Size Matters: The Structural Effect on Social Relations

Siwei Cheng
University of Michigan

Yu Xie
University of Michigan

Population Studies Center Research Report 12-756

March 2012

Direct all correspondence to Siwei Cheng or Yu Xie, P. O. Box 1248, 426 Thompson Street, Population Studies Center, University of Michigan, Ann Arbor, MI 48106-1248 (Email: chengsw@umich.edu or yuxie@umich.edu). The research is supported by a grant from the National Institute of Child Health and Human Development (R21-HD-050-259) and internal grants from the Population Studies Center, the Survey Research Center, and the Center for Chinese Studies of the University of Michigan. The authors are grateful to Elizabeth Bruch, Zheng Mu, Zhen Zeng, and Xiang Zhou for suggestions and comments on earlier drafts of the paper.

ABSTRACT

This paper follows the literature on separating out the structural effects of social contexts from individuals' own preferences in determining social relationship formations. While prior work has almost exclusively focused on compositional measures of social context, we posit that, net of the effect of group composition, total size has a distinct effect on the level of segregation in a population. In this paper, we first formalize a theoretical framework for the analysis of the size effect. Under the assumptions of (1) maximization of preference in choosing a friend, (2) multidimensionality of preference, and (3) racial homophily, we conducted analyses using agent-based modeling techniques and yielded three main findings. First, increasing total population size decreases the likelihood of forming intergroup relations. Second, the size effect increases with the number of preference dimensions. Third, the size effect decreases with the strength of preference. In addition, we empirically tested our theory using friendship nomination data from Add Health and found results consistent with our prediction that increasing total school size reduces the share of interracial friends in a school.

INTRODUCTION

In modern societies, the individual's freedom is widely respected, honored, and indeed privileged. The question of with whom to form social relationships seems to be a matter of personal choice. Yet, social scientists have long emphasized the critical role of social context for social relationship formation (e.g., Allport 1954, 1966; Blau 1974, 1977, 1994; Blau and Schwartz 1984; Granovetter 1985; Merton 1972; Moody 2001; Mouw and Entwisle 2006; Smith and Semin 2007; Williams 1947; Wimmer and Lewis 2010; Zeng and Xie 2008). This emphasis is based on the recognition that social context is a pre-condition for social relations, as it provides the milieu for social interaction and exchange (Dovidio, Gaertner, and Kawakami 2003; Goffman 1971; Granovetter 1973; Granovetter 1985), the establishment of social identity and role relations (Brewer and Weber 1994; Turner and Reynolds 2001), and the social differentiation among people (Blau 1974; Blau and Schwartz 1984; Simmel 1955). As such, it also segregates individuals into different physical or social segments so that they face different *potential* persons with whom to form relationships. These social segments have been described theoretically as "a multidimensional space of social positions" by Blau (1977), "cliques" by Granovetter (1973), "social fields" by Bourdieu (1989), and "the insiders and outsiders" by Merton (1972). In this paper, we gloss over subtle differences in these conceptualizations and lump them together under the catch-all heading of "social contexts."

To understand the role of a social context, we invoke a common but simplistic conceptualization in sociology that the formation of individuals' social relations results from a combination of two mechanisms: individual preference and structural constraint.[1] The former refers to individuals' subjective attitudes regarding potential relationships. Of course, subjective attitudes do not exist in a vacuum. To a large extent, they reflect the institutional and cultural environments in which people interact (Allport 1954,

[1] We say that this conceptualization is simplistic because individual preference and structural constraint can be confounded. In Wimmer and Lewis's (2010) work, for example, triadic relationship and reciprocity, under their label "balancing," are mechanisms for friendship formation that are neither individual preference nor structural constraint. Balancing can be best understood as a causal process leading to an aggregate outcome resulting from both individual preference and structural constraint.

1966; Dovidio et al. 2003; Hewstone et al 2002; Tajfel and Turner 1979; Williams 1947).[2] The latter refers to the structural constraints imposed by a social context on individuals' exposure to potential alternatives. This mechanism emphasizes social structure, a term defined by Blau (1974) as "the social positions that govern the social relations among their incumbents," which segregates people into circles. Local barriers separating these circles limit people's chances of meeting each other and constrain relationship formation to persons in contact with each other. A large empirical literature has already demonstrated the important role of structural constraints in studies of marriage (e.g. Blau 1977, 1994; Blau and Schwartz 1984; Lichter et al. 1995; Mare 1991; Oppenheimer 1988; Qian 1998) and friendship (e.g. Currarini, Jackson, and Pin. 2010; Hallinan and Williams 1989; Moody 2001; Mouw and Entwisle 2006; Quillan and Campbell 2003; Zeng and Xie 2008). We note that the effects of structure and preference are impossible to disentangle in empirical research based on observed data, because individuals' past friends structurally alter their subsequent exposures in a dynamic process, known as the phenomenon of "friend of a friend" (Kossinets and Watss 2009; Snijders 2001; Suitor and Keeton 1997; Watts 1999). The confounding of different sources of influences in reality however does not diminish the sociological significance of the conceptual separation between structure and preference in determining friendship choice. In this paper, we separate out the structural effects of social contexts from individuals' preferences and model social relationship formation as being solely a function of structural variations, thereby holding individual preferences fixed.

The seminal work by Blau (1974; 1977; 1994) and Blau and Schwartz (1984) already revealed the importance of group sizes in social relationship formation. Our study shares with their work a focused concern with purely structural effects. However, Blau and Schwartz failed to further unpack group sizes into their structural components. As we will show below, a group size is necessarily a product of a total size and a group proportion, which should have distinct structural effects on social relationship formation. Indeed, much of the recent empirical literature on friendship choice has, either explicitly or implicitly, gone beyond Blau's work and drawn the distinction between total size and group proportions.

[2] For example, World War II, a historical context, brought about a strong emphasis on national unity "on the basis of personal association of individuals as functional equals" (Williams 1947).

This distinction is evident in empirical research on social relationship formation, where scholars have often characterized a social context with two measurements: total size and group proportions (Currarini et al. 2010; Hallinan and Williams 1989; Joyner and Kao 2000; Moody 2001; Quillian and Campbell 2003; Zeng 2004). However, most of the theoretical interest in the past literature has been concerned only with the structural effects of group proportions, overlooking the important role of total size as a fundamental source of structural effects. As a result, total size as a distinct, structural mechanism affecting the formation of social relationships has so far eluded sociological understanding.

In this paper, we depart from the conventional emphasis on the effect of contextual composition and focus instead on the structural effect of total size. Although our theoretical insights hold true in general for social relationship formation, we focus on interracial friendship as our substantive application. We derive a formal theoretical model from commonly accepted assumptions and conclude that, net of group composition, the total size of a social context increases the level of racial segregation. This is true because a larger total size increases the chances of finding a good match on multidimensional preference for social relationship and reduces the likelihood that an individual will compromise his/her in-group preference.

In the remainder of the paper, we first justify (Sections 1and 2) the need to consider the structural effect of size on social relationship formation. We follow it (Section 3) with a formal theoretical model, the core results of the paper in an agent-based model (Section 4), and an empirical demonstration using nationally representative survey data (Section 5). Finally (Section 6), we sum up the theory and the evidence presented in this paper and point out the broader implications of our model.

1. Structural Effects of Social Context

What is a social context? While a commonly used concept in sociology, social context is not easy to define. One possible approach is to define it in terms of the boundaries around it. This approach views social context as a closed system with confined historical, geographical, and cultural specificities. Within a social context, different people are assumed to experience the same events (e.g. Williams 1947), live in the same place (e.g. Wilson 1987; Massey and Denton 1993), participate in the same activities (e.g.

Allport 1954 & 1966; Feld 1981; Kao and Joyner 2004; Moody 2001; Mouw and Entwisle 2006), or share the same norm and culture (e.g. Harding 2007; Hannerz 1969; Merton 1972). With these boundaries defined, the common practice is to quantify cross-context differences by measuring context-specific structural attributes, such as size and group composition. However, this simplistic approach essentially assumes homogeneity within the boundaries, social or physical, that define a social context. In reality, no matter how small a context is, the possibility always remains that positions within the context differ in their *exposure* - or *physical propinquity*, in Blau's language – to contextual elements. Such within-context heterogeneity in opportunity structure, if not properly addressed, can confound inferences about social behavior (Blau 1977, 1994; Wimmer and Lewis 2010; Zeng and Xie 2008). Although the homogeneity assumption is naïve, it greatly simplifies the analysis of structural constraints and thus serves as our starting point for modeling the influences of social context. In more realistic applications, we would want to modify the assumption by parameterizing within-context heterogeneity as a function of social attributes (Blau 1974, 1977, 1994). For example, we may quantify exposure differences by using statistical techniques such as a distance decay function (Giles-Corti and Donovan 2003; Laurie and Jaggi 2003; Zeng and Xie 2008).[3]

Social research has long been concerned with how social context exerts structural influences on a variety of social relations, such as marriage, friendship, family ties, work relations, and residential neighbors. These social relations differ markedly, as they pertain to different social institutions. Yet, it is possible to propose a common framework that encompasses the role of all structural constraints in these diverse settings. We achieve this by decomposing opportunities in a social context into the following two orthogonal components[4]:

[3] Although we realize the necessity of offsetting the effect of exposure in modeling reality (Zeng and Xie 2008), we will not incorporate this into our model, as the core analysis and conclusion of our theoretical model on size effect is unaffected by the differential exposure.

[4] Our imposition that size and group composition are orthogonal components is for analytical purpose. We impose this assumption so as to analyze the distinct effect of total size net of the effect of group composition. In reality, size and group proportion may be correlated: larger school may also tend to have a more diverse student body. However, our analytical simplification does not undermine the theoretical argument of this paper that size matters regardless of group composition.

(1) The relative representations of different social groups in the context. For group g (g = 1, …G), let us denote its proportion by P_g, with $\sum_{g=1}^{G} p_g = 1$.

(2) The total number of persons across all social groups in the context. Let us denote the number by *N*.

We call the first component "group composition" and the second "total size." In the current literature, the effect of contextual group composition on relationship choices is well recognized. Why does group composition matter? Theoretically, "social associations depend on opportunities for social contacts" (Blau 1977, p.27), so that, controlling for group preference, the likelihood of forming intergroup ties with group g should be proportional to p_g. Consider the hypothetical case in which an individual is indifferent in choosing between a black friend and a white friend; thus, his/her choice of friend completely depends on the chances of exposure. Imagine a scenario in which the population is composed of 30% blacks and 70% whites. In the case, this race-indifferent person's probabilities for choosing a black friend and a white friend would be respectively 30% and 70%, a pure function of racial composition in the population. In Wimmer and Lewis's (2010, p.590) words, "the smaller the relative size of a group, the more likely its members will form out-group ties, under ceteris paribus conditions."[5]

Besides simple group proportions, researchers have also derived measures of multiple-group composition, among which heterogeneity is the most widely used. Heterogeneity (*H*) is defined as $[1 - \sum_{g=1}^{G} p_g^2]$, where p_g denotes the proportion of each group in the choice set (Gibbs and Martin 1962). It measures the likelihood that a randomly chosen pair in the population contains people from two different groups. Heterogeneity is expected to increase the likelihood of intergroup relations, because it increases opportunities for individuals to meet out-group alternatives (Blau 1977, 1994; Blau and Schwartz 1984).

Corresponding to this theoretical understanding is now a sizable empirical literature in sociology that documents the effects of group composition. Below, we briefly review studies in two areas that attract significant sociological interests: marriage and friendship.

[5] It is implicit in this argument that total size *N* is considered among the "ceteris paribus conditions."

The pairing of marital partners with certain characteristics is an important indicator of a society's intergroup relations. A higher level of homogamy based on such attributes as socioeconomic status, education, and race/ethnicity is generally considered indicative of low social openness (Kalmijn 1991, 1998; Mare 1991; Qian 1998; Schwartz 2010; Torche 2010). Marital choices depend on people's preferences in mate selection as well as the availability of alternatives in the local marriage market. Researchers have found preferences for marital partners to be multidimensional and have accordingly modeled the structural determinants of marital choice on these multiple dimensions. For example, Blau and his associates have established and found support for the hypothesis that greater heterogeneity in racial and ethnic origins, region of birth, industry, and occupation is associated with higher rates of intermarriage (Blau et al. 1982; Blau and Schwartz 1984; Blum 1985; Blau 1994). More recently, with growing availability of micro-level and macro-level linked data, studies on marital choice have operationalized the local marriage market by considering the joint distributions of age, sex, race, education, religion, and socioeconomic status as structural determinants of individual choice (Kalmijn and Tubergen 2010; Lichter et al 1995; Lewis and Oppenheimer 2000; Oppenheimer 1988; Mare 1991; Qian 1998; Rosenfeld 2005). However, this line of research focuses exclusively on relative group size (p_g) as the key structural factor. For example, as Kalmijn and Tubergen (2010, p.471) state, "the greater the relative size of a group in a state, the more often the children of immigrants marry endogamously."

Like marital choice, friendship choice has also been considered an important area in sociology, as it provides a concrete measure of intergroup social boundaries at the level of ordinary behaviors (Feld 1981; Hallinan and Williams 1989; Moody 2001). The launching of the National Longitudinal Study of Adolescent Health (Add Health) in 1994-95 has led to a proliferation of new research on the structural effects of schools on interracial friendship choice (Currarini et al. 2010; Joyner and Kao 2000; Kao and Joyner 2004; Moody 2001; Mouw and Entwisle 2006; Quillian and Campbell 2003; Zeng 2004). An attractive feature of the Add Health study for this new wave of studies is that it included, at the baseline survey, detailed school-level contextual measurements as well as individual friendship nomination records. Table 1 gives a brief summary of results from four major articles in this line of research. From this table, we can discern two compositional measures that are most commonly adopted in this area: relative group proportion and heterogeneity. Findings on relative group size generally corroborate Blau's (1977, 1994) hypothesis (e.g., Quillian and Campbell 2003). In the case of heterogeneity, evidence from Add Health data suggests that heterogeneity positively affects interracial friendship both because it promotes same-race preference and because it increases the opportunity to meet cross-race alternatives (Moody 2001).

In sum, in both the marital and the friendship choice literatures, as well as in other related works, prior researchers have consistently focused on the structural effects of group composition to the neglect of total size as a structural determinant of social relationship formation. This is understandable, because group proportions can be interpreted as group sizes when total size is controlled. Indeed, the standard practice in the literature on friendship has been to equate group proportion effects with group size effects.

However, we hasten to add that total size has not gone totally unnoticed in past research. Friend choice within school contexts is a good case in point. For example, Hallinan and Williams (1989) long ago pointed out that "students in large schools, grades, classes, or ability groups have a wider range of peer contacts than those in smaller organizational units." More recently, Moody (2001, p.697) remarked, "Students have more potential friends to choose from in large schools; tendencies for self-selection suggest that friendship segregation will increase with school size."[6] Still, their concern with "the size of a group" was ultimately reduced to an emphasis on "the composition of the group [that] defines the characteristics of those peers with whom a student has chance encounters" (Hallinan and Williams 1989, p.68).[7] What is particularly surprising is that there has been little theorizing about the structural effects of total size. While researchers have always included total size in their empirical analyses, they have treated it more or less as a statistical nuisance or, at best, a statistical control (Joyner and Kao 2000; Moody 2001; Quillian and Campbell 2003; Wimmer and Lewis 2010).

---

[6] Moody (2001) included school size as a control variable in modeling friendship segregation, but found no significant effect.

[7] By "the size of a group" they mean the total size of the population in an organizational unit, and thus this phrase refers to the total size of a social context.

Table 1. Summary of selected studies on intergroup friendship using Add Health data

Article	Statistical Model	Outcome Variable (Y)	Selected Variable(s) for School Context (X)	(Average) Coefficient on X	Standard Error
Moody (2001)	OLS Regression	Same-race friendship preference	Racial heterogeneity (H)	-5.32**	1.88
			H^2	15.55**	4.58
			H^3	-11.72***	3.30
			School size	-0.04	0.04
		Log odds of a same-race friendship nomination	Racial heterogeneity (H)	-7.99**	2.13
			H^2	22.03***	5.57
			H^3	-15.96***	4.23
			School size	-0.10	0.07
Quillian and Campbell (2003)	P* Model	Log odds that an own-group dayd will be a friendship dyad	School 10-30% Own-Race × Receiver Receiver own-Race	-1.1***	--
			School 30-60% Own-Race × Receiver Receiver own-Race	-1.303***	--
			School 60-100% Own-Race × Receiver Receiver own-Race	-2.21***	--
			Log of school size	-0.755***	--
Zeng (2004)	Two-stage multi-level model	Coefficient of interracial preference (β) estimated from the within-school model	School diversity	-0.35*	--
			Target group size	-0.68*	--
			School size (square root)	-0.182*	0.059
Mouw and Entwisle (2006)	P* Model	Log odds of a friendship dyad	School diversity × interracial dyad	-1.884***	--
			Log of school size × Reciprocity	0.184**	--
			Log of school size × Transitivity	0.027*	
			Log of school size × Expansiveness	-0.048**	--
			Log of school size × Popularity	-1.414**	--

* P<0.05; ** p<0.01; ***p<0.001

Notes: Some studies present multiple models predicting the same outcome variable with different sets of controlling variables. To keep parsimonious, in our table, we only present the results from the model with the most controlling variables. For school context measures, we only select the independent variables that are particularly relevant to our major focus: group proportion, heterogeneity and school size. Some studies stratify samples into subgroups; in those cases, for simplicity we only report the means of the significant coefficients across all subgroups, and we mark the significant levels by the highest significance level reported in the article. Standard errors are not reported for the means of coefficients.

For example, in the P* model, Mouw and Entwisle (2006) interacted school size (measured in the log of the number of students) with network variables, but these do not tell about the interaction between school size and the likelihood of interracial dyads. [8] In an unpublished paper, Zeng (2004) included the square root of school size and found a significant negative effect, but she interpreted the result to mean that size affects in-group preference rather than exerting a structural effect.[9] In recent years, dynamic agent-based models in network analysis enable researchers to assess the validity of various theories concerning the size effect (Snijders 2001; Snijders, Bunt and Steglich 2010). Grund (2010), in an unpublished paper using a dynamic model of network formation, reports that the absolute network size has a direct impact on the observed homophily. His results are further confirmed by Flashman (2012), who found that the pool of possible friends in a community enhances the chances of adolescence in selecting similarly achieving friends. In a study of residential segregation using dynamic agent-based models, Fossett and Dietrich (2009) found that segregation outcomes do not vary much by city size. Our purpose in this paper is to amend this intellectual imbalance by arguing for a structural effect of size while holding everything else constant. That is to say, our study departs from an exclusive emphasis in the prior literature on group composition and shows a distinct effect of total size on intergroup relations.

2. Why Worry About Size?

To derive the effect of total size, our point of departure is the proposition that individuals' preferences in social relations should be based on factors that are multidimensional in nature. Among these preference dimensions, some are observable by both the individual actor and the researcher, such as age, race and educational attainment, while some are known to the individual actor but not to the

[8] The authors interpret coefficients on these interactions as such: "The log of the number of students in the school (in the full in-school sample, standardized to the mean) is included as a predictor of the coefficients of the social network variables, under the assumption that network effects will be larger in big schools"(Mouw and Entwisle 2006, p.420).

[9] Another article by Currarini et al. (2010) that decomposed the sources of homophily into biases in preferences and biases in meeting rates also parameterized meeting opportunity by relative proportion of each group and found a significant effect. They interacted school size with the two sources of bias and pointed out that "it is interesting to note that larger schools exhibit significantly higher biases in the rates at which students meet students of their own race." Certainly, this finding suggests that future studies should recognize school size as a non-trivial factor, but like the others, it did not go beyond this finding and examine the distinct effect of total size.

researcher, such as personalities, hobbies, cultural tastes, and political ideology. In the following paragraphs, we first explain the motivation and necessity for assuming multidimensional preference, particularly the inclusion of unobserved characteristics in relationship formation. Then we explicate the link between multidimensional preference and the effect of total size.

A line of social theories motivates the rationale for assuming multidimensional preference. Sociologists like Simmel (1955), Goffman (1971), and Breiger (1974) have all emphasized, though from different perspectives, that members of a society are affiliated with a spectrum of social groups on multiple social dimensions. Relationship formation is a process in which people choose other people, their choices being based on the consideration of these dimensions. In line with theoretical arguments, empirical studies have incorporated multiple dimensions, such as race, age, educational attainment, socioeconomic status, religion, and cultural taste in marriage models (e.g. Kalmijn and Tubergen 2010; Rosenfeld 2005), friendship (e.g., Quillian and Campbell 2003; Rytina and Morgan 1982; Wimmer and Lewis 2010), residential choice (e.g., Massey and Denton 1993). On the flip side, failure in recognizing the multidimensional nature of preference can lead to misinterpretation of observed patterns (McPherson, Smith-Lovin, and Cook 2001; Wimmer and Lewis 2010). Consider the possibility that people's cultural taste is associated with their education level. In such cases, we should not attribute education homogamy in marriage to educational preference alone, because it may result as a mixture of in-group preferences for education and for cultural taste.[10] The recognition of multidimensional preference proves crucial to our theoretical model of relationship formation.

[10] Two other potential confounders in multidimensional preference are the tendency to reciprocate friendship and the tendency to befriend someone with whom one shares a common friend. The former is called "reciprocity" and the latter is called "triadic closure." Wimmer and Lewis (2010) called both tendencies "balancing mechanisms." In fact, our multidimensional preference assumption encompasses such possibilities, as we can view reciprocity and triadic closure as people simply including "whether the alternative treats me as a friend" and "whether the alternative and I have a friend in common" as two additional dimensions in their preference space which we call "effects of network structure." Models in social network analysis, among which P*(exponential random graph) models are best-known, have offered methods of incorporating these effects (see Wasserman and Faust 1994 and Snijder 2002). While we recognize the importance of effects of network structure in relationship formation, we exclude them from our theoretical model so as to avoid digression from our core analysis. Still, we believe that the size effect proposed in this paper remains valid, if not amplified, in the presence of such effects.

Not only is preference multidimensional, but some dimensions important to the individual decision maker shall remain unknown to the researcher. We make this assumption because we do not believe that standard social science measures shall ever exhaust all the different variables taken into account by the social actor, including, for example, personality, physical appearance, hobbies, cultural tastes, religious affiliations, and political beliefs. For this reason, there will always be unobserved dimensions. For K-dimensional preference, we assume that we only observe a strict subset of J dimensions, so K is always greater than J. Let us write $L=K-J$, in which L denotes the number of unobserved dimensions. Such unobserved traits may be elicited by the individual actor through repeated interactions with a potential partner, such as dating and participating in common social activities, so that an informed choice can be made, which distinguishes them from pure noise in choice behavior. Yet, like random noise, these traits are not readily observable by the researcher, as they represent unobserved population heterogeneity, "the toughest challenge in today's quantitative social science and statistics" (Xie 2007).

The consideration of unobserved dimensions in preference is not just of pedantic interest for modeling social relation formations. It actually has practical implications. Suppose a policy-maker commits him- or herself to enhancing interracial friendship in high schools, with a specific interest in discovering what type of school context can best promote racial integration. Admittedly, other dimensions besides race are involved in friendship choice. Yet, it is neither practical nor necessary to study how school context influences friendship choice in all relevant dimensions. The objective for this policy-maker is concerned solely with how school characteristics affect the observed level of segregation on the dimension of race, aggregated over all other dimensions.[11] In this sense, all other dimensions except race are considered unobserved.

[11] Statistically speaking, the policy maker's analysis is "marginalized" over other dimensions. For example, in analyzing categorical data, the row and column totals in the contingency table present the marginal distribution of the row and column variables, and a log-linear model allows the calculation of "marginal means" (Xie and Powers 2008).

As we have posited earlier, multidimensional preference is essential to our proposition of total size effect. We present the intuition of our argument as follows. An individual starts by assigning a weight to each dimension that represents its relative importance in determining the individual's preference. Through weighting, multiple dimensions of relevant factors form a single summary measure of utility associated with choosing each potential alternative as a partner. We further assume that the individual forms a relationship with the alternative who yields the highest utility *(rational actor assumption)*. A small community restricts the individual's choice set of alternatives, so matching on all dimensions of preference is unlikely to be realized, and some compromise is likely. If occurring, a compromise is likely to be on all relevant dimensions. In fact, finding a perfect match is nearly impossible in reality if preference is of high dimensions.[12] As a result, compromises are inevitable for all social actors. When total size increases, given the rationality to maximize utility, individual actors will take advantage of the enlarged choice set by forming relationships that yield the highest utility for themselves. Thus, chances of compromise decline in all relevant dimensions. In this way, total size acts to minimize compromises that go against individuals' preexisting relationship preferences. Viewed from a population perspective, an increase in total size raises the average share of highly satisfactory matches on all relevant preference dimensions.

Let us consider an example of friendship formation. Suppose one's preference for a friend has two dimensions: the observed dimension, race, and the unobserved dimension, personality. Further suppose that individuals all prefer friends who are similar to themselves along both dimensions (homophily). Let racial segregation be our primary concern. We assume for simplicity that race is binary, white versus black, but there are many categories of personality. In a small community, the chances of finding a good match on personality within one's own racial group is low, giving rise to the possibility of a compromise, either in the dimension of race or in the dimension of personality, in exchange for a match in the other dimension. In a larger community with the same racial composition, a person faces a larger

[12] To see this, consider a hypothetical case in which a person wants to find a match on twenty independent binary dimensions. There are as many as 2^{20}=1,048,576 different combinations of traits, which may far exceed the size of his/her actual choice set.

choice set, so the chances are better for him/her to find a good match on personality within his/her own race. As a result, everything else being equal, a higher level of racial segregation would occur in larger communities.

In the remainder of this paper, we will first construct our theoretical framework and then illustrate its implications for realistic situations in two ways. First, we will demonstrate the total size effect on the aggregate level of friendship racial segregation. Note that the number of potential dyads expands rapidly with an increase in size – for example, from 90 to 9900 when total size increases from 10 to 100. It is thus not possible to express aggregate outcomes through closed mathematical formulas. Thus, we use micro-level simulation, also called "agent-based modeling," to explore the aggregate outcome resulting from individual choices. Agent-based modeling not only facilitates the computation of aggregate measures from individual behaviors, but also allows us to systematically explore how the size effect depends on parameter specifications, such as the number of preference dimensions and the strength of preference. In Section 4 we present agent-based modeling results based on our theoretical model.

Second, in Section 5, we use actual friendship nomination data from the National Longitudinal Study of Adolescent Health in-school samples to evaluate our theoretical proposition about a size effect. We report results consistent with our prediction that, after controlling for group compositions, total school size is negatively associated with one's likelihood of choosing an interracial friend in a school.

3. Size Matters: A Theoretical Model

Our assertion that total size matters is a theoretical statement that can be made and defended on theoretical grounds alone. In this section, we establish a formal theoretical model that shows the effect of total size on relationship formation. For the purpose of illustrating the size effect, we focus on friendship formation, with the primary outcome of interest being friendship racial segregation, measured by the share of interracial friends in a closed context. We simplify the process of friendship formation so that friendship results from each agent making a choice regarding with whom to form a relationship.[13] We call the decision maker the *agent*, and the people at the risk of being chosen the *alternatives*. While we realize that reciprocal ties are also important to social relations, we focus on directional ties in this study to ensure parsimony of our demonstration. This means that multiple agents can choose the same person as a friend, and that the person being chosen does not have the right to reject the choice. Although the second implication seems naive, it actually corresponds to the situation of the empirical data that we will analyze later in the paper. One way to consider the role of rejection is that the agent incorporates all information, including discouraging signals, about each alternative when evaluating each alternative's desirability as a friend.

With this simplification of friendship formation as a choice problem, we further distinguish between latent preference and realized choice. Latent preference represents an agent's subjective evaluation of utility gained from social relations over a set of alternatives. Realized choice is the actual partner whom the agent chooses out of the set of alternatives. The former is unobserved, and the latter observed. As Zeng and Xie (2008) argued, the major reason for distinguishing latent preference and realized choice involves the role of structural constraints. It is commonly assumed that latent preference is independent of structural constraints and invariant over the choice process, as its existence precedes one's entrance into any structural context. Unlike latent preference, realized choice is subject to the availability constraint imposed by social structure. In the remainder of this section, we establish a theoretical framework that derives realized choice from latent preference and structural constraint. We close this section with two propositions regarding the effect of total size to be tested by agent-based modeling in the next section.

3.1. Latent Preference

Independent of structural constraints, latent preference represents the agent's inherent assessments of other persons' desirability as a friend. We can define it through a thought experiment:

---

[13] While greatly simplifying the model, the assumption of directional choice is inconsequential for the core result in this paper.

Suppose that the experimenter offers a high school student a hypothetical choice between a white friend and a black friend in his or her high school who are identical on every dimension except race. This student's choice reflects his/her latent racial preference, and it is not affected by who is available to be a friend in the school context.

As in standard choice models in economics, we represent latent preference by a utility function.[14] The main reason for invoking the utility function is to quantify the desirability (i.e., utility), affected by multiple dimensions, in a uni-dimensional space. In other words, the utility is the quantitative expression of preference. We say that the agent "prefers A over B" if and only if the utility of A exceeds that of B. The agent is "indifferent between A and B" if and only if the utilities associated with the two alternatives are identical. As such, choice decision results from the comparison of the relative utilities associated with different alternatives. In our model, we specify utility as a function of two main components: a systematic utility *V* and an idiosyncratic disturbance ϵ. We further assume that the two components are additive and separable.[15] Thus, we write the utility for "agent *i* choosing alternative *j*" as:

$$U_{ij} = V_{ij} + \epsilon_{ij} \ . \tag{1}$$

The parameterization of the systematic utility V_{ij} is central to our model. As we have argued earlier, the agent's preference in social relations should be multidimensional. Therefore, V_{ij} should be a function of components in multiple dimensions. Suppose that the agent's preference contains *K* dimensions, then we can represent the utility as a function of a vector $Z_{ij} = (Z_{ij1}, \dots Z_{ijK})'$. We call a "covariate" each element in Z_{ij}, measured specifically for the agent (*i)* and the alternative (*j)*. We further

[14] In formal language, a preference relation can be represented by a utility function only if it is rational (Mas-colell, Whinston and Green 1995). Since it is not the focus of this article to distinguish preference from utility, we use the two terms interchangeably.

[15] We assume that in the population, ϵ's are independently and identically distributed and independent of the systematic component of utility. In statistical models of discrete choice, the scale of utility is normalized with respect to the variance of the random disturbance. In the case of conditional logit models, for example, the random disturbance follows the type I extreme value distribution with the probability density function: $f(\epsilon) = \exp(-\epsilon) \cdot \exp(-\exp(-\epsilon))$. The variance of ϵ is normalized to $\pi^2/6$.

add a parameterization of the uni-dimensional measure of utility as a function of the *K*-dimensional covariates. We follow a long tradition in both economics and sociology in employing a linear function:

$$V_{ij}(z_{ij}) = \sum_{k=1}^{K} \beta_k z_{ijk} \ . \qquad (2)$$

In this equation, z_{ijk} is the value of the *k*th covariate, and the corresponding β_k is the weight of this covariate in determining the utility. We note that the linear function form is more powerful and flexible than it may appear, as a variety of functional forms can be easily expressed as a linear function after proper transformation, such as polynomial functions, spline functions, and step-functions. Alternatively, we may express equation (2) as:

$$V_{ij}(z_{ij}) = V_{ij1} + V_{ij2} + \cdots + V_{ijK}, \qquad (3)$$

where $V_{ijk} = \beta_k z_{ijk}$, representing the contribution of the *k*-th dimension to the overall utility.

We reiterate that covariates Z_{ijk} $(k = 1, \ldots K)$ all pertain to measures specifically for agent *i* and alternative *j*. The model is general enough in allowing various forms of Z_{ijk}. If, as a special case, there is a "fixed effect" of agents (or alternatives) independent from the attributes of alternatives (or agents), then the corresponding covariate z_{ijk} shall depend exclusively on agent *i* (or alternative *j*).

Corresponding to our primary interest in friendship racial segregation, we parameterize a covariate ($Z_{ij\text{race}}$) to represent racial homophily, with $Z_{ijrace} = 1$ if agent *i* and alternative *j* share the same race, and $Z_{ijrace} = 0$ otherwise. For our main argument about a positive size effect on racial segregation in friendship, all that is needed is the assumption of racial homophily. For other dimensions, it is sufficient that they contribute to the utility in some specific forms as in equation (2). However, for the sake of discussion, we assume in this paper homophily in all other dimensions and thus use only one specific form of covariates -- the absolute social distances in attributes between the agent and the alternative. That is, $Z_{ijk} = |Z_{ik} - Z_{jk}|$, where Z_{ik} and Z_{jk} are respectively attributes of the *i*th agent and the *j*th alternative in the *k*th dimension. We base this specification of covariates as social distances on a prevailing sociological understanding that homophily -- the preference for alternatives with similar social attributes -- exists in a range of social dimensions, including race/ethnicity (McPherson et al. 2001;

Shrum, Cheek, and Hunter 1988), religion (Fischer 1977; Laumann 1973), age (Fischer 1977; Fischer 1982; Burt 1990), education (Marsden 1987; Yamaguchi 1990), and social class (Wright 1997). Yet we note that our parameterization of all other covariates as social distances in attributes is a mere illustration of other factors existing in addition to racial homophily. Instead of homophily, alternative forms of preference may be parameterized in other dimensions, such as preference for friends of higher socioeconomic status than one's own. Alternative functional forms for them would not change our main conclusion.

For the *k*th covariate in equation (2), there is a corresponding β_k for mapping the covariate onto the uni-dimensional utility. Note that β_{k} measures the weight attached to the *k*th preference dimension. Under homophily, we assign negative weights to all dimensions, because utility decreases with the increase in social distance between the agent and the alternative. The ratio between any pair of β's, say β_k and $\beta_{k'}$, with $k \neq k'$, quantifies the tradeoff in preference between covariates k and k'. That is, the ratio $|\beta_k/\beta_{k'}|$ measures the amount of utility in covariate k' that the agent is willing to sacrifice in exchange for a unit increase of utility in covariate *k*.[16] To be concrete, suppose the ratio in the value of β between racial distance and SES distance, $(\frac{\beta_{race}}{\beta_{SES}})$, equals three. This implies that the agent is willing to give up a same-race friend in exchange for an other-race friend if the latter is at least three units closer to him/her on the SES dimension than the former.

Finally, let us discuss a key assumption on latent preference in our theoretical model. We assume that the agent's utility associated with an alternative is unaffected by other alternatives in the choice set. In discrete choice models, this is called the "independence of irrelevant alternatives" (IIA) assumption. In friendship choice models, the IIA assumption also means the absence of peer interferences across

[16] This ratio has the same property as *the marginal rate of substitution* in the economic demand theory. In economic policy problems, the ratio $\beta_{\mathrm{k}}/\beta_{\mathrm{k}'}$ is called *willingness to pay* (McFadden 2001), which represents the maximum amount a person would be willing to pay, sacrifice or exchange in order to receive a good or avoid something undesired. While the ratio among $\beta_{\mathrm{k}}'s$ carries substantial meanings, the absolute values of $\beta_1, \ldots \beta_K$ are not informative, because multiplying $\beta_1, \ldots \beta_K$ and ϵ_{ij} by the same scalar does not affect ranking of utility among alternatives.

difference persons. That is, not only is each person's utility associated with an alternative determined by unique, pairwise-measured covariates $Z_{ijk}'s$, the utility is also unaffected by the presence and preferences of other agents. Statistically, we make use of this assumption by requiring that the $\epsilon_{ij}'s$ are mutually independent across all *i*'s and *j*'s.

3.2. Realized Choice

In our theoretical framework, we require that latent preference remain invariant across social contexts. However, latent preference is unobservable in reality. Instead, we observe actual choices realized by agents and make inferences about latent preferences through realized choices. This is often called the "revealed" preference approach. One important sociological insight is that realized choices actually differ from latent preferences, because realized choices are affected by opportunity constraints imposed by the contextual social structure (Zeng and Xie 2008). To illustrate, let us go back to our earlier example of a thought experiment, in which an experimenter would ask a respondent to choose between a black and a white as a friend. Now we add the realism that the person may also prefer many other attributes besides race in a friend. Due to this demand for a high-dimensional match, this person may not realize his racial preference while satisfying his/her other preferences for a friend because social context constrains his/her exposure to limited alternatives to choose from so that there is no perfect match. More importantly, he/she may resolve the structural situation of no perfect match on all dimensions by compromising his/her racial preference, i.e., befriending a person from the less-preferred race. In this paper, we show that total size affects interracial friendship formation by imposing this structural constraint.

How does total size affect realized choice? Formally, suppose there are *N* people in a closed context, and agent *i* chooses alternative $j^*(j^* \in (1, \dots N)\ \&j^* \neq i)$ as his/her friend. Let us use ij^* to denote a *matched pair*, in which agent *i* chooses alternative j^*, with its associated utility U_{ij^*}. For convenience, we further denote $U_{ij}^* = U_{ij^*}$. We invoke the behavioral assumption that is essential to the size effect (*rational actor assumption*): the agent pursues optimal utility in social relations and chooses the alternative with the highest utility in his/her choice set:

$$U_{ij}^* = U_{ij^*} = \max\{U_{ij} | j \in \{1, \ldots N\}, j \neq i\}. \quad (4)$$

Recall again that our model of friendship choice does not preclude different agents from choosing the same person. Thus, all agents maximize their utility in choosing their friends regardless of other agents' choices. Let us now ask a question about the expected value of U_{ij^*} in the closed context. Holding everything else unchanged, how does the expected utility of *the matched pairs* in the population, expressed by $E(U^*)$, vary with population size *N*? The answer is a positive relationship. To explain, we draw a distinction between the *unconditional* expectation of U_{ij} in the population ($E(U)$) and the expectation of U_{ij} *conditional* on the matched pairs ($E(U^*)$). The unconditional expectation is the average utility over all *potential combinations of pairs*, which add up to a total of $(N \cdot (N-1))$ directional pairs; the conditional expectation is the average utility over a total of *N matched pairs*:

$$E\left(U_{ij}^*\right) = E\left(U_{ij} \middle| (i, j) \text{ is a matched pair}\right). \quad (5)$$

Due to our rational actor assumption, the agents always make their choices at the highest utility level available. It follows that the expected utility conditional on matched pairs always exceeds the unconditional utility, i.e. $E(U^*) > E(U)$. An increase of the total size *N* raises the agent's likelihood for finding a more preferable friend – i.e., one that yields higher utility within the context. Once an agent befriends a more preferable friend, the utility associated with this matched pair increases. Due to such a mechanism, increasing total size *N* raises the expected utility conditional on realized matches ($E(U^*)$), because the agent can only improve his match with a larger choice set, while the unconditional expected utility over the choice set ($E(U)$) remains unaffected.

Next, let us follow with a second question: given that $E(U^*)$ increases with total size N, where does the increase in utility come from? The answer is that an overall increase in utility is shared over all relevant dimensions in the preference function. To see this, recall equation (3), which decomposes the systematic component of utility (V_{ij}) into additive dimensions of covariates. The summation of the *K* dimensions is commutative and associative, meaning that neither alternating their order nor changing the sequence of performing the summations affects the overall utility. If there is an increase in $E(U^*)$, it shall come from the increase in utility on every dimension. In our case, because we specified the systematic components of the utility function with homophily functions, $V_{ijk} = \beta_k . |z_{ik} - z_{jk}|$, a utility increase can only come from the shortened distance between z_{ik} and z_{jk}. As a result, we expect to observe higher likelihood of in-group ties on every dimension, including race, in a larger context.[17]

Now we summarize our answers to the first and second questions. Our answer to the first question suggests that the larger the total size, the higher the expected utility conditional on the matched pairs. Our answer to the second question suggests that the increase in overall utility for the matched pairs should be attributed to utility increases in all preference dimensions. Combining the two answers together, we arrive at the conclusion that under racial homophily, an increase in total size *N* decreases the share of interracial friends, and thus increases the level of friendship racial segregation in the context.[18]

What will happen if we increase size *N* to infinity? Surprisingly, we will not see complete segregation. Instead, the limiting share of intergroup ties depends on the distribution assumption about the error term. Following the tradition of conditional logit model, which assumes the error term ϵ to follow i.i.d. type I extreme value distribution (i.e. Gumble distribution), we now examine the limiting share of intergroup friendship as *N* approaches infinity. Consider the case with two racial groups, whites with group proportion p, and blacks with group proportion $(1-p)$. First, we transform the behavior model of latent utility into a statistical model of choice probability. It has been proven that the probability that agent (*i*) befriends an alternative (*j*) can be expressed as (McFadden 1974):

$$\Pr(Y_i = j | Z_i) = \frac{\exp(Z'_{ij}\beta)}{\sum_{l=1, l\neq i}^{J} \exp(Z'_{il}\beta)} \quad . \tag{6}$$

[17] An attentive reader may have noticed that our analysis here does not concern the random component of utility function. In fact, when the rational actor maximizes the utility from a social tie, he/she maximizes both the systematic utility and the random utility. As we set the two components to be additive and separable in the utility function, the same analysis applies to both. But the random disturbance is unobserved, so it remains marginal to our major analysis.

[18] Certainly, if we do not assume homophily in dimensions other than race, then we can still conclude that an increase in total size *N* increases the level of racial segregation, regardless of other dimensions.

Although there are multiple covariates in the conditional logit model, our focus is on interracial friendship. Let covariate Z_{ijrace} equal zero if the agent and the alternative share the same race, and one if their races differ. When we increase size to infinity, the share of interracial friends will approach a limit that depends on only two quantities: β_{race}, the weight of the race covariate, and *p*, the proportion of whites in the population. The likelihood of a white choosing a black friend (π_1) approaches:

$$\pi_1 = \frac{(1-p)\cdot\exp(\beta_{\text{race}})}{p+(1-p)\cdot\exp(\beta_{\text{race}})}\,. \tag{7}$$

Blacks' likelihood of choosing a white friend (π_2) approaches:

$$\pi_2 = \frac{p\cdot\exp(\beta_{\text{race}})}{(1-p)+p\cdot\exp(\beta_{\text{race}})}\,. \tag{8}$$

As for the population average, the overall share of interracial friends (π) under infinite total size is the weighted sum of π_1 and π_2 by their respective group proportion:

$$\pi = p\cdot\pi_1 + (1-p)\cdot\pi_2\,, \tag{9}$$

where π_1 and π_2 take the values as expressed in Equation (7) and (8). This results shows that if we increase size to infinity, a situation when total size no longer imposes structural constraint on in-group matching, there will not be complete segregation. Noise in our model plays such a role that a nonzero share of ties shall remain interracial.

3.3 Testable Propositions

We have established above that total size *N* negatively affects interracial friendship. Next, we explicate two testable propositions about the strength of this size effect.

First, how does the effect of total size depend on the dimensionality of covariates that affect preference? As we have suggested earlier, agents consider multiple dimensions in making a choice. The higher the preference dimensions, the more difficult for an agent to find a perfect match on all dimensions, and thus the more likely it is that he/she will compromise over the dimensions in making a choice. When total size is increased, this structural constraint is relaxed, resulting in a reduction in the likelihood of making such compromises. Thus, we shall see a larger size effect associated with higher dimensions in the comprising of agents' preference.[19] We summarize this argument in Proposition 1:

> ***Proposition 1:*** The effect of total size increases as the number of dimensions in individual preference increases.

Second, how does the effect of total size depend on relative weights of different covariates affecting preference? In our setting, we are primarily interested in interracial friendship. Suppose that the preference is strongly affected by racial homophily, relative to other factors. This is represented by a large absolute value of β_{race} relative to other β's in our model. In this situation, when the agent does not find a perfect match and needs to make a compromise in choosing a friend, it is less likely that he/she will give up racial preference, because the agent gives this racial homophily dimension more weight than other dimensions. Thus, size effect on racial segregation would be small. So we come to the second proposition:

> ***Proposition 2:*** The effect of total size decreases as the weight of the covariate representing the dimension of primary interest becomes higher.

In the next section, we use agent-based modeling based on our theoretical framework to test the two propositions stated above. We will show that the agent-based modeling results are consistent with our theoretical arguments.

[19] Yet, the effect of total size exists as long as there are two or more dimensions in the latent preference.

4. Agent-Based Modeling

We conducted two experiments of agent-based simulation according to our theoretical model so as to test the two propositions about size effect. We followed the five steps below.

First, we set up the choice set. The context contains *N* people. Consider four dimensions of covariates: race, age, family background, and academic performance. We assume two racial groups, with the proportion of the majority group set at 80% and the minority group at 20%. The individual attributes on the latter three dimensions are assumed to follow a continuous scale and are identically distributed. Thus, the names of the other three dimensions are interchangeable and inconsequential. In reality, covariates on different dimensions may be correlated with each other (Blau and Schwartz 1984; Wimmer and Lewis 2010), but their correlations are inconsequential to our model. To preserve parsimony, we assume these covariates to be mutually independent.[20]

Second, we set up the individual utility. As discussed earlier, we specify that all the ovariates affect friendship choice through homophily. The latent utility can be written as:

$$U_{ij} = \beta_0 + \beta_1 Z_{ij1} + \beta_2 Z_{ij2} + \beta_3 Z_{ij3} + \beta_4 Z_{ij4} + \epsilon_{ij}, \qquad (10)$$

where $Z_{ijk} = |Z_{ik} - Z_{jk}|$ for *k* =1,2,3,4, representing the absolute distance between the agent's and the choosee's individual attribute. β_0 is an intercept that stays the same for every *ij* pair.[21] Because we assume homophily on every dimension, $\beta_k \leq 0$ for *k* =1,2,3,4. Panel A in Table 2 gives the covariates distribution in the population, and Panel B and Panel C in Table 2 gives the preference coefficients for the two experiments respectively. In the utility function, the random component ϵ_{ij} follows i.i.d. type I extreme value distribution. The variance of ϵ_{ij} equals $\frac{\pi^2}{6}$. It represents the non-deterministic component in choice model.

[20] Strictly speaking, their correlation is inconsequential to our model so long as the dimensions are not perfectly collinear. In fact, if two dimensions are perfectly collinear, in this model we will not consider them as two distinct dimensions.

[21] Although β_0 has meaning in a latent utility function, it will not affect the choice outcome, because it does not affect the ranking of utility among the alternatives for any agent.

Table 2. Covariates distribution and preference coefficients for agent-based models predicting average share of interracial friends by school size

Panel A Covariates distribution in agent-based models

Covariate	Distribution	Distribution Parameters	Parameter Specification
Race (Z_1)	Binomial	*p* (proportion of the minority group)	*p*=0.2
Age (Z_2)	Normal	μ (mean) ; σ (standard deviation)	$\mu = 0; \sigma = 3$
Family Background (Z_3)	Normal	μ (mean) ; σ (standard deviation)	$\mu = 0; \sigma = 3$
Academic Performance (Z_4)	Normal	μ (mean) ; σ (standard deviation)	$\mu = 0; \sigma = 3$

Panel B Preference coefficients in Experiment 1 representing different numbers of preference dimensions

	β_1	β_2	β_3	β_4
Round 1.1	-6	0	0	0
Round 1.2	-6	-1	0	0
Round 1.3	-6	-1	-1	0
Round 1.4	-6	-1	-1	-1

Panel C Preference coefficients in Experiment 2 representing different strength of racial preference

	β_1	β_2	β_3	β_4
Round 2.1	-2	-1	-1	-1
Round 2.2	-4	-1	-1	-1
Round 2.3	-6	-1	-1	-1
Round 2.4	-8	-1	-1	-1

Notes:
1. The four covariates in Panel A are independent and identically distributed.
2. The four coefficients $\beta_1, \beta_2, \beta_3$ and β_4 in Panel B and Panel C correspond to Z_1, Z_2, Z_3 and Z_4 in Panel A respectively.

Third, we simulate the choice behavior of every person in the context using an agent-based model. For each agent, his/her choice set of friend contains every other person in the population. Consistent with our assumption of rational actors, each and every agent *i* chooses friend *j* who yields the highest U_{ij} , which is calculated uniquely for every combination of *i* and *j*. Because their friend choices are non-exclusive -- i.e., multiple agents can choose the same friend -- the sequence of agents' choices does not matter.

Fourth, we iterate the agent-based modelling process 100 times for each setup. For each round of iterations, we calculated the share of interracial friends, which equals the number of interracial ties divided by the number of people in the context. Then we average the share of interracial friends over the 100 iterations as an estimate of the expected share of interracial friends under this setup.

Last, we repeat this process for different setups. The primary independent variable is total size *N*. We vary N from 50 to 1000, by an interval of 50, with a total of 21 data points. We then plot a curve of expected share of interracial friends by total size *N*. The slope of the curve is an indicator of the size effect. In all setups, we fix the relative group proportion p_g as a constant. We present model specifications for the two experiments in Panel B and Panel C of Table 2. In Experiment 1, we fix the weights but vary the number of dimensions from 1 to 2, 4 and 4, represented by Round 1.1, Round 1.2, Round 1.3 and Round 1.4. In Experiment 2, we fix the number of preference dimensions at 4 but vary the weight on race from -2 to -4, -6 and -8, represented by Round 2.1, Round 2.2, Round 2.3 and Round 2.4.

We present the results of the two experiments in Figures 1 and 2. In both figures, the horizontal axis is context size, and the vertical axis is the average share of interracial friends. First, let us look at the general trend. A total of eight curves in the figures support our theoretical prediction of size effect, as they all slope downward, indicating that the expected share of interracial friends decreases with total size. Moreover, they also show that the relationship between interracial friendship and total size is non-linear: there is a diminishing return on total size. As total size increases, the share of intercial friends approaches a limit discussed in the previous section.

Next, we turn to Propositions 1 and 2. The four curves in Figure 1 represent results for Experiment 1, which models varying numbers of dimensions in the utility function. The dash-dotted line indicates only one dimension of preference: racial homophily. Thus in the utility function of Equation (10), $\beta_2 = \beta_3 = \beta_4 = 0$. This curve is almost flat.[22] When we increase the number of dimensions from one to two, three, and four, the slope of the curve becomes increasingly steeper, which means that size effect becomes larger with the increase in the number of dimensions. Further, the four curves converge to the same limit, although by different rates, as total size increases.[23] Hence, our agent-based modeling results support Proposition 1.

[22] The reason that the curve does not reduce to a horizontal flat line is that the random component, although small in variance, still exists in the utility function, and the noise of this random component diminishes with increased size.

[23] Theoretically, Equation (9) can help us determine that the limiting shares of interracial friends in the four cases as N approaches infinity should be 0.002459.

Figure 1. Average share of interracial friends in Experiment 1 by numbers of preference dimensions

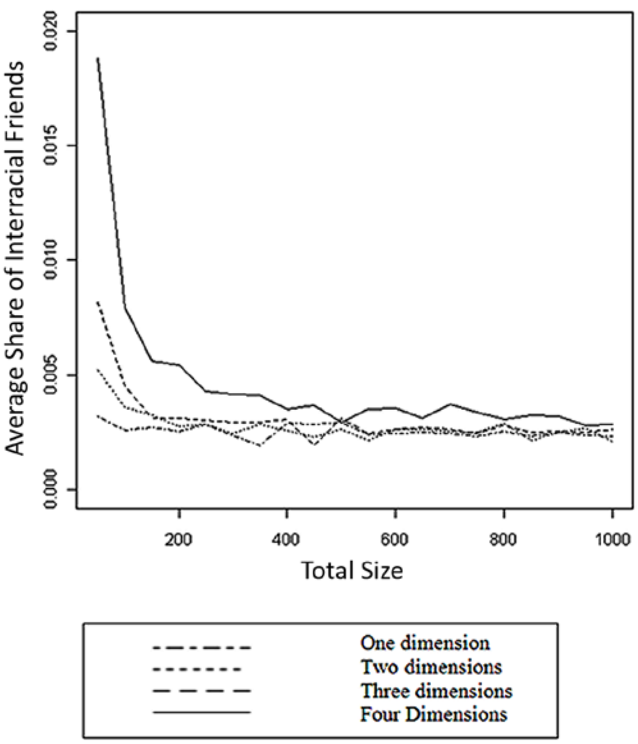
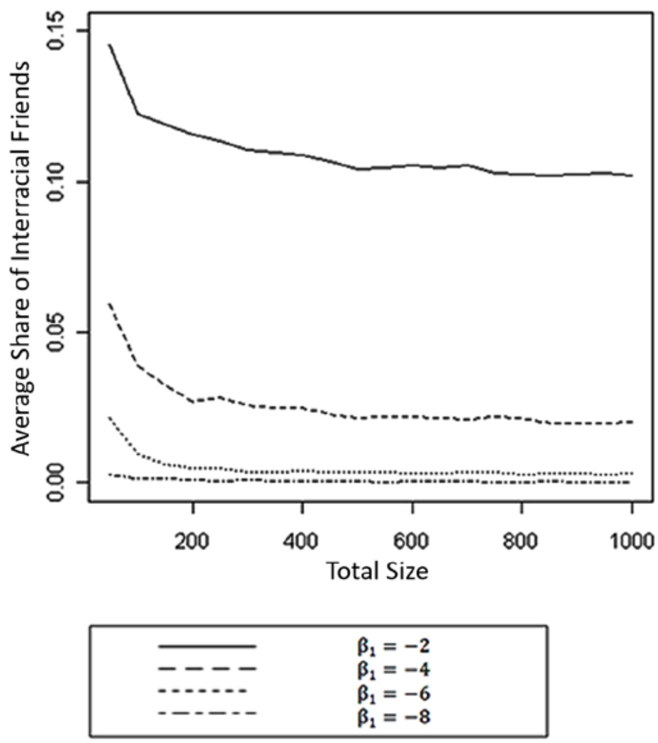


Figure 2. Average share of interracial friends in Experiment 2 by strength of racial preference



The four curves in Figure 2 represent results for Experiment 2. Four coefficients of weights (β_{race}) are considered:-2, -4,-6 and -8, with the first being the weakest and the last the strongest. The results show that the stronger the agent's preference for racial homophily, the lower the level of interracial relations, as he/she is more reluctant to befriend an other-race person as opposed to a same-race person. The comparison of the curves corroborates our Proposition 2. As is evident from the figure, the slope of the curve gets flatter, or the share of interracial friends is affected less by size, as we increase the weight for racial preference. We observe that the slope for $\beta_{race} = -2$ is the steepest, while for $\beta_{race} = -8$ the flattest. This pattern confirms Proposition 2, i.e., the effect of total size on racial segregation is negatively associated with the relative importance of preference for racial homophily.

5. Friendship Segregation in American Schools: An Empirical Test

The primary objective of this paper is to propose a theoretical argument that context size, *ceteris paribus*, exerts a positive structural effect on group segregation in social relationship. We have presented the theoretical model under some well-accepted assumptions and demonstrated the model with results from two experiments using agent-based modeling techniques. Ideally, we would also like to put our theory to empirical tests. However, it is not easy to find appropriate empirical data with which to test the theory. The main difficulty lies in the condition "*ceteris paribus*," i.e., everything else being equal. When we compare friendship patterns across social contexts of different sizes, how can we be sure that the social contexts are indeed comparable either in other contextual characteristics or in characteristics of persons who are located in different contexts?

To date, we believe that Add Health data are the best data source for an empirical test of the theory. With Add Health, we focus on youth's interracial friendship as the outcome of interest, with schools as social contexts. In this section, we present our effort to use the Add Health data to test the theory. We conduct the empirical analyses at two levels, first at the school level and then at the individual level.

5.1. Add Health Data

We analyze data from the Wave I in-school survey of the National Longitudinal Study of Adolescent Health (Add Health), which provides a school-based sample of adolescents in grades 7-12 in the United States during the 1994-95 school year. A total of 172 schools participated in the survey. The survey collected information on demographic backgrounds and in-school activities from about 90,000 students in grades 7 through 12 in a 45- to 60-minute class period between September 1994 and April 1995. The survey instrument asked each respondent to nominate up to five best male friends and five best female friends in order of closeness from a school roster. The roster-based nomination enables us to match the attributes of each respondent (agent) with the attributes of his/her friends. For simplicity, we focus on the agent's choice of best friend for the illustration of our model.

In testing our theory, we add the following three restrictions to the data. First, while adolescents make friends within the focal school (their own school) as well as in other social contexts, such as in their sister school[24] or their neighborhood,[25] we restrict context to be the focal school to ensure that we have the demographic information of both the agent and his/her alternatives.[26] Hence, conceptually, we model respondents' choice of the "best in-school friend." In operation, if the best friend of the respondent is not in the focal school, we substitute the highest-ranked friend within the focal school. Second, because the nomination of the best opposite-sex friend includes boyfriend or girlfriend, who is likely based on a different preference from that for a non-romantic friend, we limit our analysis to same-sex friends. Third, we focus on the interracial friendship formation of four mutually exclusive racial groups: non-Hispanic white, non-Hispanic black, non-Hispanic Asian and Hispanics, denoted as W, B, A and H respectively in our analysis.[27] Table 3 gives the summary result cross-classifying the race of the best friend by the race of the agent. The diagonal cells reveal that on average, whites are most likely to make a same-race best friend (88.90%), followed by blacks (81.31%), Hispanics (58.52%), and then Asians (54.75%).

Table 3. Racial composition of the agent's best friend, by the agent's race

Agent's Race	Race of agent's best friend (percentage)					
	White	Black	Asian	Hispanic	Total	N
White	88.90	2.20	2.49	6.41	100	38,578
Black	7.70	81.31	1.33	9.66	100	10,736
Asian	29.72	4.41	54.75	11.12	100	3,281
Hispanic	26.68	10.59	4.22	58.52	100	9,868
Total	62.97	16.93	5.34	14.77	100	62,463

Data Source: The National Longitudinal Study of Adolescent Health Wave I in-school sample.

[24] In Add Health, a sister school is a middle school or a high school that has a feeder relationship with the focal school. It either supplies the focal school with incoming students or enrolls most of its graduates.

[25] For example, Mouw and Entwisle (2006) found that about a third of the level of interracial friendship segregation in schools is attributable to residential segregation.

[26] The Add Health dataset collects the students' demographic information, including race and ethnicity only for respondents in the focal school. Since this information is key to our theoretical model, we need to restrict the choice set to the focal school. We do recognize the potential bias in neglecting friendship from the neighborhood or other sources, but to the extent that the association between school-based and neighborhood-based friends does not wipe out the effect of school size completely, this restriction is not consequential to the testing of our theoretical model.

[27] We realize that Hispanic origin is often treated as an ethnicity rather than a race. To keep the language simple for this section, we ignore the ethnic-racial distinction and call the four groups "racial groups." Consistent with the commonly assumed one-drop rule, we categorize a white and Asian respondent as black even if he/she also reports being partly black. Given that white-Asian biracials are equally likely to be classified as either white or Asian (Xie and Goyette 1997), we randomly assign a white-Asian respondents to be white or Asian with equal probability.

5.2. School-Level Analysis

The primary interest of our theoretical model is the agent's likelihood of choosing an interracial friend as a function of school size. This association can be analyzed at two levels. First, we can evaluate the validity of our theoretical argument at the *school level*, because the aggregate level of racial segregation in friendship, the outcome of our primary interest, is a school-level characteristic. In our first set of analyses, we measure the aggregate level of racial friendship segregation by the average share of interracial friends in the school. To account for friendship choice pattern for different racial groups, we further break down the measure of "interracial friends" into 12 interracial combinations for the agent and his/her chosen friend, which are denoted by 12 ordered two-letter pairs: *WB*, *WA*, *WH*, *BW*, *BA*, *BH, AW*, *AB*, *AH*, *HW*, *HB*, and *HA*, with the first letter representing the race of the agent, and the second letter representing the race of the agent's chosen friend. For example, *WB* denotes a dyad in which a white person chooses a black friend, and *BW* denotes a dyad in which a black person chooses a white friend. The average share of interracial friends is calculated specifically for each school in the sample. School size is the total number of students in the school roster (in thousands). For simplicity, we use ordinary linear regression (OLS) and regress the proportions of such interracial friendship ties on school size, proportions of racial groups, type of school (public or private), urbanity of school (urban, suburban, or rural). A summary of the school-level variables is given in Table 4.

The primary goal of this paper is to argue for a distinct effect of school size. Thus, we focus our discussion of empirical results on the coefficients of school size, while treating compositional measures and other school attributes as controls. In Table 5, we present the main results from our school-level analysis pertaining to our study, namely the coefficients of school size on the proportions of the 12 aforementioned types of interracial friendship in a coefficient matrix. The coefficients should be interpreted row-wide, with racial homophily (say a white choosing a white friend for the first row) as the reference. See Table A1 for the full regression results. The results shown in Table 5 overall corroborate our theory, as seven significant coefficients on school size are all negative, indicating that school size negatively affects the likelihood of making interracial friends. For example, the coefficient for a black agent choosing a white best friend is -0.0363, meaning that an increase of one thousand students in school size is associated with a decrease of 0.0363 in the average share of whites being chosen by black agents as best friends. The other five coefficients in Table 5 are not statistically different from zero.

Table 4. Summary statistics for school-level variables in Add Health data

School-Level Variable	Mean	Std. Dev.
Number of Schools in the Sample		172
School Size	1239.96	752.32
Racial Group Proportions		
Proportion of Whites	59%	0.30
Proportion of Blacks	16%	0.22
Proportion of Asians	5%	0.09
Proportion of Hispanics	15%	0.19
Urbanity	Frequency	Percentage
Urban Schools	56	32.56%
Suburban Schools	94	54.65%
Rural Schools	22	12.79%
School Type		
Public Schools	155	90.12%
Private Schools	17	9.88%

Data Source: The National Longitudinal Study of Adolescent Health Wave I in-school sample.

Table 5. Selected coefficients for school size predicting school-level share of interracial friends

			Race of agent's best friend		
		White	Black	Asian	Hispanic
Agent's race	White	---	-0.0059***	-0.0005*	-0.0075***
	Black	-0.0363***	---	0.0005	-0.0610***
	Asian	0.0058	-0.0341***	---	0.0066
	Hispanic	-0.0024	-0.0225***	0.0001	---

Data Source: The National Longitudinal Study of Adolescent Health Wave I in-school sample.
Notes:
1. * P<0.05; ** p<0.01; ***p<0.001.
2. Coefficients are selected from twelve models, one for each combination of matched pair. All of these models are OLS regression with the outcome variable being the average share of pairwise-measured interracial friendship dyads by school. School size is measured by the number of students on the school roster in thousands. Full models are presented in Table A1.

5.3. Individual-Level Analysis

One shortcoming of the school-level analysis is essentially an ecological analysis and as such could be biased from an aggregation of individual-level characteristics associated with school size. Thus, the results in Table 5 may be confounded by the differential composition of individual characteristics across school contexts. To check the robustness of our earlier results against their sensitivity to compositional differences, we conduct additional analysis examining the individual-level likelihood of choosing an interracial friend as a function of a mixture of individual and school-level attributes. Similarly as we did for the school-level analysis, we break down "interracial friend" ties into a total of 12 interracial combinations. Hence, the interracial friend outcomes for a white agent are *WB*, *WA* and *WH*, for a black agent are *BW*, *BA* and *BH*, for an Asian agent are *AW*, *AB* and *AH*, and for a Hispanic agent are *HW*, *HB* and *HA*. We estimate multinomial logit regressions for each racial group. We parameterize the regression coefficients in such a way that the coefficients represent the logged odds-ratio of the agent's likelihood of choosing an interracial friend in each racial group, with the baseline outcome category being the agent choosing a same-race friend. The independent variables include school size measured in thousands, the agent's gender, age, grade in school, indicators for the chooser's mother's education level, the group proportion for the agent's race in the focal school, and the indicator for the school's being in a urban area.

Tables A2 presents the full results for the four multinomial models. Table 6 gives the key results for the effects of school size on interracial friendship for all combinations. Of seven coefficients that are statistically significant from zero, five are negative, in the direction in favor of our theoretical model. For example, the logged odds ratio is -0.415 (or the odds ratio is 0.66) for a black agent choosing a white friend (a *BW* pair). This means that holding other conditions consistent, if school size increases by 1000 students, a black agent's odds of choosing a white friend relative to a black friend declines by 34%. We note that the two statistically significant coefficients not supporting our theory both pertain to the selection of Hispanics as best friends. The deviant cases for Hispanics may result from an ambiguity in American society as to whether Hispanics should be treated as an ethnicity overlapping with race or, as we do in this paper, effectively as a distinct racial group. Overall, the close resemblance in results between Tables 5 and 6 indicates that our earlier findings from school-level analysis are not sensitive to the compositional variance of individual attributes in the school. To sum up, analysis of Add Health data on both the school and individual consistently supports our theoretical argument that holding everything else constant, total school size negatively affects the likelihood of interracial friends, and hence positively affects the level of friendship racial segregation.

Table 6. Selected coefficients for school size predicting the agent's likelihood of choosing an interracial friend

		Race of the agent's best friend				
		White	Black	Asian	Hispanic	N
Agent's race	White	Baseline	-0.180*	-0.0907	0.276***	25638
	Black	-0.415***	Baseline	0.257	0.0908	6281
	Asian	-0.476***	-0.229	Baseline	0.295*	1857
	Hispanic	-0.170**	-0.463***	-0.205	Baseline	4946

Data Source: The National Longitudinal Study of Adolescent Health Wave I in-school sample.
Notes:

1. * P<0.05; ** p<0.01; ***p<0.001.
2. Coefficients are selected from four models, one for each racial/ethnic group. All of these models are multinomial logit regression with the categories of choice outcome being choosing a best friend from a racial group, and the baseline category being choosing a same-race friend. School size is measured by the number of students on the school roster in thousands. Full models are presented in Table A2.

6. Conclusion

A cornerstone of sociology is the recognition that an individual's agency is always constrained by his/her surrounding context. This paper builds on this long-standing sociological tradition that embeds an individual's "rational" choice in a social context. Of course, the importance of social context is now so well entrenched in sociology that it is no longer sensible to ask *whether* it matters, but *how* it matters. In the past several decades, a large body of sociological research has firmly established group composition as a structural determinant for the formation of social relations. However, this body of work has overlooked the similar structural role of total size.

The main objective of this paper was to argue for the structural influence of total size in a social context on social relationship formation. In the preceding sections, we showed, both with agent-based modeling and with empirical data analysis, that an increase in size, everything else being equal, leads to a higher level of racial segregation in friendship. However, we believe that the contributions of this paper are not limited to the demonstration of the mere existence of a size effect. Our core result has important wider implications.

First, our model contributes to the understanding of the interaction between individual agency and social structure. Several schools of social scientists, among which neoclassical economists are perhaps the most fervent, assume that the social world is made of rational actors with free agency. Our model recognizes the individual's "free" agency by introducing latent preference as an intrinsic disposition and invoking the rational actor assumption as a context-invariant behavioral rule in realizing preference. However, we demonstrated, with the example of the size effect on the choice of social relations, that the realization of predetermined preference under "rational" individual agency is subject to structural constraints.

Second, we draw attention to multidimensional preference in relationship formation. While existing theories have well alluded to the multidimensionality of individual preference, they have not identified its significance for the structural effects of social context. To our knowledge, our study is the first to integrate multidimensional preference and structural effects under a common framework. We did so by establishing that multidimensionality in preference is a necessary condition for a size effect, in that a large total size acts as a buffer against the compromises across preference dimensions.

Aside from its contribution to the structural effect of social context, our model leaves some interesting extensions open for future research. First, in this paper, we assume that alternatives in the context are equally accessible to the agent, without adjustment for differential exposure within the choice set. To our relief, an emerging line of work has offered ways to account for such exposure differences. For example, Zeng and Xie (2008) proposed a preference-opportunity-choice framework that separates individual-level opportunity structure out from individual-level preference. Hence, we recommend that future research extend the findings from our theoretical model to incorporate varying opportunity structures within the context. Second, in investigating the structural effect, we have assumed that individual preference is fixed. In reality, however, latent preference varies across context. Thus, it stands as a challenge for future work to identify, either using theoretical modeling or empirical data, the distinction of the structural effects from the variability in preference effects across contexts. Third, we sympathize with the Granovetter's (1985) contention that social relationship formation is a dynamic process in which agents are always "embedded" in their network relations. We believe that our model can be seen as a static simplification, but it is possible to extend this model so as to account for the dynamics of the network structure.

We close with a note on the broad implications of our theoretical argument. As we have mentioned earlier, the recognition of the size effect reshapes our understanding of the impact of social context on certain social outcomes, such as friendship racial segregation, marital choice, and the choice of residential neighborhood. In our study, we confined our operationalization of a social context to conventional arenas, such as schools. However, our model may have implications for the emergence of very large, complex, and usually intangible communities such as Facebook and Twitter in cyber space. As a simple illustration, Facebook increases the opportunities for a person to meet a potential friend of specific characteristics that are hard to find through actual contact. In doing this, cyber space enlarges the size of the individual's choice set, and thereby promotes in-group matches in terms of multiple social dimensions. Thus, will the emergence of cyber-space communities promote social segregation and hinder social integration? Yes, if size matters, as we think it does.

References

Allport, Gordon W. 1954. *The Nature of Prejudice*. Reading, MA: Addison Wesley.

Allport, Gordon W. 1966. "The Religious Context of Prejudice." *Journal for the Scientific Study of Religion* 5:447-57.

Blau, Peter M. 1974. "Presidential Address: Parameters of Social Structure." *American Sociological Review* 39:615-35.

Blau, Peter M. 1977. "A Macrosociological Theory of Social Structure." *American Journal of Sociology* 83:26-54.

Blau, Peter. M. 1994. *Structural Contexts of Opportunities.* Chicago: The University of Chicago Press.

Blau, Peter. M. and Joseph E. Schwartz. 1984. *Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations.* New York: Academic Press.

Blau, Peter M., Terry C. Blum, and Joseph E. Schwartz. 1982. "Heterogeneity and Intermarriage." *American Sociological Review* 47:45-62.

Blum, Terry C. 1985. "Structural Constraints on Interpersonal Relations: A Test of Blau's Macrosociological Theory." *The American Journal of Sociology* 91:511-21.

Bourdieu, Pierre. 1989. "Social Space and Symbolic Power." *Sociological Theory* 7:14-25.

Breiger, Ronald L. 1974. "The Duality of Persons and Groups." *Social Forces* 53:181-90.

Brewer, Marilynn B. and Joseph G. Weber. 1994. "Self-evaluation Effects of Interpersonal Versus Intergroup Social Comparison." *Journal of Personality and Social Psychology* 66:268-75.

Burt, Ronald S. 1990. *Kinds of relations in American discussion networks.* Pp. 411-51, Structures of Power and Constraint, edited by C. Calhoun, MW Meyer, WR Scott. New York: Cambridge.

Currarini, Sergio , Matthew. O. Jackson and Paolo Pin, 2010. "Identifying the Roles of Race-based Choice and Chance in High School Friendship Network Formation." *Proceedings of the National Academy of Science* 107: 4857.

Dovidio, John F., Samuel L. Gaertner , and Kerry Kawakami. 2003. "Intergroup Contact: The Past, Present, and the Future." *Group Processes & Intergroup Relations* 6(1): 5-21.

Feld, Scott L. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86:1015-35.

Fischer, Claud S. 1977. *Networks and Places: Social Relations in the Urban Setting.* New York: Free Press. Fischer, Claud S. 1982. *To Dwell among Friends.* Chicago: University of Chicago Press.Gibbs, Jack P. and Walter T. Martin. 1962. "Urbanization, Technology, and the Division of Labor:International Patterns." *American Sociological Review* 27:667-77.

Flashman, Jennifer. 2012. "Academic Achievement and Its Impact on Friend Dynamics." *Sociology of Education* 85:61-80.

Fossett, Mark and Dietrich David R. 2009. "Effects of city size, shape, and form, and neighborhood size and shape in agent-based models of residential segregation: are Schelling-style preference effects robust?" *Environment and Planning B: Planning and Design* 36(1) 149 – 169.

Giles-Corti, Billie and Robert J. Donovan. 2003. "Relative Influences of Individual, Social Environmental, and Physical Environmental Correlates of Walking." *American Journal of Public Health* 93(9): 1583-89.

Goffman, Erving. 1971. *Relations in Public: Micro-Studies of the Public Order*. New York: Harper & Row.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78:1360-80.

Granovetter, Mark S. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology 91(3):481-510.*

Grund, Thomas. 2010. "Emergence of Homophily: A Micro-Mechanism Approach Focusing on Triadic Closure." Unpublished manuscript.

Harding, David J. 2007. "Cultural Context, Sexual Behavior, and Romantic Relationships in Disadvantaged Neighborhoods." *American Sociological Review* 72(3): 341-364.

Hallinan, Maureen T. and Richard A. Williams. 1989. "Interracial Friendship Choices in Secondary Schools." *American Sociological Review* 54:67-78.

Hannerz, Ulf. 1969. *Soulside: Inquiries into Ghetto Culture and Community.* New York: Columbia University Press.

Hewstone, Miles, Mark Rubin, and Hazel Willis. 2002. "Intergroup Bias." *Annual Review of Psychology* 53: 575–604.

Joyner, Kara and Grace Kao. 2000. "School Racial Composition and Adolescent Racial Homophily." *Social Science Quarterly* 81(3):810+

Kalmijn, Matthijs. 1991. "Status Homogamy in the United States." *American Journal of Sociology* 97: 496-523.

Kalmijn, Matthijs. 1998. "Intermarriage and Homogamy: Causes, Patterns, Trends." *Annual Review of Sociology* 24:395-421.

Kalmijn, Matthijs and Frank van Tubergen. 2010. A Comparative Perspective on Intermarriage: Explaining Differences Among National-Origin Groups in the United States. *Demography*, 47(2), 459-79.

Kao, Grace and Kara Joyner. 2004. "Do Race and Ethnicity Matter among Friends? " *Sociological Quarterly* 45(3): 557-73.

Kossinets, Gueorgi and Duncan.J. Watts. 2009. "Origins of Homophily in an Evolving Social Network1." *American Journal of Sociology* 115(2):405-50.

Laumann, Edward O. 1973. *Bonds of Pluralism: The Form and Substance of Urban Social Networks.* New York: Wiley.

Laurie, Alexander J. and Narendra K. Jaggi. 2003. "Role of 'Vision' in Neighbourhood Racial Segregation: A Variant of the Schelling Segregation Model." *Urban Studies* 40 (13): 2687-2704.

Lewis, Susan K. and Valerie K. Oppenheimer. 2000. "Educational Assortative Mating across Marriage Markets: Non-Hispanic Whites in the United States." *Demography* 37(1): 29-40.

Lichter, Daniel T., Robert Anderson, and Mark Wayward. 1995. "Marriage Markets and Marital Choice." *Journal of Family* 16: 412–31.

Mare, Robert D. 1991. "Five Decades of Educational Assortative Mating." *American Sociological Review* 56(1): 15-32.

Marsden, Peter V. 1987. "Core Discussion Networks of Americans." *American Sociological Review* 52(1):122-131.

Mas-Colell, A.. Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory*. Oxford, UK: Oxford University Press.

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

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." Pp. 105–42 in *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press.

McFadden, Daniel. 2001. "Economic Choices." *The American Economic Review* 91:351-378.

McPherson, J. Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Merton, Robert K. 1972. "Insiders and Outsiders: A Chapter in the Sociology of Knowledge." *American Journal of Sociology* 78:9-47.

Moody, James. 2001. "Race, School Integration, and Friendship Segregation in America." *American Journal of Sociology* 107(3): 679-716.

Mouw, Ted and Barbara Entwisle. 2006. "Residential Segregation and Interracial Friendship in Schools." *American Journal of Sociology* 112:394-441.

Oppenheimer, Valerie K. 1988. "A Theory of Marriage Timing." *The American Journal of Sociology* 94(3): 563-91.

Qian, Zhenchao. 1998. "Changes in Assortative Mating: The Impact of Age and Education, 1970-1990." *Demography* 35(3): 279-92.

Quillian, Lincoln and Mary E. Campbell. 2003. "Beyond Black and White: The Present and Future of Multiracial Friendship Segregation." *American Sociological Review* 68(4): 540-66.

Rosenfeld, Michael J. 2005. "A Critique of Exchange Theory in Mate Selection1." *American Journal of Sociology* 110:1284.

Rytina, Steve and David L. Morgan. 1982. "The arithmetic of social relations: the interplay of category and network." *American Journal of Sociology* 88(1):88-113.

Schwartz, Christine R. 2010. "Earnings Inequality and the Changing Association between Spouses' Earnings." *American Journal of Sociology* 115(5): 1524–57.Shrum, Wesley, Neil H. Cheek, Jr., and Saundra MacD Hunter. 1988. "Friendship in School: Gender and Racial Homophily." *Sociology of Education* 61(4):227-239.

Simmel, Georg. 1955. *Conflict and the Web of Group-Affiliations.* New York: Free Press.

Smith, Eliot and Gun Semin 2007. "Situated Social Cognition." *Current Directions in Psychological Science* 16(3): 132-135.

Snijders, Tom A. B. 2001. "The Statistical Evaluation of Social Network Dynamics." *Sociological Methodology* 31:361-395.

Snijders, Tom A.B., Gerhard G. van de Buntb and Christian E.G. Steglichc. 2010. "Introduction to stochastic actor-based models for network dynamics." *Social Networks* 32(1):44-60.

Suitor, Jill and Shirley Keeton. 1997. "Once a friend, always a friend? Effects of homophily on women's support networks across a decade." *Social Networks* 19(1):51-62.

Tajfel, Henry and John C.Turner. 1979. An Integrative Theory of Intergroup Conflict. Pp. 33-47 in *The Social Psychology of Intergroup Relations*, edited by W. G. Austin and S. Worchel. Monterey, CA: BrooksCole.

Torche, Florencia. 2010. “Educational Assortative Mating and Economic Inequality: A Comparative Analysis of Three Latin American Countries.” *Demography* 47(2): 481–502.

Turner, John C. and Kenneth Reynolds. 2001. “The Social Identity Perspective in Intergroup Relations: Theories, Themes, and Controversies.” Pp. 133-52 in *Blackwell Handbook of Social Psychology: Intergroup Processes*, Vol. 4, edited by R. Brown and S. L. Gaertner. Oxford, UK: Blackwell.

Wasserman, Stanley and Katherine Faust. *Social Network Analysis: Methods and Applications.* Cambridge, UK: Cambridge University Press. Williams, Robin M., Jr. 1947. *The Reduction of Intergroup Tensions.* New York: Social Science Research Council.

Watts, Duncan J. 1999. "Networks, Dynamics, and the Small‐World Phenomenon." *American Journal of Sociology* 105(2):493-527.

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

Wimmer, Andreas and Kevin Lewis. 2010. "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook." *American Journal ofSociology* 116:583-642.

Wright Erik O. 1997. *Class counts: Comparative studies in class analysis.* New York: Cambridge University Press.

Xie, Yu. 2007. “Otis Dudley Duncan’s Legacy: The Demographic Approach to Quantitative Reasoning in Social Science.” *Research in Social Stratification and Mobility* 25: 141-56.

Xie, Yu and Kimberly Goyette. 1997. “The Racial Identification of Biracial Children with One Asian Parent: Evidence from the 1990 Census.” *Social Forces* 76: 547-70.

Xie, Yu and Daniel Powers. 2008. Statistical Methods for Categorical Data Analysis. Bingley, UK: Emerald Group Publishing Limited.

Yamaguchi, Kazuo. 1990. "Homophily and Social Distance in the Choice of Multiple Friends: An Analysis Based on Conditionally Symmetric Log-Bilinear Association Model." *Journal of the American Statistical Association* 85:356-366.

Zeng, Zhen. 2004. "The Contextual Determinants of Interracial Preferences in Adolescent Friendship." PhD dissertation, Department of Sociology, University of Michigan, Ann Arbor, MI.

Zeng, Zhen and Yu Xie. 2008. "A Preference-opportunity-choice Framework with Applications to Intergroup Friendship." *American Journal of Sociology* 114:615–648.

Appendix

Table A1. OLS regression predicting school-level average share of interracial friends using school-level characteristics, by the race of the agent and the race of the agent's chosen friend

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	WB	WA	WH	BW	BA	BH	AW	AB	AH	HW	HB	HA
Proportion of the agent's race	-0.0420***	-0.0310***	-0.0700***	-0.197***	-0.0179***	-0.0660***	-0.3800***	-0.0891***	-0.0141	-0.0575***	-0.0872***	-0.0254***
	(0.0013)	(0.0009)	(0.0018)	(0.0075)	(0.0012)	(0.0031)	(0.0221)	(0.0095)	(0.0165)	(0.0073)	(0.0055)	(0.0030)
Proportion of the chosen friend's race	0.1800***	0.6330***	0.5630***	0.190***	0.2580***	0.547***	0.7920***	0.3480***	0.428***	0.8100***	0.8670***	0.663***
	(0.0018)	(0.0025	(0.0032)	(0.0071)	(0.0052)	(0.0063)	(0.0135)	(0.0095)	(0.0173)	(0.0055)	(0.0058)	(0.0055)
School size	-0.0059***	-0.0005*	-0.0075***	-0.0363***	0.0005	-0.0610***	0.0058	-0.0341***	0.0066	-0.0024	-0.0225***	0.0001
	(0.0059)	(0.0002)	(0.0004)	(0.0018)	(0.0005)	(0.0014)	(0.0043)	(0.0024)	(0.004)	(0.0014)	(0.0013)	(0.0007)
School type (reference category: public school)												
Private school	0.0071***	0.0205***	-0.0090***	0.0375***	0.0517***	-0.0540***	0.0054	-0.0176***	-0.0088	0.0189***	0.0215***	0.0140***
	(0.0006)	(0.0005)	(0.0009)	(0.0057)	(0.0017)	(0.0042)	(0.0074)	(0.0042)	(0.0073)	(0.0057)	(0.0053)	(0.0031)
School urbanity (reference category: urban school)												
Suburban school	-0.0076***	-0.0067***	0.0095***	-0.0048*	-0.0034***	-0.0004	0.0260***	-0.0002	-0.0035	0.0087**	-0.0219***	-0.0107***
	(0.0004)	(0.0004)	(0.0006)	(0.0023)	(0.0007)	(0.0018)	(0.0062)	(0.0033)	(0.0059)	(0.0027)	(0.0026)	(0.0015)
Rural school	-0.0093***	-0.0052***	-0.0002	-0.0111**	-0.0025*	-0.0344***	0.0163	0.0601***	-0.0525***	0.0515***	0.0041	-0.0097**
	(0.0006)	(0.0005)	(0.0009)	(0.0038)	(0.0017)	(0.0028)	(0.0145)	(0.0078)	(0.014)	(0.0058)	(0.0053)	(0.0031)
Constant	0.0568***	0.0404***	0.0898***	0.1820***	0.0173***	0.1790***	0.0852***	0.0781***	0.0688***	0.0354***	0.0771***	0.0229***
	(0.0011)	(0.0007)	(0.0016)	(0.0069)	(0.0010)	(0.0028)	(0.0115)	(0.0043)	(0.0079)	(0.0046)	(0.0031)	(0.0016)
N	38467	38463	38488	10724	10700	10723	3259	3252	3266	9851	9851	9845

<u>Data Source:</u> The National Longitudinal Study of Adolescent Health Wave I in-school sample.

<u>Notes:</u> (1) Standard deviation in parentheses. * p<0.05, ** p<0.01, *** p<0.001.
(2) The twelve models (1)-(12) represents twelve combinations of the agent's race and the race of the agent's chosen friend, indicated by the two-letter pair at the first row of each column. School size is measured in thousands. The number of observations (*N*) for each model represents the number of observations from the agent's race, which is used to calculate the average share of interracial friends for that specific race.

Table A2. Multinomial logit models predicting the likelihood of choosing an interracial friend using individual-level and school-level characteristics, by the race of the agent and the race of the agent's chosen friend

	(1)			(2)			(3)			(4)		
Agent's race	White			Black			Asian			Hispanic		
Race of the agent's best friend	Black	Asian	Hispanic	White	Asian	Hispanic	White	Black	Hispanic	White	Black	Asian
Gender	-0.404***	-0.129	-0.231***	-0.429***	-0.876***	-0.349***	-0.152	-0.550*	0.0435	-0.151	-0.00531	-0.216
	(0.0888)	(0.0812)	(0.0538)	(0.106)	(0.221)	(0.0901)	(0.122)	(0.247)	(0.163)	(0.0794)	(0.108)	(0.148)
Age	0.0603	-0.173*	0.0897	-0.102	-0.585**	-0.0106	-0.203*	-0.324	-0.191	-0.128*	0.0339	-0.417***
	(0.075)	(0.0753)	(0.0464)	(0.0819)	(0.18)	(0.0645)	(0.0942)	(0.194)	(0.119)	(0.0544)	(0.0697)	(0.101)
Grade	-0.101	0.224**	-0.174***	0.0512	0.425*	-0.167*	0.230*	0.369	0.083	0.154**	-0.0179	0.413***
	(0.0784)	(0.0779)	(0.0488)	(0.086)	(0.187)	(0.069)	(0.101)	(0.206)	(0.13)	(0.0588)	(0.0756)	(0.108)
Mother's Education (Reference Category: never went to school)												
Eighth grade or less	-1.206	11.05	0.842	-3.078*	11.36	-1.128	1.376	14.66	-0.0929	-0.229	-0.957	0.64
	(0.9)	(489.6)	(1.082)	(1.445)	(1663.8)	(1.283)	(0.901)	(1766.3)	(0.778)	(0.443)	(0.544)	(1.048)
Above eighth grade, but did not graduate from high school	-1.251	11.14	0.556	-3.703**	10.6	-1.916	1.314	15.04	-0.305	0.601	-0.111	0.544
	(0.824)	(489.6)	(1.063)	(1.371)	(1663.8)	(1.251)	(0.875)	(1766.3)	(0.767)	(0.434)	(0.521)	(1.049)
High school graduate	-1.454	11.48	0.408	-3.395*	10.86	-1.64	1.662	15.53	-0.561	1.052*	0.473	1.157
	(0.814)	(489.6)	(1.06)	(1.358)	(1663.8)	(1.243)	(0.85)	(1766.3)	(0.728)	(0.429)	(0.509)	(1.034)
Completed a GED	-1.642	11.06	0.41	-3.377*	9.729	-1.39	1.726	14.93	-0.172	0.642	0.554	0.586
	(0.844)	(489.6)	(1.067)	(1.378)	(1663.8)	(1.255)	(0.98)	(1766.3)	(0.971)	(0.459)	(0.542)	(1.109)
Went to a business, trade, or vocational school after high school	-1.476	11.66	0.269	-3.350*	11.36	-1.877	2.113*	15.33	-0.201	1.118*	0.542	1.404
	(0.829)	(489.6)	(1.065)	(1.369)	(1663.8)	(1.253)	(0.905)	(1766.3)	(0.817)	(0.456)	(0.551)	(1.07)
College, but did not graduate	-1.605	11.67	0.376	-3.520**	10.99	-1.784	2.042*	14.62	-0.677	1.150**	0.671	1.698
	(0.823)	(489.6)	(1.062)	(1.363)	(1663.8)	(1.246)	(0.865)	(1766.3)	(0.765)	(0.44)	(0.524)	(1.042)
Graduated from college/university	-1.732*	11.99	0.0672	-3.297*	11.13	-1.754	1.528	15.06	-0.637	1.397**	0.828	1.73
	(0.817)	(489.6)	(1.061)	(1.359)	(1663.8)	(1.244)	(0.846)	(1766.3)	(0.719)	(0.436)	(0.519)	(1.039)

Table A2. Continued

	(1)			(2)			(3)			(4)		
Agent's race	White			Black			Asian			Hispanic		
Race of the agent's best friend	Black	Asian	Hispanic	White	Asian	Hispanic	White	Black	Hispanic	White	Black	Asian
Professional training beyond a four-year college	-1.6	12.12	-0.0769	-2.861*	11.74	-2.112	2.202**	15.42	-0.786	1.651***	1.109*	2.111*
	(0.822)	(489.6)	(1.063)	(1.363)	(1663.8)	(1.254)	(0.852)	(1766.3)	(0.748)	(0.464)	(0.554)	(1.064)
Proportion of the agent's race	-3.308***	-3.438***	-2.880***	-5.967***	-3.298***	-1.160***	-9.302***	-7.664***	-1.463*	-6.023***	-5.224***	-2.062***
	(0.202)	(0.186)	(0.132)	(0.323)	(0.505)	(0.168)	(0.566)	(1.211)	(0.632)	(0.27)	(0.35)	(0.453)
School size	-0.180*	-0.0907	0.276***	-0.415***	0.257	0.0908	-0.476***	-0.229	0.295*	-0.170**	-0.463***	-0.205
	(0.0766)	(0.0673)	(0.044)	(0.0857)	(0.162)	(0.0742)	(0.104)	(0.2)	(0.142)	(0.0629)	(0.0877)	(0.113)
School urbanity (reference category: urban school)												
Suburban school	-0.378***	0.18	-0.0844	0.0954	-0.404	-0.138	0.702***	-0.411	0.0135	-0.176	-0.646***	0.473*
	(0.1)	(0.0958)	(0.0643)	(0.112)	(0.229)	(0.0941)	(0.146)	(0.276)	(0.216)	(0.0937)	(0.124)	(0.189)
Rural school	-0.375*	0.0281	-0.469***	0.217	-0.519	-0.343*	0.780*	0.00507	-1.583	0.662**	0.545	-0.0212
	(0.163)	(0.161)	(0.113)	(0.191)	(0.483)	(0.164)	(0.308)	(0.55)	(1.042)	(0.254)	(0.285)	(0.628)
Constant	0.843	-12.4	-0.717	4.480**	-8.934	2.026	0.507	-14.22	0.879	1.175	0.298	-0.396
	(0.97)	(489.6)	(1.115)	(1.471)	(1663.8)	(1.329)	(1.078)	(1766.3)	(1.124)	(0.605)	(0.759)	(1.288)
N	25638	25638	25638	6281	6281	6281	1857	1857	1857	4946	4946	4946

Data Source: The National Longitudinal Study of Adolescent Health Wave I in-school sample.

Notes: (1) Standard deviation in parentheses. * p<0.05, ** p<0.01, *** p<0.001.
(2) For each of the four multinomial logit models (1)-(4), the baseline category is the agent choosing a friend from his/her own race. The coefficients that this table reports are the log odds ratios on the independent variables for choosing each of the other three racial groups relative to choosing a friend from his/her own race. School size is measured in thousands.