree of black geographic isolation in these regions, rather than by a higher distaste for travel (in fact, the opposite is true; the distaste for travel is lowest in the Midwest and South.)

Of course, these results depend on the existing pattern of residential segregation, which is likely to be generated in part by preferences over social interactions. It is important to emphasize that this sorting does not affect the interpretation of my results. My results suggest that if we eliminated racial preferences and all other influences on social interaction decisions, but left the pattern of residential segregation the same, we would still see a substantial deviation from the perfectly integrated ideal based purely on individuals' desire to avoid travel. In other words, given the revealed preference for avoiding travel, the existing degree of residential segregation is sufficiently high to induce substantial deviations in behavior. This tells us about the potential importance of travel-avoidance in driving interaction behavior.

6 Conclusion

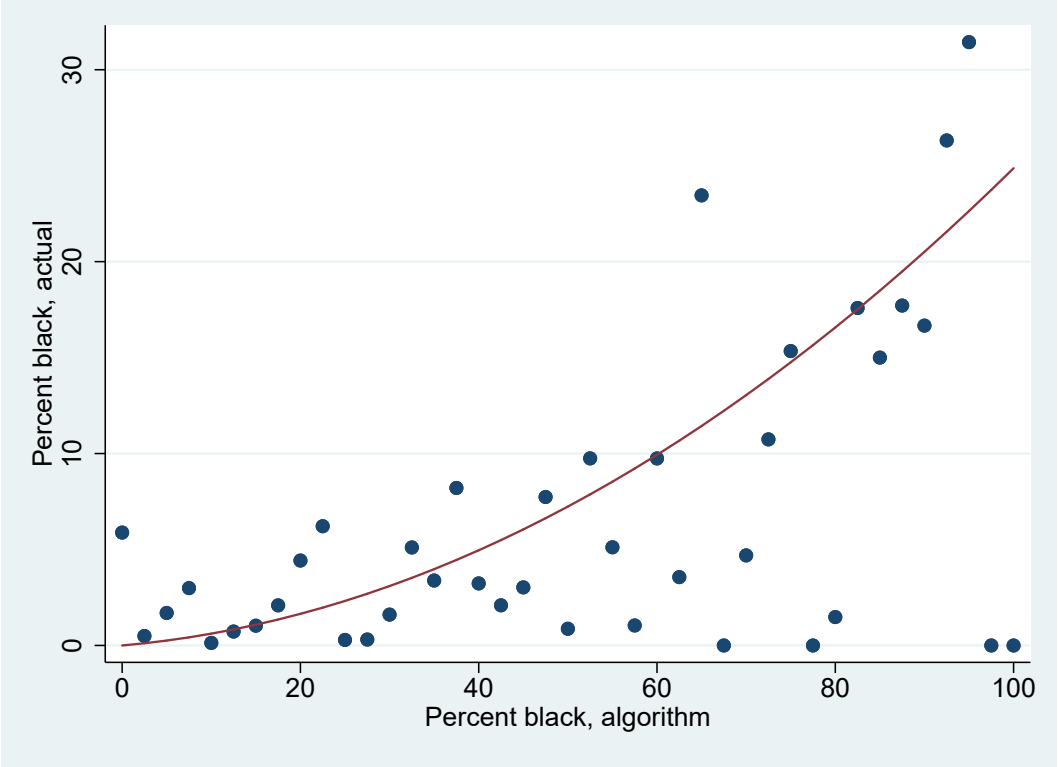
This paper highlights an important, and (to my knowledge) previously unrecognized fact: that residential sorting can *increase* inter-racial contact, even when it results in racial segregation. As I argue, and show in my secondary simulation exercise, there is at least one important causal link between neighborhoods and social interactions: the effect of physical distance. This effect is sufficiently important that it can account for a 15-20% reduction in cross-racial interactions from the perfectly integrated ideal, given the existing pattern of residential segregation. Despite this, integrating cities would have a *negative* effect on the number of cross-racial interactions. This occurs because the existing process of residential sorting encourages cross-racial interactions by ensuring

that individuals pay a low "price" for interacting with the other-race individuals they like best.

While it is unlikely that policy makers would attempt the type of complete desegregation exercise I consider in this paper, my results are also relevant for assessing the more marginal programs that attempt to change the distribution of people within cities. While these programs may serve many worthwhile goals, including more evenly distributing access to good schools and other public services, they may also have unintended consequences for the pattern of social interactions in cities. As my analysis highlights, these consequences may not be unambiguously positive, and may actually serve to reduce interactions between members of different social groups.

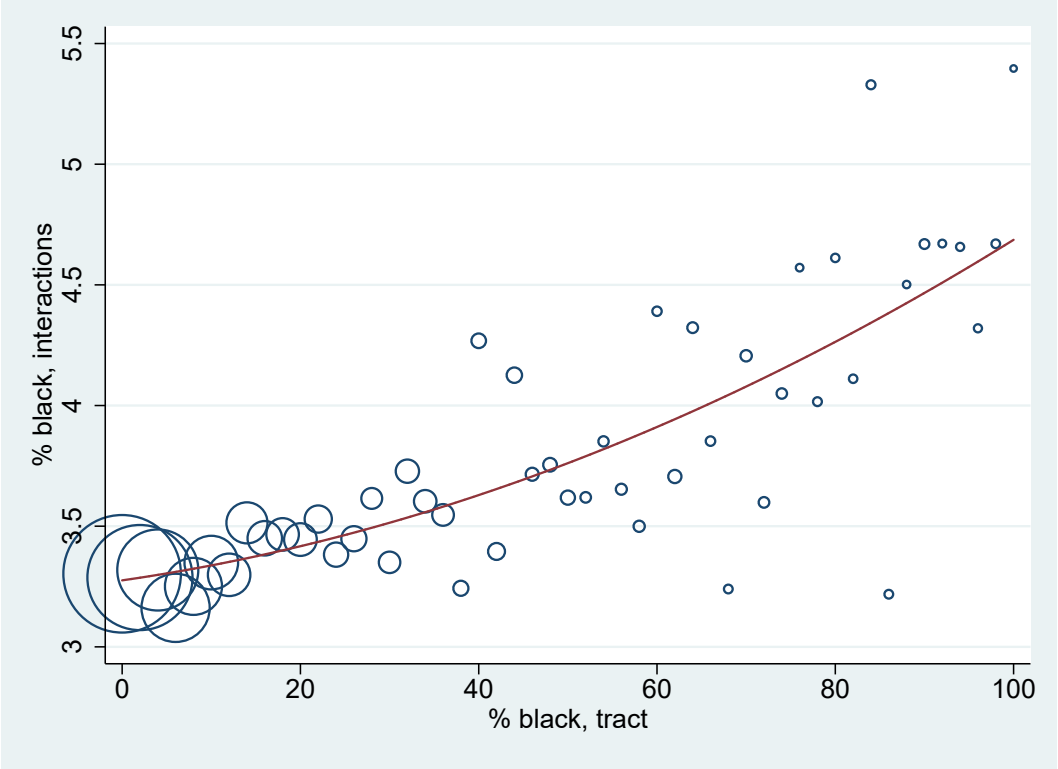
7 Figures

Figure 1: Relationship between black faces, algorithm vs. hand-coded



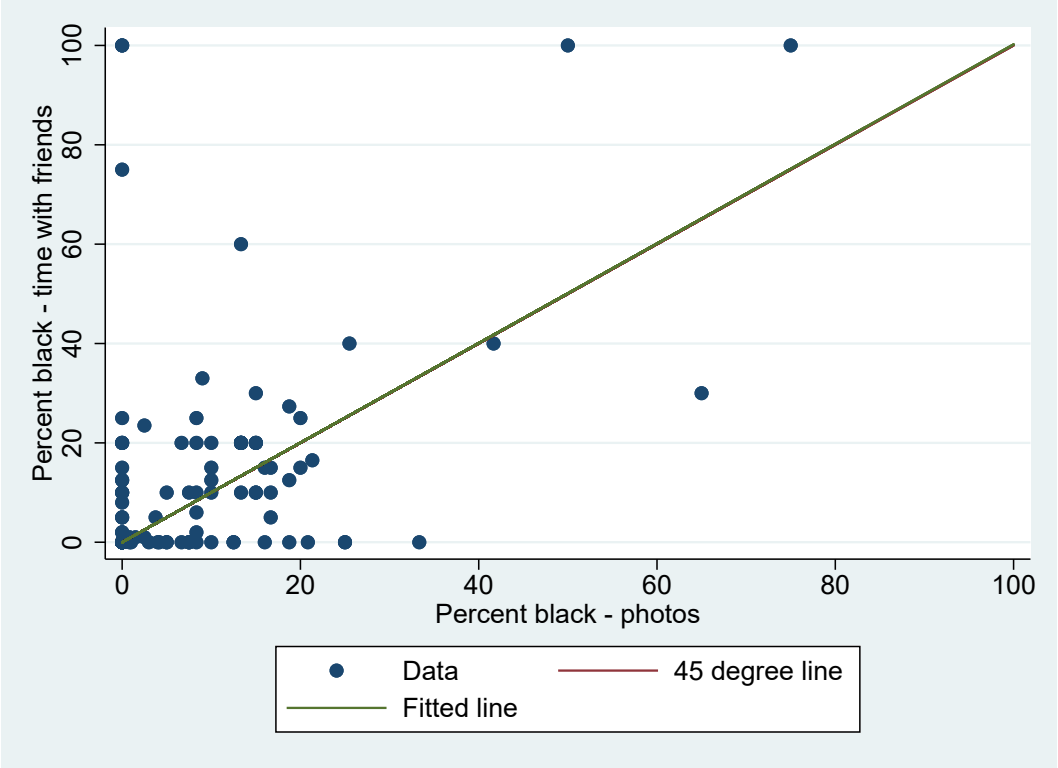
This figure plots the actual fraction of black faces for a set of hand-coded photographs against the fraction of black faces found by the race classification algorithm. Each circle represents a 2.5-percentile group for the fraction of black faces found by the algorithm (0-2.5%, 2.5-5%, etc.); the vertical height shows the mean of the actual fraction black for that group. The size of each circle represents the total number of faces found in the photographs in that group. The red line shows the fitted quadratic relationship between the algorithm's predictions and the actual frequency of black faces.

Figure 2: Relationship between fraction of black interactions and tract demographics



This figure plots the proportion of interaction time spent with black friends against the proportion of the population that is black in a user's assigned home tract. The circles represent the mean within each 2-percentage point cell on the x-asis, with the circle size indicating the number of users in this group. The red line shows the quadratic fit between the two variables. The sample is the set of approximately 93,000 Flickr users living in one of the 202 CBSAs in my main sample.

Figure 3: Relationship between time spent with black friends and black faces in social media photos: MTurk



This figure plots the proportion of interaction time spent with black friends against the proportion of black faces in an MTurk worker's last social media photograph. The sample is a set of 170 non-black MTurk workers who responded to my survey and spent any time alone with friends on the day prior to the survey. The green line is the fitted line from a regression of the interaction measure on the photo measure (with the constant supressed); this overlaps almost precisely with the 45 degree line.

8 Tables

Table 1: Key parameters by region

	Duncan index	SSI	Avg. dist. b/w tracts	δ	α_{ww}	α_{wb}	Relative distaste for black interactions $\left(\frac{\alpha_{wb}}{\alpha_{ww}}\right)$
All	0.521	0.577	41.1 km	0.458	-3.996	-8.384	2.098
Northeast	0.563	0.474	44.9 km	0.523	-4.914	-8.385	1.706
Midwest	0.596	0.642	37.6 km	0.429	-4.597	-9.604	2.089
South	0.502	0.700	39.6 km	0.419	-3.463	-8.409	2.428
Pacific	0.457	0.271	37.5 km	0.506	-3.689	-7.250	1.965

This table shows the mean level of the key parameters in my model, for all 202 cities in my main analysis sample, and separately by region. The Duncan index is computed directly from Census data. The SSI was provided by the authors Echenique and Fryer (2007); note that this variable is available for 167 of my 202 CBSAs only. The average distance between tracts was calculated using shapefiles provided by the U.S. Census Bureau. For details on the estimation of the other parameters, please see the Data section.

Table 2: Distance to the average white/black person, by race

	Distance to the average:	
	Non-black person	Black person
Non-black	27.8 km	26.7 km
Black	26.9 km	22.7 km

This table shows the mean distance to the average black/non-black person within the same CBSA, for blacks and non-blacks separately. These figures were calculated using great-circle distance between Census tracts, based on shapefiles provided by the U.S. Census Bureau, as well as information on the population of each tract by race from the 2010 Census. The sample is the set of all individuals living in one of the 202 cities in my main analysis sample.

Table 3: Previous estimates of the disutility of distance

	Context	Year(s) of observation	Estimated cost of travel, per minute*	Ratio of travel cost to average hourly wage*
Thomadsen (2005)	Fast food, Santa Clara County	1999	0.49	2
Davis (2006)	Movie theatres, 36 cities	1996	0.23&	1
McManus (2007)	Coffee shops, University of Virginia	2000	0.10	0.5-1#
Manuszak and Moul (2009)	Gas stations, Chicago & surrounding area	2001	0.18-0.24	0.68-0.91
Houde (2012)	Gas stations, Quebec City	1991-2001	0.10-0.57@	0.75-2.50@
Seim and Waldfogel (2013)	Liquor stores, Pennsylvania	2005	0.46	1.95

* All dollar estimates are in 2002 USD. Where possible, I use the authors' reported estimates of hourly wages to construct the ratio shown in column (4). Where this is not possible, I use the national hourly wage for the appropriate year, multiplied by the ratio of median income in the relevant geographic area to the median income of the United States.

& Davis (2006) estimates a non-linear function of distance; following Seim and Waldfogel (2013), the reported coefficient is the estimated cost of travelling 3.2 km.

The estimated coefficient is equal to approximately the average wage for students in the relevant geographic market; it is equal to about 0.5 times the average wage for adults in Virginia.

@ The initial estimates reported by (Houde, 2012) are larger than this. His preferred estimates suggest that a time valuation of 4 times the average hourly wage. However, these estimates do not account for traffic. Once I adjust for the average speed of traffic in Quebec City at rush hour (the relevant time, since the estimates examine consumers' willingness to deviate from commute paths), the estimates are reduced to those shown in the table.

Table 4: Relationship between estimated disutility of travel and travel patterns in Flickr

	Fraction of photos taken within indicated distance of home			
	1 km	3 km	5 km	10 km
Coefficient on $\hat{\delta}$	0.158*	0.168**	0.154**	0.089***
	(0.095)	(0.085)	(0.069)	(0.043)
N	202	202	202	202

This table shows the results from a regression of the mean fraction of photos taken within the indicated distance of Flickr users' homes on the estimated disutility of travel. A increase in the disutility of travel implies that users living within that city or tract dislike travel more. The sample is the set of 202 CBSAs in my main analysis sample.

Table 5: Frequency of social interactions: ATUS

	Dependent variable: indicator for spent any time with friends on diary day	
	Non-black sample	Black sample
Age	-0.017***	-0.011***
	(0.001)	(0.001)
Age squared	0.000***	0.000***
	(0.000)	(0.000)
No high school	-0.112***	-0.051**
	(0.008)	(0.020)
High school	-0.083***	-0.065***
	(0.006)	(0.018)
Some college	-0.059***	-0.068**
	(0.006)	(0.018)
Bachelor's	-0.025***	-0.059***
	(0.006)	(0.020)
Region - Northeast	0.004	-0.009
	(0.004)	(0.019)
Region - Midwest	0.010*	0.002
	(0.005)	(0.019)
Region - South	0.002	-0.005
	(0.005)	(0.017)
Constant	0.722***	0.596***
	(0.005)	(0.038)
N	56,048	9,073
R^2	0.028	0.019
Mean of dep. variable	0.245	0.246

This table shows the results from a regression of an indicator for spending any time with friends on demographic characteristics, in the American Time Use Survey sample. The sample is the set of all diary respondents aged 15-85 who filled out a diary on a weekend day.

Table 6: Demographics: MTurk survey respondents compared to U.S. population

	MTurk	U.S. population (over 18)
% age 18-25	15.5%	15.0%
% age 26-45	67.9%	36.0%
% age 46-65	14.5%	33.3%
% age 65	2.1%	15.7%
% male	52.1%	48.5%
% white	73.5%	69.3%
% black	7.1%	11.8%
% Asian	6.5%	5.0%
% Hispanic	11.7%	13.9%
% No high school	0.1%	13.3%
% high school	10.4%	38.0%
% some college	33.3%	23.3%
% Bachelor's	41.3%	16.4%
% post-grad	14.9%	9.0%
N	1,628	11,572,214

This table shows demographic information for my MTurk survey sample, and for the U.S. population over the age of 18. Information on the U.S. population comes from the 2006-2010 American Community Survey.

Table 7: Tract demographics: comparison to U.S. population

	Flickr users	Population - same cities	Population - all
CBSA population	5,348,751	4,688,542	3,834,430
CBSA SSI*	0.702	0.711	0.700
Tract level:			
Density (pop/sq. km)	3,338	2,604	2,197
Median age	37.8	37.0	37.1
Median income	$34,171	$30,461	$29,046
% white	72.5	71.0	73.3
% black	10.3	13.6	12.7
% Asian	8.5	5.8	4.9
% Hispanic	13.3	17.8	16.4
% No high school	11.4	14.8	15.2
% high school	20.5	27.0	28.3
% some college	25.0	28.0	28.2
% Bachelor's	25.2	19.0	17.8
% post-grad	17.9	11.2	10.5
Number of individuals	87,640	231,874,255	285,098,410
Number of tracts	27,279	54,336	66,970
Number of cities	202	202	933

This table shows average home CBSA and tract demographics for users in my sample, compared to the averages for the U.S. population living in the same set of cities (column 2), and to the entire U.S. population (column 3.) Demographic information is taken from the 2006-2010 American Community Surveys. * This variable is calculated for the 167 CBSAs in my main analysis sample for which it is available.

Table 8: Race classification confusion matrix

	Classified as:	
	Non-black	Black
Actual race: Non-black	87%	13%
Black	25%	75%

This table shows the proportion of non-black/black faces that were classified as non-black/black by the race classification algorithm. The sample is a subset of faces from 20,000 randomly selected Flickr photographs, 10% of which are set aside for testing purposes.

Table 9: Relationship between city characteristics and relative distaste for black interactions

	Dependent variable: $\frac{\alpha_{wbc}}{\alpha_{wwc}}$				
	(1)	(2)	(3)	(4)	(5)
Northeast	0.162				-0.130
	(0.309)				(0.315)
Midwest	0.311				-0.141
	(0.290)				(0.332)
South	0.666***				-0.017
	(0.256)				(0.323)
SSI		0.580**			0.512
		(0.291)			(0.470)
% black			2.934***		2.208*
			(0.882)		(1.300)
Log population				-0.247***	-0.345***
				(0.093)	(0.109)
Mean of dep. var.	2.434	2.434	2.434	2.434	2.434
N	167	167	167	167	167

This table shows the results from a regression of relative distaste for black interactions on city characteristics. The relative distaste term is constructed as $\frac{\alpha_{wbc}}{\alpha_{wwc}}$, where α_{wbc} captures non-black people's preferences for interacting with black people and α_{wwc} captures non-black people's preference for interacting with other non-black people. As both terms are negative for all cities in the sample, more positive values of this term indicate a stronger relative *dislike* for black interactions. See the text for details on how the α terms are estimated. The sample is the set of 167 cities in my main analysis sample that have information on the black SSI from Echenique and Fryer (2007).

Table 10: Results: main simulation exercise

	δ low		δ high	
	Non-black	Black	Non-black	Black
Interaction rate				
Actual	25.1%	26.5%	25.1%	26.5%
Simulated	11.4%	14.5%	5.2%	6.7%
% of interactions with blacks				
Actual	3.5%	Unknown	3.5%	Unknown
Simulated	2.3%	93.2%	2.3%	93.2%
% of time interacting with blacks				
Actual	0.9%	Unknown	0.9%	Unknown
Simulated	0.3%	13.5%	0.1%	6.2%

This table shows the results from my second simulation exercise, in which individuals are randomly distributed throughout the city but maintain their racial preferences. The first two columns show the results of the exercise when I peg δ, the disutility of travel, to be equivalent to the hourly wage in the city; the third and fourth columns show the results when I peg δ to be equivalent to 2 times the average hourly wage. Row 1 shows the actual and simulated interaction rate for blacks and non-blacks. Row 2 shows the actual and simulated proportion of all interactions that take place with black people. Row 3 combines the first two rows and calculates the percentage of time that each group spends interacting with black people.

Table 11: Results: main simulation, by region (non-blacks)

	Northeast	Midwest	South	Pacific
Interaction rate				
Actual	24.1%	25.5%	24.9%	25.7%
Simulated - δ low	5.3%	15.4%	14.6%	9.6%
Simulated - δ high	2.4%	6.8%	6.9%	4.1%
% of interactions with blacks				
Actual	3.6%	3.3%	3.5%	3.5%
Simulated - δ low	3.9%	3.9%	1.3%	1.2%
Simulated - δ high	3.9%	4.0%	1.3%	1.2%
% of time interacting with blacks				
Actual	0.9%	0.8%	0.9%	0.9%
Simulated - δ low	0.2%	0.6%	0.2%	0.1%
Simulated - δ high	0.1%	0.3%	0.1%	0.0%

This table shows the results from my second simulation exercise, in which individuals are randomly distributed throughout the city but maintain their racial preferences, separately by region and for non-blacks only. Row 1 shows the actual and simulated interaction rate. Row 2 shows the actual and simulated proportion of all interactions that take place with black people. Row 3 combines the first two rows and calculates the percentage of time that each group spends interacting with black people.

Table 12: Results: secondary simulation exercise

	δ low		δ high	
	Non-blacks	Blacks	Non-blacks	Blacks
Black int. rate:				
Actual	3.5%	Unknown	3.5%	Unknown
Random	13.1%	19.4%	13.1%	19.4%
Social segregation (Actual - random)	9.6 pp	Unknown	9.6 pp	Unknown
Simulated	11.1%	37.1%	10.3%	42.0%
% change, random to simulated	-15.3%	91.2%	-21.4%	116.4%
% explained by distance	20.8%	Unknown	29.2%	Unknown

This table shows the results from my first simulation exercise, where individuals are assumed to maintain their existing location in the city but make interaction decisions based on the disutility of travel only. The first row shows the actual proportion of interactions with black people, based on the Flickr data; because my Flickr data do not contain a sufficient number of black individuals, I report this for non-blacks only. The second row shows the proportion of black interactions that would occur if individuals matched randomly within cities. It is equal to the proportion of black people within an individual's CBSA. The difference between the random and actual rates is my index of social segregation, reported in row 3. Row 4 reports the frequency of black interactions that would occur if individuals made decisions based only on the disutility of travel. The fifth row reports the percentage difference between the predicted and random rate, while the sixth row reports the % of the total social segregation index that can be accounted for by distance alone.

Table 13: Results: secondary simulation, by region (non-blacks)

	Northeast	Midwest	South	Pacific
Actual	3.6%	3.3%	3.5%	3.5%
Random	12.7%	13.6%	19.2%	5.5%
Segregation	9.1 pp	10.3 pp	15.7 pp	2.0 pp
Simulated (δ low)	11.2%	11.1%	15.7%	5.5%
Simulated (δ high)	10.2%	9.7%	14.6%	5.4%
% change, random to simulated	-11.8 to -19.7%	-18.4 to -28.7%	-18.2 to -24.0%	0.0 to -1.8%
% explained by distance	16.5 to 27.5%	24.3 to 37.9%	22.3 to 29.3%	0.0 to 4.5%

This table shows the results from my first simulation exercise, where individuals are assumed to maintain their existing location in the city but make interaction decisions based on the disutility of travel only, with the results broken down by region. The interpretation of each line is the same as in Table 12, with the exception that I report ranges for the last two rows (based on the simulated results with δ low and δ high).

9 Appendix

9.1 Additional tables

Table A1: Number of visits to home and other locations in CBSA

	Mean number of visits	Fraction of all visits
Home tract	14.0	42.3%
Other tracts	2.6	4.5%

This table shows the number of unique visits a Flickr user makes to his or her assigned home tract, compared to other tracts she visits. The sample is a set of approximately 87,000 Flickr users who live in one of the 202 CBSAs in my main analysis sample.

Table A2: Number of Foursquare venues around photo locations: home tract vs other visited tracts

	Number of venues		
	Home tract	Other visited tracts	Difference
Food & drink	0.780	0.929	-0.149***
			(0.016)
Entertainment	0.550	0.567	-0.016
			(0.014)
Stores	0.342	0.448	-0.107***
			(0.011)
Offices	0.138	0.156	-0.018***
			(0.005)
Other	1.962	2.092	-0.129***
			(0.026)
All venues	3.773	4.191	-0.419***
			(0.036)

This table shows the mean number of Foursquare venues within 25 m of a photograph's location, depending on whether that location is within the Flickr user's home Census tract or not. The sample for these calculations is a sample of approximately 40,000 photos taken by about 20,000 Flickr users who live in one of the 202 CBSAs in my main sample. Users were randomly sampled for inclusion in this set, and two photos were randomly sampled for each user: one in the user's "home" census tract, and one in another tract that the user visits.

Table A3: Relationship between demographics predicted by last name and home tract demographics

	Dependent variable:		
	% same race/eth, CBSA	% same race/eth, visited tracts	% same race/eth, home tract
Prob. user	0.073***	0.099***	0.107***
is white	(0.004)	(0.005)	(0.008)
N	10,172	10,172	10,172
R^2	0.027	0.033	0.016
Prob. user	0.033***	0.053***	0.082***
is black	(0.007)	(0.008)	(0.012)
N	10,172	10,172	10,172
R^2	0.002	0.004	0.004
Prob. user	0.116***	0.136***	0.140***
is Hispanic	(0.004)	(0.004)	(0.006)
N	10,172	10,172	10,172
R^2	0.061	0.096	0.057
Prob. user	0.045***	0.099***	0.100***
is Asian	(0.003)	(0.004)	(0.005)
N	10,172	10,172	10,172
R^2	0.019	0.059	0.036

The table shows the results from a regression of the proportion white, black, Hispanic or Asian in i) a Flickr user's CBSA, ii) all tracts a Flickr user visits (excluding the home tract), and iii) the home tract on the user's probability of being that race, based on his or her last name. For example, Column (1) in Panel 1 regresses the proportion white in a user's CBSA on the probability that she is white based on her last name. The probability distribution over race by last name comes from the Census Bureau (2010). The sample is the set of Flickr users in one of the 202 CBSAs in my main sample who have a last name appended to their profile.

Table A4: Relationship between black interactions and city/tract characteristics: Flickr

	Dependent variable: black interaction rate
Ln CBSA population	-0.002
	(0.031)
CBSA segregation	0.146
	(0.461)
CBSA % black	0.006
	(0.005)
Ln tract population	-0.102**
	(0.041)
Tract % black	0.012***
	(0.002)
Tract % no high school	-0.003
	(0.004)
Tract % high school	-0.001
	(0.004)
Tract % some college	-0.011***
	(0.004)
Tract % college	-0.006
	(0.005)
Tract median age	-0.043*
	(0.023)
Tract median age squared	0.001**
	(0.000)
Ln tract median income	-0.012
	(0.076)
Tract density	0.000***
	(0.000)
Northeast	0.002
	(0.097)
Midwest	-0.168
	(0.114)
East North Central	-0.094
	(0.103)

10 References

References

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis*.

Brueckner, Jan K., and Ann G. Largey. 2008. "Social Interaction and Urban Sprawl." *Journal of Urban Economics*, 64(1): 18–34.

Choo, Eugene, and Aloysius Siow. 2006. "Who Marries Whom and Why." *Journal of Political Economy*, 114(1): 175–201.

Davis, Donald R., Jonathan I. Dingel, Joan Monras, and Eduardo Morales. forthcoming. "How Segregated is Urban Consumption." *Journal of Political Economy*.

Davis, Peter. 2006. "Spatial Competition in Retail Markets: Movie Theaters." *RAND Journal of Economics*, 37(4): 964–982.

Echenique, Federico, and Roland G. Fryer. 2007. "A Measure of Segregation Based on Social Interactions." *Quarterly Journal of Economics*, 122(2): 441–485.

Galichon, Alfred, Scott Duke Kominers, and Simon Weber. 2016. "Costly Concessions: An Empirical Framework for Matching with Imperfectly Transferable Utility."

Glaeser, Edward L., and Joshua D. Gottlieb. 2006. "Urban Resurgence and the Consumer City." *Urban Economics*, 43(3): 1275–1299.

Han, Hu, and Anil K. Jain. 2014. "Age, Gender and Race Estimation from Unconstrained Face Images." *MSU Technical Report*, , (MSU-CSE-14-5).

Hellerstein, Judith, and David Neumark. 2008. "Workplace Segregation in the United States: Race, Ethnicity and Skill." *Review of Economics and Statistics*, 90(3).

Houde, Jean-Francois. 2012. "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline." *American Economic Review*, 102(5): 2147–2182.

Huff, Connor, and Dustin Tingley. 2015. "'Who are these people?' Evaluating the demographic characteristics and political preferences of MTurk survey respondents." *Research and Politics*, 2(3): 1–12.

Jeffries, Adrianne. 2013. "The Man Behind Flickr on Making the Service 'Awesome Again'." *The Verge*.

Kazemi, Vahid, and Josephine Sullivan. 2014. "One Millisecond Face Alignment with an Ensemble of Regression Trees." *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1867–1874.

Krysan, Maria, and Kyle Crowder. 2017. *Cycle of segregation: social processes and residential stratification.* New York:Russell Sage Foundation.

Manuszak, Mark D., and Charles C. Moul. 2009. "How Far for a Buck? Tax Differences and the Location of Retail Gasoline Activity in Southeast Chicagoland." *Review of Economics and Statistics*, 91(4): 744–765.

Marmaros, David, and Bruce Sacerdote. 2006. "How do Friendships Form?" *Quarterly Journal of Economics*, 121(1): 79–119.

Massey, Douglas S., and Nancy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass.* Cambridge, MA:Harvard University Press.

McManus, Brian. 2007. "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee." *RAND Journal of Economics*, 38(2): 512–532.

Morris, Mike. 2018. "HUD, Houston come to agreement on city's affordable housing efforts." *Houston Chronicle*, March 9.

Patacchini, Eleonora, Pierre M. Picard, and Yves Zenou. 2015. "Urban Social Structure, Social Capital and Spatial Proximity."

Seim, Katja, and Joel Waldfogel. 2013. "Public Monopoly and Economic Efficiency: Evidence from the Pennsylvania Liquor Control Board's Entry Decisions." *American Economic Review*, 103(2): 831–862.

Thomadsen, Raphael. 2005. "The Effect of Ownership Structure on Prices in Geographically Differentiated Industries." *RAND Journal of Economics*, 36(4): 908–929.

Wilson, William Julius. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy.* Chicago, IL:University of Chicago Press.