# A PREFERENCE-OPPORTUNITY-CHOICE FRAMEWORK WITH APPLICATIONS TO INTERGROUP FRIENDSHIP*

Zhen Zeng

University of Wisconsin-Madison

Yu Xie

University of Michigan

June 27, 2007

*Running Head*: Choice Framework

Word Count: 14,253

---

* Direct all correspondence to Zhen Zeng (e-mail: zzeng@ssc.wisc.edu), Center for Demography and Ecology, University of Wisconsin-Madison, 4406 Social Science Building, 1180 Observatory Drive, Madison, WI 53706, or Yu Xie (e-mail: yuxie@umich.edu), Population Studies Center, Institute for Social Research, 426 Thompson Street, University of Michigan, Ann Arbor, MI 48106. An earlier version of this paper was presented at the 2002 meeting of the American Sociological Association Section on Methodology, Princeton, NJ (April). This research was supported by a research grant from NICHD to Yu Xie and a traineeship from the Hewlett Foundation to Zhen Zeng.

# A PREFERENCE-OPPORTUNITY-CHOICE FRAMEWORK WITH APPLICATIONS TO INTERGROUP FRIENDSHIP


## Abstract

A longstanding objective of friendship research is to identify the effects of personal preference and structural opportunity on intergroup friendship choice. Although past studies have used various methods to separate preference from opportunity, researchers have not yet systematically compared the properties and implications of these methods. We put forward a general framework for conceptualizing the relationships of preference, opportunity, and choice. To implement this framework, we propose the conditional logit model with opportunity for estimating preference parameters free from the influence of opportunity structure. We also compare our approach to several alternatives—the conventional conditional logit model, the unconditional logit model, the loglinear model, and the log rate model for analyzing friendship choice. As an empirical example, we test hypotheses of homophily and status asymmetry in friendship choice using data from the National Longitudinal Study of Adolescent Health. The example also demonstrates the approach of conducting a sensitivity analysis to examine how parameter estimates vary by opportunity structure.

# A PREFERENCE-OPPORTUNITY-CHOICE FRAMEWORK WITH APPLICATIONS TO INTERGROUP FRIENDSHIP

The tendency for friends to be similar to each other has long been noted as a universal phenomenon. The adage, "Birds of a feather flock together," is believed to date back to the Roman historian, Livy. As contemporary sociologists see it, the predominance of homogeneous associations in friendship networks is due to the fact that the most significant rules operating in friendship choice are those of homophily and propinquity. Homophily is a preference principle that refers to the tendency for people to seek out and bond with others who are like themselves. Propinquity in this context is a structural principle based on the observation that social activities tend to bring people of similar status and attributes into contact with one another (Feld 1982). Consequently, people have a greater opportunity to make friends with similar others. Many researchers have observed that the pattern of homogeneous associations in interpersonal relationships such as marriage and friendship is a result not only of personal preference, but also of social structure (McPherson and Smith-Lovin 1987; McPherson et al. 2001; Quillian and Campbell 2003).[1]

The separation of the effects of preference and opportunity on friendship choice has been a longstanding objective in friendship research. There are at least two reasons for this interest. First, sociologists study intergroup friendship choice as a window into intergroup relations; for this purpose homophily is a purer indicator of social distance than observed homogeneous

---

[1] McPherson and Smith-Lovin (1987) used "induced homophily" and "choice homophily" to refer to the level of homogeneous association due to social structure and in-group bias respectively. Because the literal meaning of homo·phily is "*love* of the same kind," we use this word to strictly refer to in-group preference as a psychological disposition. That is, our use of "homophily" is equivalent to McPherson and Smith-Lovin's "choice homophily."

association due to both homophily and propinquity. Second, the separation of preference and opportunity allows researchers to compare patterns of preference across social contexts.

Past studies have attempted to separate preference and opportunity in a number of ways. Recognizing that the likelihood of having a friend in an outgroup is directly influenced by the relative size of the outgroup, some studies have focused on dyads (i.e., pairs of people) as the units of analysis and modeled the conditional probability of dyads being friends (e.g., Hallinan and Teixeira 1987; Moody 2001; Quillian and Campbell 2003; Mouw and Entwisle 2006). The dyadic approach effectively accounts for one particular aspect of opportunity—the group size factor—in intergroup friendship choice. Another approach is to statistically control for opportunity by adding variables that capture opportunity. For example, in predicting dyadic friendship, Mouw and Entwisle (2006) added residential distance between a pair of potential friends to control for physical proximity. Furthermore, at an aggregate level, researchers have used loglinear analysis to model contingency tables, where friendship ties are classified by respondents' and their friends' characteristics (Yamaguchi 1990). A well-known advantage of loglinear analysis is that its estimates of the association between a pair of categorical attributes (e.g., two age groups) are invariant to the marginal distributions of the attributes.

Although these disparate methods have been found useful in practice, researchers have not yet systematically compared their properties and implications. The purpose of this paper is to propose a unified conceptual framework for analyzing preference, opportunity, and choice, suitable for analysis at both the individual and aggregate levels. Within our framework, choice is conceptualized as resulting from the interaction of preference and opportunity. To disentangle the two, we incorporate opportunity into discrete choice models, symmetric to preference in functional form. We will discuss conditions under which preference and opportunity can be separated and conditions under which a clean separation cannot be achieved, but sensitivity analysis may nonetheless be conducted to examine how preference estimates vary by assumptions about the opportunity structure.

Our framework generalizes beyond friendship choice to a class of discrete choice situations where choice is constrained by opportunity structure. In this paper, we focus on the application to interracial friendship choice for two reasons. First, there has been much interest in friendship research on the separation of preference and opportunity. In the school desegregation literature, for example, researchers want to know the effect of diversity on racial homophily among adolescents (Mouw and Entwisle 2006). Because racial composition directly affects the opportunity of making interracial friends, we must account for this effect when interpreting the observed association between diversity and interracial friendship choice. Second, interracial friendship choice has been analyzed at both the individual level, with the choice set consisting of individuals, and at the group level, with the choice set consisting of racial groups. This application thus offers us an opportunity to demonstrate that the two levels of analysis can be unified through choice set aggregation in our framework.

We realize that friendship formation is a complicated social process, one that perhaps should not be boiled down to a discrete choice exercise. A more realistic model of friendship formation might take into account reciprocity of relationships, influence of common associates, and the time dimension, all of which are beyond the analytical power of standard discrete choice models. Let us state at the outset that our primary goal is not to offer a realistic behavioral model of friendship choice, but to propose a way to decompose friendship choice into preference and opportunity, which in turn can be incorporated into more sophisticated models of friendship formation in future research.

The rest of this paper is organized as follows. Section I sets up the preference-opportunity-choice (POC) framework and introduces the conditional logit model with opportunity (CLO) for estimating parameters of preference separate from opportunity. Section II proposes extensions of the CLO for analyzing ordered and unordered selection of multiple friends, followed by a discussion on choice set aggregation and a comparison of various related models. Section III demonstrates these methods through an empirical example with data from the National

Longitudinal Study of Adolescent Health. Finally, section IV draws conclusions and makes recommendations for the usage of the POC framework.

## I. The Preference-Opportunity-Choice Framework

*Unconstrained Choice versus Constrained Choice*

We begin with a distinction between two types of choice situations: unconstrained choice and constrained choice. In *unconstrained choice*, choice is based *purely* on preferences for alternatives under consideration. A prime example of unconstrained choice is a consumer survey of product preference, where respondents are presented with a hypothetical choice situation and asked to make one or more selections from a list of products. For example, they may be given a choice of Coke and Pepsi and asked which soft drink they prefer. We call it unconstrained choice because in a hypothetical choice situation like this, it can be assumed that choosers can pick any item they please from the choice set; the exercise of their preferences is free from external influences such as which product is more available on the local market.

*Constrained choice* is the situation where a choice decision is influenced not only by intrinsic preference but also by external factors such as availability, abundance, and affordability of the items in the choice set. In this paper, we refer to the influence of all these external factors on choice as *opportunity*. Real-world choices are always made under constraints. A classic example of constrained choice is a transportation study where people are surveyed on their means of transportation to work, that is, whether they go to work by car, bus, subway, bike, or on foot. In exercising their preferences for means of transportation, people are constrained by the “supply” factors. For example, some do not live on a bus or subway line, some cannot afford a vehicle, and still others live so far away from work that biking or walking to work is not feasible.[2] In a real-world choice situation like this, alternatives in the choice set are usually not equally accessible and, furthermore, accessibility varies across decision makers due to their own unique circumstances. As a result, people do not always end up choosing what they like best. Note that in both constrained and unconstrained choice, respondents are restricted by the survey instrument to the alternatives presented to them; the distinction we emphasize is whether the alternatives in the choice set can be regarded as equally accessible *a priori*.

The distinction between constrained choice and unconstrained choice has implications for the inference of preference from observed choice. We regard *preference* as the relative importance (called "utility" in economics) an individual attaches to the characteristics of choice alternatives, and we regard choice as the exercise of preference in a given context. Preference, as an underlying psychological attribute, is not directly observable and must be deduced from choice. In an unconstrained choice situation, choice directly expresses the underlying preference. That is to say, a higher choice probability for alternative A than for alternative B *always* indicates that A is preferred to B. In a constrained choice situation, choice does not correspond directly to preference. For example, suppose that the sales of Brand A milk exceed those of Brand B milk. Can we deduce from this *prima facie* evidence that consumers prefer brand A to Brand B? While this is a plausible interpretation, it may also be the case that consumers are indifferent to the two brands and that the sales difference is simply due to the better distribution and the resultant wider availability of Brand A.

The unconstrained and constrained choice situations correspond respectively to the stated preference and revealed preference methods in econometric analysis of choice. Stated preference

---

[2] In order to treat accessibility to the various means of transportation as an exogenous "supply" factor, we need to assume that access to public transportation does not influence people's decisions about where to live.

is a survey technique where respondents are asked to make choices between alternative services or products. By varying the attributes (e.g., packaging and taste) of the alternatives, often with the use of a factorial design, researchers explore the importance people attach to the various product attributes (Louviere et al. 2000). Revealed preference analysis deals with choices and decisions that have already been made in the real world. To study revealed preference, researchers must pay close attention to the context of choice—in addition to the attributes of the alternatives—in order to correctly deduce preference from choice. The same choice problem can be researched with either method. If we are able to disentangle opportunity and preference in revealed preference analysis, we should expect a high level of correspondence between stated and revealed preference. The current paper deals with how to deduce revealed preference from real world choices.

*The POC Framework*

By *opportunity* we broadly refer to the influence of all factors other than preference on choice. Depending on the context, opportunity may encompass different factors. As an example, let us consider what would be relevant external factors in intergroup friendship choice. Sociologists have long recognized that intergroup friendship choice depends not only on people's preferences but also on opportunities for intergroup interaction in their social environment. It is difficult to lay precise boundaries around a person's social environment; one common operationalization is to delimit it to either a physical space or an institution within which a social activity takes place, such as a metropolitan area or a school. The population composition of the social environment determines with whom individuals interact in that environment. Furthermore, interpersonal interactions in a social environment are structured. For example, schools are organized by grades and classes, and sometimes also by academic tracks. Students belonging to the same grades, classes, and tracks have more opportunity to interact with each other (Kubitschek and Hallinan 1998), and as a result, are more likely to become friends *a priori*. In general, opportunities for intergroup friendship choice depend on both the population composition and the organizational structure of a social environment.

The following simple example illustrates the effect of population composition on intergroup friendship choice. Students in a middle school were asked to nominate their best male friends and best female friends from the school roster. Table 1a presents female students' nominations of same-sex friends cross-classified by respondents' race and friends' race. As the table shows, for example, 140 out of the 181 white female students nominated best female friends, with 73 nominations going to whites, 24 to Hispanics, 34 to blacks, and 9 to Asians. We will analyze these data with regression models later. For now, let us make two qualitative observations: (1) within each row, the diagonal cells are larger than the off-diagonals, indicating an in-group bias; (2) larger groups (in this case, blacks and whites) receive more friendship nominations than smaller groups. While the former indicates the effect of preference on intergroup friendship choice, the latter reveals the influence of group size.

[Table 1 About Here]

To better understand the relationships among preference, opportunity, and choice, let us further consider the above friendship nomination data in two hypothetical situations. The first situation we examine is the state of indifference. Specifically, let us assume that students in this school are race-blind (and also blind to all other characteristics) in picking their best friends. In this situation, we would expect that, for any respondent, the probability that the best friend is of a particular race should be proportional to the size of that racial group. Correspondingly, the racial composition of all nominated friends should approximate the racial composition of the school.[3]

---

[3] In this example we make a simplifying assumption that this school is a homogeneous environment, where friendship is equally likely between any pair of schoolmates *a priori*. Consequently, opportunity for intergroup friendship depends on the population composition only. We will discuss this assumption later in the paper.

This scenario of indifference leads to the first property of our framework generalizeable to all discrete choice situations:

*POC 1 (State of Indifference) — If the chooser is indifferent to all alternatives in the choice set, choice probability is proportional to opportunity*.

The second hypothetical situation we consider is that of equal opportunity. Let us now suppose that the racial groups in this school are equally numerous, but that students are no longer race-blind in choosing friends. In this situation, we would expect that the likelihood of interracial friendship choice directly reflects the chooser's preferences for various races. Aggregated across all students, the number of nominated friends belonging to a particular race should be proportional to the average preference for that race in this school. The scenario of equal opportunity leads to the second property of our framework:

*POC 2 (Equal Opportunity) —If all alternatives in the choice set have an equal opportunity to be chosen, choice probability is proportional to preference.*

In actuality, we can assume neither the state of indifference nor that of equal opportunity except in artificially created situations. For general choice situations, we propose the following multiplicative assumption:

*POC 3 (Multiplicative Assumption) —Choice probability is proportional to the product of preference and opportunity.*

We now present this assumption in a more formal way. Let $i$ denote the chooser and $J$ denote the set of alternatives.[4] Let $p_{ij}$ be the probability that $i$ chooses alternative $j$ out of $J$, with $\sum_{j \in J} p_{ij} = 1$.

Let $o_{ij}$ and $a_{ij}$ be the opportunity and preference respectively for $i$ to choose $j$ out of choice set $J$. Both $o_{ij}$ and $a_{ij}$ are non-negative real values. For a given chooser $i$, $(p_{i1}, \ldots, p_{in})$, where $n$ is the size

---

[4] The choice set may vary by the chooser. In order to simplify notation, we assume the same choice set for all choosers and omit the subscript $i$ for $J$.

of $J$, is as a vector of choice probabilities, $(o_{i1},\ldots, o_{in})$ is a vector characterizing the opportunity structure $i$ faces, and $(a_{i1},\ldots, a_{in})$ is a vector characterizing $i$'s preferences for the alternatives in $J$. From now on, we use $\mathbf{p}_i$, $\mathbf{o}_i$, and $\mathbf{a}_i$ to denote the vectors of choice probability, opportunity, and preference respectively. The multiplicative assumption specifies the following relationship of $p_{ij}$, $o_{ij}$, and $a_{ij}$:

$$p_{ij} \propto o_{ij}a_{ij} . \tag{1}$$

After normalization we have:

$$p_{ij} = \frac{o_{ij}a_{ij}}{\sum_{k\in J} o_{ik}a_{ik}} . \tag{2}$$

It is easy to see that expressions 1 and 2 satisfy the aforementioned properties POC 1 and POC 2. When the chooser is indifferent to all alternatives in the choice set, i.e., $a_{ij} = a_{ik}, j \neq k$, the choice probabilities $\mathbf{p}_i$ are determined up to a scaling factor by the opportunity vector $\mathbf{o}_i$, that is, $p_{ij} \propto o_{ij}$ . When opportunity is equal, i.e., $o_{ij} = o_{ik}, j \neq k$, $\mathbf{p}_i$ is determined up to a scaling factor by the preference vector $\mathbf{a}_i$, that is, $p_{ij} \propto a_{ij}$ . Expression 2 is reminiscent of Luce's (1959) choice theorem $P_J(j) = \frac{v(j)}{\sum_{k\in J} v(k)}$,[5] where function $v(j)$ represents a response strength associated with response $j$, and choice probability is proportional to response strength. In standard discrete choice models, $v$ is interpreted as a utility function. In our framework, however, $v$ is decomposable to opportunity and preference, with the latter corresponding to utility in semantics.

Although opportunity and preference are mathematically symmetric in expression 1, the two quantities assume distinct roles in our choice framework. In particular, preference is a trait of the decision maker, whereas opportunity characterizes the circumstances under which choice

---

[5] In Luce's notation, $P_J(j)$ denotes the probably that $j$ is chosen out of choice set $J$.

occurs. By *circumstances*, we mean not only the social environment in which the decision maker is situated, but also the particular position he/she occupies in that environment. Let us consider the example of friendship choice among schoolmates again. Imagine that two students, *i* and *i'*, who attend two different schools, trade places with each other. In this exchange, they bring their preferences to the new environments, but leave behind their opportunity structures. If *i* and *i'* exchange roles perfectly—in terms of class scheduling and extracurricular activities—they will inherit each other's opportunity structure. This thought experiment illustrates how preference is an intrinsic characteristic of the decision maker, whereas opportunity is an extrinsic characteristic of the decision maker's social environment and position.

*The Conditional Logit Model with Opportunity*

The vector $\mathbf{a}_i$ represents chooser *i*'s preference for each alternative in the choice set. For both substantive and statistical reasons, we do not estimate $\mathbf{a}_i$ for each chooser-alternative combination but instead estimate parameters that characterize $\mathbf{a}_i$ through *preference function*. Substantively, we are interested not in particular individuals' preferences of concrete choice alternatives but in the pattern of association between choosers' characteristics and the characteristics of alternatives. For statistical considerations, we need to constrain the dimension of the parameter space to be smaller than that of the data space for identification. Since only one binary outcome is observed for each combination of chooser *i* and alternative *j*—either *i* chooses *j* as a friend, or *i* does not—we could not estimate $\mathbf{a}_i$ even if we wanted to. Therefore, we express the scalar $a_{ij}$ as a function of the characteristics of the chooser and of the alternatives. Let $\mathbf{z}_{ij}$ be a vector of characteristics pertaining to chooser *i* and alternative *j* and $\boldsymbol{\beta}$ be a vector of parameters. We specify the following preference function:

$$a_{ij} = \exp(\mathbf{z}_{ij}^{'}\boldsymbol{\beta})\,. \tag{3}$$

Substituting for $a_{ij}$ in (2) leads to the following expression for choice probability $p_{ij}$ in POC:

$$p_{ij} = \frac{o_{ij} \exp(\mathbf{z}_{ij}^{'}\boldsymbol{\beta})}{\sum_{k \in J} o_{ik} \exp(\mathbf{z}_{ik}^{'}\boldsymbol{\beta})}. \tag{4}$$

The exponential function for preference ensures that choice probabilities are non-negative. We call (4) the conditional logit model with opportunity (CLO). CLO is a weighted form of the standard conditional logit model, where utility $\exp(\mathbf{z}_{ij}^{'}\boldsymbol{\beta})$ is weighted by opportunity $o_{ij}$. In the situation of equal opportunity, (4) reduces to the standard conditional logit model:

$$p_{ij} = \frac{\exp(\mathbf{z}_{ij}^{'}\boldsymbol{\beta})}{\sum_{k \in J} \exp(\mathbf{z}_{ik}^{'}\boldsymbol{\beta})}. \tag{5}$$

CLO can be estimated using computer programs written for conditional logit model with $\ln(o_{ij})$ on the right-hand side as an offset variable, whose coefficient is not estimated but fixed at 1.

*Conceptualization of Opportunity in Economic and Sociological Analyses*

While consistent with the sociological literature on interpersonal associations, our conceptualization of opportunity as unequal accessibility of choice alternatives diverges from the notion of opportunity as constraints on the *composition* of choice sets prevalent in economic analysis of discrete choice. There, a distinction is made between a universal set of alternatives and the particular choice set for an individual. The latter is a subset of the former, consisting of "alternatives that are both feasible to the decision maker and known during the decision process" (Ben-Akiva and Lerman 1985, Pp. 33). Opportunity is then defined as the discrepancy between the choice set tailored to the individual decision maker and the universal set, and can be represented by indicators capturing the inclusion or exclusion of particular alternatives. Much research has been conducted to understand the effects of environmental and personal constraints on the composition of choice sets (e.g., Siddarth et al. 1995).

A special case of discrete choice models is the Two-Sided Logit (TSL) model (Logan 1996a; 1998; Logan et al. 2001), developed for matching problems such as college choice and job choice. In TSL, matched choices involve decisions of two parties, e.g., students and colleges, or workers and employers. Here, opportunity refers specifically to the fact that the choice set available to a decision maker in TSL is constrained by choices made by members of the other population. For example, a college's choice set consists of students who decided to apply, and a student's choice set consists of colleges that decided to extend an admission offer to him/her. The choice set faced by a college (student) is thus a very small subset of the universal set of students (colleges). As in one-sided discrete choice models, opportunity in TSL is a binary concept: individuals either have or do not have an opportunity to choose an alternative.

Distinct from this binary concept, we view opportunity as continuous values that quantify the accessibility of alternatives in a given choice set. Our notion of opportunity is akin to its usage in sociological research of intergroup associations (e.g., Moody 2001; Mouw and Entwisle 2006). Since Blau's (1977a; 1977b) seminal work on social structure and intergroup associations, sociologists have paid much attention to the problem of disentangling preference and opportunity for intergroup associations. Blau pointed out that social structure exerts a predictable influence on intergroup associations independent of sociopsychological dispositions. In particular, "[g]roup size governs the probability of intergroup relations…Inequality impedes and heterogeneity promotes intergroup relations." Guided by these ideas, studies of intergroup associations view opportunity as the pattern of associations predicted under the assumption of random mixing and preference as deviation from that pattern (Verbrugge 1977; Mayhew et al. 1995; McPherson et al. 2001; Mouw and Entwisle 2006). Consistent with this sociological tradition, we formalize opportunity and preference as choice probabilities of the indifferent chooser and choice probabilities in the state of equal opportunity respectively.

*Operationalization of Opportunity*

The form of (4) suggests that in order to estimate preference parameters based on observable choices, we must know the opportunity structure $\mathbf{o}_i$ *a priori*. The problem is that opportunity structure—defined as the choice probabilities of the indifferent chooser—is not directly observable. Therefore, it is often necessary to derive it from knowledge and assumptions about the choice context. Hence, the successful separation of opportunity and preference depends on how well we know the choice context and whether our assumptions are plausible. Next, we discuss the operationalization of opportunity, again using friendship choice as an example.

From prior literature, we know that opportunity for friendship choice in a social environment is affected by its demographic composition and organizational structure. The simplest opportunity structure is that of equal opportunity: all alternatives in a given choice set have the same choice probability *a priori*. Equal opportunity structure for friendship choice among schoolmates may arise from a homogeneous school environment, where the amount of interaction *induced by the structure of school activities* is equal for each pair of students (although the actual amount of interaction may differ as a result of students *choosing* to spend time with their friends). Equal opportunity structure is used implicitly in most studies of interpersonal relations, not because researchers believe that the social environment under study is homogeneous, but because this is the most natural assumption to make when no information on its organizational structure is available. For example, studies of interethnic marriage that use the U.S. Census as their data source usually assume an equal opportunity marriage market at the national or metropolitan level (Harris and Ono 2005), because the Census provides no information on the social contexts under which people picked their spouses. Often, making no assumption about the structure is practically equivalent to assuming a homogeneous environment and hence equal opportunity—if parameters of the choice model are interpreted as effects of preference.

The school environment has a number of features that would afford the researcher a more refined operationalization of an opportunity structure of friendship choice. As mentioned before, students taking classes together or participating in the same extracurricular activities spend more time with one another and consequently have a greater opportunity to make friends. If data on school activities are available, we should incorporate this information into the opportunity structure. For example, by assuming that opportunity increases in direct proportion to the amount of contact induced by the school structure, we can approximate opportunity with interpersonal exposure and measure it by, say, the number of classes two students share in a week. Although still imperfect, this operationalization of opportunity structure approximates reality more closely than the equal opportunity assumption. In section III, we demonstrate how to construct opportunity structure as a function of the grade level difference between dyads.

In the literature on interpersonal associations, *exposure* and *opportunity* are often used interchangeably. We treat them as separate concepts, with opportunity as an abstract quantity, not tied to any particular environmental constraint, and exposure as a particular type of constraint which results from unequal *contact* with choice alternatives. While opportunity is proportional to choice probability in our framework, we expect exposure not to be related to choice probability in the same way. In friendship and mate choice, for example, choice probability increases with the amount of interpersonal contact, but perhaps not at the same pace. There may exist a saturation point, beyond which greater exposure will not have any additional effect on choice probability. On this account, approximating the opportunity structure of friendship choice with shared class time works only up to a certain level of exposure.

It is often the case that researchers have some knowledge about the structure of a social environment but not the exact quantitative form of opportunity structure. For example, school-based surveys commonly collect data on grade levels, but rarely on classes and activities students share with peers. In a situation like this, it is reasonable to assign a higher opportunity score for students in the same grade level than for those in different grade levels. The question is, how much difference should there be? One possible approach is to parameterize $o_{ij}$ as an exponential function of the attributes of chooser *i* and potential friend *j*, $o_{ij} = \exp(\mathbf{w}_{ij}^{'}\boldsymbol{\delta})$—just as preference is expressed as $a_{ij} = \exp(\mathbf{z}_{ij}^{'}\boldsymbol{\beta})$. This leads to the following expression for choice probability:

$$p_{ij} = \frac{\exp(\mathbf{w}_{ij}^{'}\boldsymbol{\delta} + \mathbf{z}_{ij}^{'}\boldsymbol{\beta})}{\sum_{k \in J} \exp(\mathbf{w}_{ik}^{'}\boldsymbol{\delta} + \mathbf{z}_{ik}^{'}\boldsymbol{\beta})}, \quad (6)$$

which is the exact form of the standard conditional logit model. Equation 6 suggests that as long as the two sets of attributes, $\mathbf{z}_{ij}$ and $\mathbf{w}_{ij}$, are disjoint—that is, there is no variable affecting both preference and opportunity—we can estimate parameters that characterize preference and opportunity. If the two sets of variables are not disjoint, then only the total effects of the overlapping variables are estimated and it is not possible to determine the unique portions attributable to preference and opportunity.

Although a clean separation of preference and opportunity cannot be obtained when the same variables influence both preference and opportunity, it is still beneficial to conduct sensitivity analysis to examine the extent to which inference of preference varies by assumptions about the opportunity structure. Researchers can either estimate preference parameters under various opportunity structures—as we shall demonstrate in Section III—or estimate the total effect of the overlapping variable and then use the multiplicative assumption of POC 3 to arrive at a range for the estimate of preference. Suppose that the probability of an average white student selecting a same-race peer as friend is $c$ times that of selecting a black peer if the two potential friends are otherwise identical. The ratio $c$ is thus the total effect of race on friendship choice (for white choosers), consisting of a portion due to racial homophily and a portion due to in-school racial propinquity—possibly through the practice of tracking. The multiplicative assumption says that the product of the portions due to preference and opportunity equals the total effect. Thus, if the opportunity for same-race friendship choice is $d$ times that of cross-race, racial homophily is $c/d$. We can therefore speculate on the size of $d$ based on knowledge of the opportunity structure in this school (e.g., the proportions of white and black students in academic tracks) to arrive at an estimate of, or a range for, racial homophily.

It would be incorrect to assert that knowledge of opportunity structure is always unattainable. Depending on the problem under investigation, there may be a number of ways to arrive at reasonable estimates empirically. For example, in a supermarket, whether a product is placed near the check-out counters or in its usual section affects its sales. We consider shelf location an opportunity factor because it affects exposure and access to products, but is unlikely to influence customers' intrinsic preference. A simple *opportunity switching experiment* can be conducted to find out the opportunity structure associated with shelf locations. The experiment can be done in two steps. First, place two competing brands, A and B, near check-out counters and in their regular section respectively. Find out the choice probabilities $p_A$ and $p_B$ among customers who bought either A or B but not both. Then, switch A and B's shelf locations and find out the updated choice probabilities $p_A'$ and $p_B'$. The opportunity structure associated with shelf locations, $o_{\text{premium}}/o_{\text{regular}}$, is estimated by the square root of the cross ratio $p_A\, p_B'/\, p_B\, p_A'$.[6] Whether the estimated opportunity structure is specific to Brands A and B or not is an empirical question and can be resolved by multiple experiments involving other commodities. Similar experiments can also be conducted to estimate other conditions of opportunity structure such as in-store advertising. Estimates of opportunity structure from such experiments can then be imported into later studies to estimate parameters of preference. This "borrowing" of opportunity structure is

---

[6] This estimation requires the assumption that consumers' relative preferences for A and B are constant before and after the two brands switching shelves. Before switching, $a_A \propto p_A/o_{premium}, a_B \propto p_B/o_{regular}$. After switching, $a_A \propto p'_A/o_{regular}, a_B \propto p'_B/o_{premium}$.

Under the assumption of constant preference: $a_A/a_B = \dfrac{p_A/o_{premium}}{p_B/o_{regular}} = \dfrac{p'_A/o_{regular}}{p'_B/o_{premium}}$. Rearranging the terms gives $o_{premium}/o_{regular} = \sqrt{\dfrac{p_A p'_B}{p_B p'_A}}$.

made possible by the conceptual separation of preference as traits of decision makers and opportunity as properties of choice contexts in the POC framework.

To recapitulate, in order to infer preference from observed choices, the researcher must be able to answer the counterfactual question: what would the choice probabilities be if the decision maker were indifferent to the alternatives? Hence, the inference of preference requires the researcher to operationalize and quantify the opportunity structure. Methods are available for empirically estimating opportunity structures. Nevertheless, researchers are often faced with the situation where factors affecting opportunity are known, but not the exact form of dependency. We discussed two scenarios for this case: if the sets of factors affecting preference and opportunity do not overlap, the standard conditional logit model can be used to estimate parameters of preference and opportunity; if there are factors affecting both preference and opportunity, the researcher may conduct a sensitivity analysis to arrive at a range of preference estimates based on available knowledge and assumptions about the opportunity structure. The preceding discussion also suggests that adding variables to the right-hand side of the equation to capture the effects of opportunity on choice should not be taken as an automatic solution to the problem of separation. In most situations, the exact form of the opportunity structure must be known in order to infer preference; and the correct adjustment to preference estimates is to add $\ln(o_{ij})$ as an offset variable. In sum, the separation of preference and opportunity is not free knowledge and can only be obtained by supplying the necessary information, i.e., the opportunity structure.

## II. Application of POC to Friendship Choice

In this section we apply the preference-opportunity-choice framework to intergroup friendship choice, focusing on the following issues: the extension of CLO to selections of multiple friends, measurement and estimation of intergroup preference, choice set aggregation, and comparison of CLO with alternative models for friendship choice.

*Three Types of Friendship Choice Data*

Friendship choice data can be collected through roster-based friendship nominations. This is where a survey is conducted in a well-defined social environment such as a school or a workplace, and each respondent is asked to nominate his/her friends from a roster of individuals in that social environment.[7] Using roster-based nomination data enables us to model friendship choice out of a well-defined choice set, and in that, is a much more satisfactory survey method than using free nomination data, where respondents provide characteristics of their friends but little is known about the context of choice (e.g., who could potentially be chosen as friends). In general, there are three ways in which researchers ask respondents to nominate friends from a roster:

1. In *best-friend selection*, respondents are asked to name their single best friends.
2. In *ordered selections*, respondents are asked to nominate up to a predetermined number of friends in order of closeness.
3. In *unordered selections*, respondents are asked to name up to a predetermined number of friends without specifying the order of closeness.

We regard best-friend selection as a basic choice problem, which can be directly modeled with CLO. Selections of multiple friends can be handled as extensions of the basic selection problem. Next we derive statistical models for ordered and unordered selections.

Ordered Selections

Ordered selections may be modeled as a sequence of best-friend selections. First, *i* selects the best friend from choice set *J*. The selected individual is then removed from the choice set, and *i* selects

---

[7] A question arises as to whether students also have friends outside their circle of schoolmates. Under the independence of irrelevant alternatives (IIA) assumption, however, this does not affect the relative choice probabilities among schoolmates.

the best friend from the remaining potential friends. This process is repeated until finally the last friend is selected. Let *m* be the number of rank-ordered friends *i* selects from choice set *J*. We model the probability of rank-ordered selections conditional on *m*. Each selection is regarded as a best-friend choice problem and modeled with CLO. In modeling ordered selections, we assume that the choice probabilities of selecting the *m* friends are independent. We call this assumption *irrelevancy of past choices*.

Let $p_i(j_1, j_2,..., j_m \mid J)$ denote the probability that chooser *i* selects $j_1$ as the best friend, $j_2$ as the second best friend,…, out of choice set *J*. $p_i(j_1 \mid J)$ is the probability of best friend selection. This notation refers to the same quantity as the previous $p_{ij}$ except that we now make the choice set explicit. Under the irrelevancy-of-past-choices assumption, $p_i(j_1, j_2,..., j_m \mid J)$ can be written as the product of *m* choice probabilities, each modeled by CLO:

$$\begin{aligned} & p_i(j_1, j_2,..., j_m \mid J) \\ & = p_i(j_1 \mid J) p_i(j_2 \mid J - \{j_1\}) ... p_i(j_m \mid J - \{j_1,..., j_{m-1}\}) \\ & = \frac{o_{ij_1} \exp(\mathbf{z}'_{ij_1}\boldsymbol{\beta})}{\sum_{h \in J} o_{ih} \exp(\mathbf{z}'_{ih}\boldsymbol{\beta})} \cdot \frac{o_{ij_2} \exp(\mathbf{z}'_{ij_2}\boldsymbol{\beta})}{\sum_{h \in J - \{j_1\}} o_{ih} \exp(\mathbf{z}'_{ih}\boldsymbol{\beta})} \cdot ... \cdot \frac{o_{ij_m} \exp(\mathbf{z}'_{ij_m}\boldsymbol{\beta})}{\sum_{h \in J - \{j_1,..., j_{m-1}\}} o_{ih} \exp(\mathbf{z}'_{ih}\boldsymbol{\beta})} \end{aligned} \quad (7)$$

In this sequence of selections, the choice set *J* reduces to *J*-{$j_1$} at the second selection, to *J*-{$j_1$, $j_2$} at the third selection, etc.

Under the assumption of equal opportunity, (7) simplifies to

$$\begin{aligned} & p_i(j_1, j_2,..., j_m \mid J) \\ & = \frac{\exp(\mathbf{z}'_{ij_1}\boldsymbol{\beta})}{\sum_{h \in J} \exp(\mathbf{z}'_{ih}\boldsymbol{\beta})} \cdot \frac{\exp(\mathbf{z}'_{ij_2}\boldsymbol{\beta})}{\sum_{h \in J - \{j_1\}} \exp(\mathbf{z}'_{ih}\boldsymbol{\beta})} \cdot ... \cdot \frac{\exp(\mathbf{z}'_{ij_m}\boldsymbol{\beta})}{\sum_{h \in J - \{j_1,..., j_{m-1}\}} \exp(\mathbf{z}'_{ih}\boldsymbol{\beta})}, \end{aligned} \quad (8)$$

which is of the form of *rank-ordered logit model* or *exploded logit model* (for a detailed discussion of this model see Allison and Christakis 1994). The rank-ordered logit model can be written as the product of multiple conditional logit models. When the chooser makes exactly one choice, (7) reduces to CLO in (4), and (8) to the standard conditional logit model in (5) respectively. The rank-ordered logit model is identical to a stratified Cox proportional hazards regression in survival analysis, where the risk set is sequentially reduced by event occurrences. Alternatives that are not selected may be regarded as censored cases. To estimate (7), we can use any standard package written for the Cox proportional hazards regression, with $\ln(o_{ij})$ included as an offset variable. When estimating the rank-ordered logit model as a Cox regression, observations need to be stratified by respondent so that choices and risk sets of different decision makers are not pooled in estimation. Readers are referred to Allison and Christakis (1994) on the estimation of this model and the extension to ties in the ranks.

Rank-ordered friendship data contain more information than best-friend nomination data, allowing us to explore a wider range of research questions. For example, we may investigate whether preferences vary by rank order or by the number of nominations respondents provide. Using loglinear methods, Yamaguchi (1990) found that while individuals in general tend to choose similar people as friends, homophily is more pronounced among those who report fewer friends.

Unordered Selections

We now turn to the case of unordered selections. Unordered selections may be considered as generated by the same selection process as ordered selections, but with missing rank order information. We use $p_i(\{j_1,..., j_m\} \mid J)$ to denote the probability of $i$ selecting $j_1$, $j_2$, … $j_m$ as an unordered set of friends from choice set $J$, distinct from the notation $p_i(j_1,..., j_m \mid J)$ for ordered selections. As in ordered selections, we model this probability conditional on the number of nominations $m$. $p_i(\{j_1,..., j_m\} \mid J)$ may be written as the sum of the choice probabilities of all possible permutations of rank-ordered selections. For example, suppose that $m = 2$ and $a$ and $b$ are the chosen friends from choice set $J$. $p_i(\{a,b\} \mid J)$ may be expressed as the sum of two rank ordered choice probabilities: $p_i(\{a,b\} \mid J) = p_i(a,b \mid J) + p_i(b,a \mid J)$. That is, either $a$ is the best friend and *b* is the second best friend, or *b* is the best friend and *a* is the second best friend. Let $G_i$ denote the set of all permutations of the *m* alternatives that *i* selects out of choice set *J*. The number of elements in $G_i$ is $n!/(n-m)!$. Let $g = (g_1, g_2, \ldots g_m)$ denote an element of $G_i$. That is, *g* is a particular permutation of $\{j_1, j_2, \ldots, j_m\}$. The choice probability of unordered selections is

$$p_i(\{j_1,\ldots,j_m\} \mid J) = \sum_{g \in G_i} \prod_{r=1}^{m} \frac{o_{ig_r} \exp(\mathbf{z}_{ig_r}^{'} \boldsymbol{\beta})}{\sum_{s=r}^{m} o_{ig_s} \exp(\mathbf{z}_{ig_s}^{'} \boldsymbol{\beta}) + \sum_{h \in J - \{j_1,\ldots,j_m\}} o_{ih} \exp(\mathbf{z}_{ih}^{'} \boldsymbol{\beta})}. \quad (9)$$

The term $\prod_{r=1}^{m} \frac{o_{ig_r} \exp(\mathbf{z}_{ig_r}^{'} \boldsymbol{\beta})}{\sum_{s=r}^{m} o_{ig_s} \exp(\mathbf{z}_{ig_s}^{'} \boldsymbol{\beta}) + \sum_{h \in J - \{j_1,\ldots,j_m\}} o_{ih} \exp(\mathbf{z}_{ih}^{'} \boldsymbol{\beta})}$ is a re-expression of (7) for the probability of selecting an ordered set of *m* friends [8]

Equation 9 can also be estimated as a conditional logit model with $\ln(o_{ij})$ included as an offset variable. As with the model for rank-ordered choice, if only one friend is selected, (9) reduces to the model for best-friend selection. Under the assumption of equal opportunity, (9) reduces to the following equation, known as the conditional logit model with multiple positive outcomes (Stata Reference Manual 2005):

$$P_i(\{j_1,\ldots,j_m\} \mid J) = \sum_{g \in G_i} \prod_{r=1}^{m} \frac{\exp(\mathbf{z}_{ig_r}^{'} \boldsymbol{\beta})}{\sum_{s=r}^{M_i} \exp(\mathbf{z}_{ig_s}^{'} \boldsymbol{\beta}) + \sum_{h \in J - \{j_1,\ldots,j_m\}} \exp(\mathbf{z}_{ih}^{'} \boldsymbol{\beta})}. \quad (10)$$

*Measuring and Estimating Intergroup Preference*

The major goal of friendship research is to infer preference patterns—i.e., who is attracted to whom—from choice behavior. In particular, sociologists are interested in intergroup (e.g.,

---

[8] The second summation in the denominator in (9) refers to the part of the choice set that is not chosen—$J$ - $\{j_1,\ldots,j_m\}$—and the first summation in the denominator refers to the remaining of $\{j_1,\ldots,j_m\}$ as *r* increases from 1 to *m*.

interracial, inter-faith, and inter-class) relations revealed in friendship choice because they reflect social distances and hierarchy among the various sociodemographic groups. We introduce the *attraction matrix* **A** as a measure of intergroup preference, with element $\mathbf{A}_{vw}$ representing the strength of attraction that members of group $v$ feel for members of group $w$. Typically, the grouping scheme is the same for both choosers and potential friends, resulting in a square matrix with diagonal cells indicating in-group preferences and off-diagonal cells indicating intergroup preferences. We define element $\mathbf{A}_{vw} \equiv \frac{p_{ij}/o_{ij}}{p_{ik}/o_{ik}}$, $i,k \in v, j \in w$. This definition scales **A** in such a way that all elements are divided by the diagonal cells. All in-group preferences are set to 1 by the scaling. With off-diagonal cells, for which $v \neq w$, $\mathbf{A}_{vw}$ is interpreted as the ratio of two choice probabilities: the probability that a typical chooser in group $v$ selects an individual in group $w$ as a friend to the probability that the chooser selects an individual from his/her own group, given that the two potential friends have equal opportunity to be chosen *a priori*.

As a measure of intergroup preference, **A** has some nice invariance properties. First, with the adjustment for opportunity, $\mathbf{A}_{vw,}$ it is essentially a measure of preference, not of choice. Hence, it is free from the influence of contextual factors such as interpersonal interaction opportunities. Second, as a relative risk (i.e., ratio of choice probability of $j \in w$ versus $j \in v$), $\mathbf{A}_{vw}$ is not distorted by total population size or choosers' *friendliness*, i.e., the number of people choosers consider as friends. Note that both population size and friendliness are, to a large extent, characteristics of survey instruments, and thus extrinsic to preference. For example, researchers may ask respondents to make single nominations or unlimited nominations of friends, from a large choice set or a small one (e.g., a school roster versus a class roster). The absolute probability of selection $p_{ij}$ is approximately proportional to the number of nominations and inversely proportional to choice set size. But the relative risk, $p_{ij}/p_{ik}$, is not directly dependent on either factor. With these two invariance properties, **A** allows for comparisons of intergroup preferences across choice contexts and data collection methods.

The intergroup attraction matrix **A** is scaled as ratios to in-group attraction. This scaling is for the convenience of comparing preference across choosers. In discrete choice models, it is straightforward to compare the same chooser's preferences for different alternatives (i.e., $\mathbf{A}_{vw}$ versus $\mathbf{A}_{vw'}$). However, comparison across choosers (e.g., $\mathbf{A}_{vw}$ versus $\mathbf{A}_{v'w}$ or $\mathbf{A}_{vw}$ versus $\mathbf{A}_{wv}$) is ambiguous, because the comparison may be made at either the absolute or the relative level. For example, the question—are whites more attracted to blacks than are blacks to whites?—may be answered differently depending on whether the absolute-level or the relative-level of attraction is being compared. It is possible that whites are less likely to regard blacks as friends than *vice versa*, and yet at the same time, they rank blacks above Hispanics and Asians as the favorite outgroup while blacks rank whites as the least favorite outgroup. (This could happen if there is a large difference in the friendliness of the two groups.) Scaling **A** by the inverse of in-group attraction eliminates such ambiguity by standardizing cross-chooser comparisons at the relative level.[9]

Now consider the case of interracial friendship choice where students fall into four groups: white, black, Hispanic, and Asian (W, B, H, and A). To estimate **A** using CLO (or its variants for multiple selections), we include a vector of sixteen indicators $\mathbf{x}_{ij} = (x_{ij}^{WW}, x_{ij}^{WB}, x_{ij}^{WH}, ..., x_{ij}^{AA})'$ as predictors to represent the race of chooser *i* and that of potential friend *j*. For example, if *i* is black and *j* is white, then $x_{ij}^{BW} = 1$ and all other 15 indicators take on the value of 0. Let $\boldsymbol{\beta} = (\beta^{WW}, \beta^{WB}, \beta^{WH}, ..., \beta^{AA})'$ be the parameters for **x** in CLO. The intergroup attraction $\mathbf{A}_{vw}$ is given by $\exp(\hat{\beta}^{vw} - \hat{\beta}^{vv})$, where *v*, *w* = W, B, H, or A. The CLO regression used to estimate **A** may include other predictors in addition to **x**. The attraction matrix

[9] Researchers interested in cross-group variation in friendliness should not use this scaling.

estimated from a model with covariates is interpreted as net intergroup preference after the effects of those covariates have been adjusted for.

Let us now return to the earlier example of interracial friendship choice presented in Table 1. Table 1b displays the intergroup attraction matrix estimated under the equal opportunity structure, which follows from the homogeneous environment assumption. The estimated intergroup attraction parameters are easily interpretable. For example, $\mathbf{A}_{WH} = 0.558$ means that a white student's relative risk of choosing a Hispanic friend versus a white friend is 0.558 under equal opportunity. This translates into a choice probability of (0.558/(1+0.558))= 36% for the Hispanic student and 64% for the white student when the choice set consists of these two alternatives only. These probabilities will reduce proportionately if the choice set includes other students as well.[10] All intergroup attractions in Table 1b are smaller than 1, indicating an in-group bias.

*Choice Set Aggregation in Friendship Choice*

As Ben-Akiva and Lerman (1985, p. 31-2) pointed out, specifying the choice set is a crucial step in discrete choice analysis. Although the actual consideration set for friendship choice consists of individuals, for reasons of both data reduction and substantive interests, sociologists often model friendship choice at a group level, in which case the set of choice outcomes consists not of individuals, but of individuals' attributes, such as membership in such groups as racial groups, age groups, or religious groups. In our interracial friendship example, the individual-level approach presents the problem of selecting one or multiple friends out of 955 students, whereas the group-level approach poses a different question: Which racial group does the best friend

---

[10] This is due to the independence of irrelevant alternatives (IIA) assumption of the conditional logit model, which implies that adding another alternative or changing the characteristics of a third alternative does not affect the relative risk ratio of the two alternatives under consideration.

belong to? A "choice" model with {white, black, Hispanic, Asian} as the choice set can be constructed to answer this question. But will inference of preference be different depending on whether the choice model is specified at the individual-level or the group-level? It turns out that the opportunity component in the POC framework has the advantage of bridging the two approaches. Indeed, under certain conditions, the same parameters of intergroup attractions can be estimated using either choice set specification. Previously, we demonstrated how to estimate interracial attractions using the individual-level approach for the friendship nomination data presented in Table 1a. Below we show how to estimate the same parameters using the group-level approach.

Let *J'* denote the group-level choice set {*W*, *B*, *H*, *A*}, $p_{iw}$ denote the probability that *i*'s best friend belongs to racial group *w*, where *w* = *W*, *B*, *H*, or *A*, and $o_{iw}$ denote *i*'s opportunity to have a best friend belonging to group *w*. The group-level choice model is given by

$$p_{iw} = \frac{o_{iw} \exp(\mathbf{x}_{iw}^{'}\boldsymbol{\gamma})}{\sum_{v \in J'} o_{iv} \exp(\mathbf{x}_{iv}^{'}\boldsymbol{\gamma})}, \qquad (11)$$

where $\mathbf{x}_{iw} = (x_{iw}^{WW}, x_{iw}^{WB}, x_{iw}^{WH}, ..., x_{iw}^{AA})'$ is a vector denoting the race of *i* and *i*'s best friend, and $\boldsymbol{\gamma}$ is a vector of parameters. The structure of $\mathbf{x}_{iw}$ is the same as $\mathbf{x}_{ij}$ in the individual-level model. As with the individual-level model, $\mathbf{A}_{vw}$ is given by $\exp(\hat{\gamma}^{vw} - \hat{\gamma}^{vv})$. The question of whether the individual-level and the group-level approaches lead to the same inference of intergroup preference can be restated as follows: are $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ the same parameters?

Under two conditions, the group-level model can be viewed as an aggregate form of the individual-level model, estimating the same parameters. The first condition is that the two models utilize the same information. That is, the individual-level model can only use group membership information on potential friends even if more detailed data are available. This requirement ensures that predictors $\mathbf{x}_{ij}$ in the individual-level model are equivalent to $\mathbf{x}_{iw}$ in the group-level model and take on the same values for all *j's* with group attribute *w*. Second, each chooser selects only one friend so that the probability that the chosen friend belongs to *w* is equal to the sum of the probabilities of each alternative in *w* being selected. With the second condition, group-level choice probability can be expressed as a function of individual-level preference and opportunity:

$$p_{iw} = \sum_{j\in w} p_{ij} = \frac{\sum_{j\in w} o_{ij} a_{ij}}{\sum_{v\in J'}\sum_{k\in v} o_{ik} a_{ik}} = \frac{\sum_{j\in w} o_{ij} \exp(\mathbf{x}_{ij}^{'}\boldsymbol{\beta})}{\sum_{v\in J'}\sum_{k\in v} o_{ik} \exp(\mathbf{x}_{ik}^{'}\boldsymbol{\beta})}. \tag{12}$$

Using the first condition, we can then simplify $\sum_{j\in w} o_{ij} \exp(\mathbf{x}_{ij}^{'}\boldsymbol{\beta})$ by replacing $\mathbf{x}_{ij}$ with $\mathbf{x}_{iw}$ and factoring out $\exp(\mathbf{x}_{ij}^{'}\boldsymbol{\beta})$. Hence, we have the following:

$$p_{iw} = \frac{\sum_{j\in w} o_{ij} \exp(\mathbf{x}_{iw}^{'}\boldsymbol{\beta})}{\sum_{v\in J'}\sum_{k\in v} o_{ik} \exp(\mathbf{x}_{iv}^{'}\boldsymbol{\beta})} = \frac{(\sum_{j\in w} o_{ij})\exp(\mathbf{x}_{iw}^{'}\boldsymbol{\beta})}{\sum_{v\in J'}[(\sum_{k\in v} o_{ik})\exp(\mathbf{x}_{iv}^{'}\boldsymbol{\beta})]}. \tag{13}$$

Equation (13) reveals the link between individual-level and group-level preference and opportunity: $\exp(\mathbf{x}_{iw}^{'}\boldsymbol{\beta})$ is *i*'s preference for *w* and $\sum_{j\in w} o_{ij}$ is the opportunity for *i* to choose *w*. Comparing (13) to the group-level choice model in (11), we see that if the latter is estimated with $o_{iw} = \sum_{j\in w} o_{ij}$,[11] it should yield the same preference parameters as the individual-level model. Hence, the individual-level model and the group-level model indeed lead to the same inference of intergroup attractions under the right conditions.

We can rewrite $o_{iw} = \sum_{j\in w} o_{ij}$ as $o_{iw} = n_w \bar{o}_{iw}$, where $n_w$ is the number of persons belonging to group *w*, and $\bar{o}_{iw} = \frac{1}{n_w}\sum_{j\in w} o_{ij}$ is the average opportunity for selecting someone in

[11] Since respondents cannot nominate themselves as friends, $o_{ij}$ should be set to 0 in calculating $o_{iw}$. For the same reason, when approximating $o_{iw}$ with group size $n_w$ (to be discussed next), use $n_w$-1 for *i*'s own group.

group $w$. This re-expression shows that group-level opportunity depends on two factors: the group size and the average individual-level opportunity. This observation provides the theoretical ground for our previous assertion that opportunity for intergroup friendship choice depends on the population composition and the structure of a social environment. If there is no systematic variation in average individual-level opportunity by group, i.e., $\bar{o}_{iw} = \bar{o}_{iv}$ for all $w, v \in J', w \neq v$, group-level opportunity $o_{iw}$ reduces to group size $n_w$. It is easy to see that in order for $n_w$ to be a proxy for $\sum_{j \in w} o_{ij}$, $\bar{o}_{iw} = \bar{o}_{iv}$ must be true as well. Hence, equal average individual-level opportunity across groups provides the necessary and sufficient condition for treating group size as a proxy for opportunity in models of intergroup friendship choice.

The result that group-level models can estimate the same intergroup preference parameters as individual-level models has two practical uses. First, while individual-level models of friendship choice are superior to group-level models as a representation of real world choice problems, they often involve very large choice sets and thus are computing-intensive. If computation time is a concern, group-level models can be used instead to estimate the same parameters of intergroup attractions. Second, some social surveys collect data on friendship choice by asking respondents to name the social groups their best friends belong to. This type of group-level data is not amenable for discrete choice modeling at the individual level because the choice set of potential friends is unknown. However, if the demographic composition of the social environment is known *and* if there is no particular reason to suspect that average individual-level opportunity varies across groups, then group-level choice models with group size treated as opportunity can be used in place of individual-level models.

*Comparison of Related Models for Friendship Choice*

CLO, the Loglinear Model, and the Log-Rate Model

The CLO is essentially a conditional logit model with an offset variable to account for opportunity structure. The relationship between the CLO and the standard conditional logit model parallels that between two familiar models to sociologists—the log rate model and the loglinear model (for a textbook treatment of these models, see Powers and Xie 2000, Pp. 87-160). The log rate model for contingency tables can be seen as an extension of the loglinear model. It includes an offset variable—usually some measure of *exposure*, such as population for crimes or person-years for deaths—on the right-hand side of the regression equation, analogous to the handling of opportunity in the CLO. Both the log rate model and the loglinear model can be used to analyze cross-tabulated data of friendship nominations. Under certain conditions, they yield the same intergroup attraction estimates as the CLO and the conditional logit models, respectively.

A loglinear model for analyzing cross-tabulated friendship nominations such as those in Table 1a has the following form:

$$\ln(f_{vw}) = \mu + \mu_v^R + \mu_w^C + \mu_{vw}^{RC}, \qquad (14)$$

where $f_{vw}$ is the expected frequency of nominations from chooser group $v$ to target group $w$, $\mu$ is the grand mean capturing the sample size, $\mu_v^R$ is the parameter corresponding to chooser group $v$, $\mu_w^C$ is the parameter corresponding to target group $w$, and $\mu_{vw}^{RC}$ is the interaction between $v$ and $w$. The parameter of interest is $\mu_{vw}^{RC}$, which describes the association between $v$ and $w$. A nice property of the loglinear model is that $\mu_{vw}^{RC}$ is invariant to marginal distributions, due to the inclusion of $\mu_v^R$ and $\mu_w^C$. The usual goal of loglinear analysis is to find the best parameterization of $\mu_{vw}^{RC}$ for a parsimonious representation of row and column association. To obtain the full intergroup attraction matrix, however, a saturated model should be fitted to yield unconstrained $\mu_{vw}^{RC}$ parameters.

A log rate model for Table 1a has the following specification:

$$\ln(f_{vw}) = \ln(E_{vw}) + \mu_{vw}^{RC}, \qquad (15)$$

where exposure $E_{vw}$ can be operationalized as the number of dyads with choosers belonging to group $v$ and potential friends belonging to group $w$, i.e., $E_{vw} = n_v n_w$, for $w \neq v$. Since respondents cannot nominate themselves, $E_{vw} = n_v(n_v - 1)$ for $w = v$. In Table 1c, we present the exposure matrix for interracial friendship choice in this school.

Previous studies (Logan 1983; Diprete 1990; Breen 1994; Xie and Shauman 1997) have shown that the loglinear model can be expressed as the conditional logit model for aggregate data. Analogously, the log-rate model can be viewed as CLO applied to aggregated data. Using the log rate model, we calculate $\mathbf{A}_{vw}$ as $\exp(\hat{\mu}_{vw}^{RC} - \hat{\mu}_{vv}^{RC}) = \dfrac{F_{vw}/E_{vw}}{F_{vv}/E_{vv}}$, where $F_{vw}$ is the observed number of nominations from $v$ to $w$. It can be easily verified that this method yields the same preference estimates as in Table 1b.

Table 2 compares the loglinear model, the log rate model, the standard conditional logit model, and CLO for analyzing friendship choice according to two criteria: whether the model utilizes individual-level data and whether it accommodates unequal opportunity. Researchers should use the loglinear model and the log rate model to analyze aggregated data in tabular form, and use the conditional logit model and the CLO to analyze individual-level data. For a school of $n$ students belonging to $k$ groups, the conditional logit model (with or without the offset of opportunity) with an individual-level choice set analyzes an $n \times n$ sociomatrix, the conditional logit model with an aggregated choice set analyzes an $n \times k$ data matrix, and the loglinear and the log rate models analyze a $k \times k$ contingency table. The three approaches lead to the same $k \times k$ intergroup attraction matrix—as long as group size is properly handled through weighting when the sociomatrix is collapsed in either dimension. As with all data aggregation, the collapse of the original sociomatrix to tabular form involves the trade-off of model specificity and computational simplicity.

[Table 2 About Here]

How is opportunity handled in the loglinear model and the log rate model? In the loglinear model, opportunity structure is represented by the row and column marginals, and the expected frequencies under the assumption of row and column independence serve as the baseline for the inference of row and column associations. The separation of *association* and *structure* (equivalent to preference and opportunity in POC) is achieved by saturating marginal distributions with row and column parameters $\mu_v^R$ and $\mu_w^C$. After saturation, association parameters $\mu_{vw}^{RC}$ are invariant to marginals. However, the approach of saturating marginals falls short of accounting for opportunity structure in intergroup friendship choice because row and column marginals in this case represent choice outcomes, rather than the context of choice. A similar problem exists in loglinear analysis of intergroup marriages, which is typically limited to data on the married population and does not consider an unmarried person as an "opportunity" for partner choice (see Schoen 1986 for a solution to this problem; also see Logan 1996b for a critique of loglinear models).

Opportunity in the log rate model and the CLO typically incorporates information external to the data contained in the contingency table. In the log rate model, exposure usually involves some measure of population, area, time, or their combinations specific to cross-classifications for aggregate data. The use of exposure as an offset term makes the log rate model suitable for rate data—number of events observed per unit of observation window. Unlike exposure in the log rate model, opportunity in CLO can incorporate any quantitative information—not just observation units—that reflects structural constraints on choice outcomes. Although this paper focuses on the application of friendship choice, the idea of adjusting for known, structured influence in the form of an offset variable has wide implications for analysis of categorical outcomes in non-choice situations.

Three Logit Models for Modeling Friendship Nomination

A variety of logit models have been used to model friendship nomination as a categorical outcome. Apart from conditional logit models (Quillian and Campbell 2001), previous studies have also employed unconditional (binary) logit models (Hallinan and Williams 1989; Moody 2001) and *p** models—a family of logit models developed for social network analysis (Mouw and Entwisle 2006). *p** models are identical to unconditional logit models in form, but characterized by the use of network characteristics such as mutuality, transitivity, cyclicity, etc. as predictors of friendships. Therefore, in the following discussion, we focus on the comparison of the conditional and the unconditional approaches to modeling friendship nomination.

The major difference between the two approaches is that the conditional logit model (alias McFadden's choice model) is more appropriate for analyzing choice. The conditional logit regression can be derived from a behavioral model of discrete choice (Luce 1959; McFadden 1974; Ben-Akiva and Lerman 1985; Pudney 1989),[12] but the unconditional logit model, as far as we know, is not similarly founded on a behavioral model. Statistically, the two models are different in that the choice probability in the conditional logit model depends on *all* alternatives in *i*'s choice set ( $p_{ij} = \frac{\exp(\mathbf{z}_{ij}^{'}\boldsymbol{\beta})}{\sum_{k \in J_i} \exp(\mathbf{z}_{ik}^{'}\boldsymbol{\beta})}$ ), whereas in the unconditional logit model, it is a function of

---

[12] There are two fundamentally different approaches to modeling choice behavior—the constant utility approach and the random utility approach. It has long been shown that the two approaches are equivalent under certain distributional assumptions of random utility (McFadden 1974; Yellott 1977).

the characteristics of *i* and *j* only ( $p_{ij} = \dfrac{\exp(\mathbf{z}_{ij}^{'}\boldsymbol{\beta})}{1+\exp(\mathbf{z}_{ij}^{'}\boldsymbol{\beta})}$ ). Hence, the conditional approach implies that friendship nomination is a choice process, which involves comparison of alternatives, while the unconditional approach implies that people make independent decisions about each potential friend without comparing the alternatives. When the number of nominations is capped low by the survey instrument—especially when only the best friends are asked for—friendship nomination is a choice process, better captured by the conditional logit model. When the number of friends is unconstrained or the limit is high, the likelihood of nomination is less likely to be constrained by other alternatives available to the decision maker, and the unconditional approach may be appropriate.

Both the unconditional and the conditional logit models focus on dyadic relationships. The advantage of the dyadic approach is that intergroup preferences estimated from models with dyads as units of analysis are undistorted by group size. This contrasts with logit models using individuals as units of analysis, e.g., a binary logit model predicting the probability of having an interracial friend, or a multinomial logit model predicting the racial group of the best friend. The multinomial logit model is as follows:

$$p(y_i = w) = \frac{\exp(\mathbf{x}_i^{'}\boldsymbol{\gamma}_w)}{\sum_{v \in J'} \exp(\mathbf{x}_i^{'}\boldsymbol{\gamma}_v)}, \tag{16}$$

where $\exp(\mathbf{x}_i^{'}\boldsymbol{\gamma}_w)$ represents the strength of attraction *i* feels for group *w*, and $\gamma_w$ is a vector of group-specific parameters. It is a well-known fact that multinomial logit models can be expressed as conditional logit models. That is, (16) is equivalent to $p(y_i = w) = \dfrac{\exp(\mathbf{x}_{iw}^{'}\boldsymbol{\gamma})}{\sum_{v \in J'} \exp(\mathbf{x}_{iv}^{'}\boldsymbol{\gamma})}$, which is exactly the group-level conditional logit model in (11) without the opportunity multiplier. The multinomial logit model thus yields preference parameters that are not corrected for the influence of opportunity on choice.[13] However, with individuals as the units of analysis, the multinomial logit model cannot incorporate outcome-specific opportunity $o_{iw}$ as an offset variable.

* * *

There are a few issues specific to *friendship* choice we do not attempt to resolve because they are beyond the central topic of this paper. One such issue is that friendship choices are affected by network dynamics such as reciprocity and transitivity, and thus the assumption of independence between respondents may not hold. The $p$* models and random effects models offer promising solutions to the problem of interdependency. $p$* models predict directed or undirected friendship ties explicitly as functions of network characteristics. Random effects models can be applied to dyadic analysis to account for correlations due to common actors (i.e., correlations among $p_{i\cdot}$) and shared targets (i.e., correlations among $p_{\cdot j}$) via error covariance structures. Readers are referred to those two bodies of literature for more information (see Wasserman and Pattison 1996; Anderson et al. 1999; Pattison and Wasserman 1999 for p* models; and see Raudenbush and Bryk 2002; Hoff 2003; Hoff 2005 for random effects models).

## III. An Empirical Example

*Data*

To illustrate the various methods discussed so far, we analyze data from the National Longitudinal Study of Adolescent Health (Add Health), a school-based study of adolescents enrolled in grades 7 through 12 in 1994-5. The Add Health study is an ideal data set for studying friendship choice. One component of the survey—the in-school questionnaire—was administered to every student in the sampled schools. In addition to collecting information on family

---

[13] A special case occurs when all respondents face the same opportunity structure. In this situation, only the intercepts, but not the other coefficients, of the multinomial logit model are biased.

background and school-related activities, the questionnaire asked students to name their friends from a school roster.[14] The roster-based nominations, coupled with individual-level data on all students, enable us to model friendship choices from well-defined choice sets.

Students were instructed to name up to five best female friends and five best male friends separately in order of closeness. Romantic partners, if respondents had any, were also included in the nominations. We limit the analysis to same-sex friends because romantic relationship is likely determined by a different selection process than friendship. Across the 132 schools with valid data on friendship choice, there are over 9 million female-female and over 8 million male-male directed dyads, and only 1% of those are friends. We took a random sample of non-friendship dyads while retaining all friendship nominations. This results in a total of 480,215 female dyads and 398,571 male dyads.[15]

*Research Design*

In the following analysis, we test two hypotheses with regard to the effects of race, age, academic achievement, and socioeconomic backgrounds on friendship choice.[16] The first hypothesis is

---

[14] The roster lists all students of the school which the respondent attends (i.e., the focal school) and its sister school if there is one. A sister school is a middle school or a high school that has a feeder relationship with the focal school. It either supplies most of the focal school's incoming students or enrolls most of its graduates. The analysis here limits the choice set to the focal school.

[15] In logit models (e.g., case control studies), selection on outcome does not bias estimates. McFadden (1978) and Parsons and Kealy (1992) have shown that the conditional logit model also has this property.

[16] Here we focus on individual-level predictors. In another paper, we incorporate school-level fixed and random effects into friendship choice models.

*homophily*—similarity in personal characteristics enhances the likelihood of friendship. The second hypothesis, which has not been tested before, is *status asymmetry*—when distance in status is equal, there is a greater tendency to nominate the person with a higher status (in terms of age, GPA, or SES) than the one with a lower status. This hypothesis predicts that, for example, a 14-year-old is more likely to nominate a 16-year-old than a 12-year-old as a friend, even though both potential friends have 2-year age gap with the chooser.

Related to status asymmetry is the issue of whether friendship choice should be treated as a one-sided or a two-sided choice. A typical two-sided choice is marriage matching where relationship is symmetric by definition: *i* is married to *i* if and only if *i* is married to *i*. Friendship nomination data are not symmetric by definition. Indeed, only 40% of the friendship nominations in Add Health were reciprocated (Mouw and Entwisle 2006). If friendship is a two-sided relationship by nature, then the 60% unreciprocated nominations must be due to huge measurement errors. Testing the status asymmetry hypothesis can shed light on this issue because random measurement errors are unlikely to cause bias in favor of higher-status friends (or lower-status friends, for that matter). Therefore, if we find strong evidence supporting status asymmetry, then friendship cannot be regarded as a symmetric relation.

To test the homophily hypothesis, we include variables indicating racial groups of respondent *i* and potential friend *j* as well as the absolute differences in age, GPA, and SES between *i* and *j*. The racial groups used in this analysis are mutually exclusive categories of non-Hispanic white, Hispanics, non-Hispanic black, and non-Hispanic Asian. GPA is calculated as the average grade of four subjects—English, math, social studies and science—with each grade first standardized within subject and school. Dyadic difference in GPA ranges from 0 to 4. SES is measured by mother's years of schooling. If the homophily hypothesis is true, the likelihood of friendship selection should decrease with status distance. Likewise, we expect the likelihood of selection to be smaller across racial boundaries than within racial boundaries.

To test the status asymmetry hypothesis with respect to age, GPA, and SES, we include interactions between the three status distance variables and indicators for the direction of status difference (e.g., $\beta_1$ |AgeDiff$_{ij}$| + $\beta_2$ $\delta_{ij}$ |AgeDiff$_{ij}$|, where $\delta_{ij} = 1$ if Age$_j$ > Age$_i$, $\delta_{ij} = 0$ if otherwise). This parameterization allows the effect of status distance to differ for dyads where the potential friend is of higher status (denoted by "alter > ego") and where the potential friend is of lower status ("alter < ego"). If both the homophily and the status asymmetry hypotheses are true, the negative effect of status distance on friendship choice should be smaller for "alter > ego" dyads than for "alter < ego" dyads.

In testing the two hypotheses, we estimate models using three types of nomination data—best friend selection, ordered selection, and unordered selection of multiple friends. As mentioned earlier, each respondent in the survey provides up to five rank-ordered friends of each gender. In modeling best-friend selection, we retain the top ranked friend and treat friends of lower ranks as non-friends along with those who were not nominated. In modeling unordered selections, we simply ignore the rank-order information on the multiple friends selected. While we compare coefficients across types of nomination, we do not formally test their differences because the models utilize different data and as such are not directly comparable.

Recall that in the POC framework, inference of preference is dependent on the specification of opportunity structure. When the researcher is unsure about the exact quantitative form of opportunity, sensitivity analysis can be conducted to illuminate how preference estimates vary by assumptions about opportunity. To demonstrate this approach, we estimate friendship choice models under three different opportunity assumptions. The first assumption is that of the homogenous environment, which leads to the equal opportunity structure (or E in short) and hence the standard conditional logit model. We expect this opportunity structure to introduce an upward bias in age homophily because it fails to account for segregation by grade levels within schools. To explore the magnitude of this bias, we also estimate models under two other opportunity structures, both constructed as functions of grade levels. The "gradient opportunity structure" (or G in short) specifies that opportunity is a continuous decreasing function of the distance in grade levels between a pair of students. Specifically, we assume $o_{ij} = \frac{1}{(d_{ij}+1)^2}$, where $d_{ij}$ is the difference in grade levels between *i* and *j*. The "dichotomous opportunity structure" (or D in short) specifies that opportunity is greater between students in the same grade than between those in different grades by a factor of 6.[17] That is, we assume the following opportunity structure: $o_{ij} = \begin{cases} 6 & if\, d_{ij} = 0 \\ 1 & if\, d_{ij} \neq 0 \end{cases}$.

We expect estimates of race, GPA, and SES to be relatively stable across the three opportunity structures tested here because the distributions of race, GPA, and SES are unlikely to vary substantially across grades within schools. The main opportunity factor influencing preference estimates of race, GPA, and SES may very well be tracking, which was implemented in more than half of the schools in the sample. Unfortunately, the Add Health survey only provides proportions of twelve-graders in various tracks at the school level; we do not know the proportions broken down by race, GPA, or SES. Therefore, we could not examine the level the in-school segregation due to tracking or the effects of tracking on intergroup relations. If individual-level data were available on tracking, it could be incorporated into the opportunity structure, analogous to grade levels.

*Results*

A total of 18 models were estimated for combinations of 3 types of nominations, 3 opportunity structures, and 2 genders. Within each gender, we label the models as E1, E2, E3, G1, G2, G3,

[17] We came up with the factor 6 by assuming that (a) opportunity is proportional to the amount of time a pair of students spend together and (b) the average student shares 6 classes with students in the same grade and one elective class with students in all other grade levels.

D1, D2, D3, with E, G, and D denoting opportunity assumptions, and 1, 2, and 3 denoting best-friend selection, ordered selection, and unordered selection, respectively. We first examine the effects of age, GPA, and SES differences on friendship choice in Table 3, and then discuss interracial attractions in Table 4. Estimates in Table 3 are relative risk ratios (rrr), interpreted as multiplicative effects on the relative risk of selection. For example, a coefficient of 0.339 for age difference means that the relative risk for any given chooser *i* to select *j* versus *k* as a friend ( $p_{ij} / p_{ik}$ ) is 0.339 if the age difference between *i* and *j* is 1 year greater than that between *i* and *k*. The rrr of 0.339 applies only to alter < ego dyads, and increases to 0.347 for alter > ego dyads. All coefficients in Table 3 are statistically significant. Asterisks indicate not the significance of the coefficients themselves, but the significance of the status asymmetry hypotheses, i.e., whether the coefficients are different for "alter > ego" and "alter < ego" dyads.

[Table 3 About Here]

In Table 3, we observe a strong pattern of homophily with respect to age and GPA. For female dyads, one year's age difference reduces the relative risk of selection by approximately 2/3 in the equal opportunity models and by about 40% in the gradient and dichotomous opportunity models. One unit of GPA is associated with an rrr between 0.61 and 0.65 across the board for girls. Homophily based on socioeconomic status is considerably weaker but still statistically significant. One unit of SES difference is associated with an rrr of about 0.95.[18] The estimates for boys are slightly higher than those for girls, indicating a somewhat lower level homophily in male-male friendship.

---

[18] The levels of homophily by GPA and SES are actually much closer. The relative risk ratios associated with one standard deviation of GPA difference and SES difference are about 0.8 and 0.9 respectively.

Furthermore, the effects of age, GPA, and SES differences as friendship barriers depend on the direction of the difference. Other things being equal, status distance has a greater negative effect when the potential friend is of lower status. The parameters testing the asymmetry hypothesis are statistically significant almost everywhere with the exception that girls exhibit the same level of discrimination against alters with lower GPA as against alters with higher GPA. Although status asymmetry is much weaker than homophily in strength, the evidence in Table 3 clearly bears it out as a general preference principle because relative preference for persons of higher status is observed with respect to all personal traits and the result is robust to variation in model specification. The finding of status asymmetry is significant also because it provides empirical grounds for treating friendship as a directed relationship. As we noted earlier, if friendship should indeed be seen as a two-sided choice, then the tendency to nominate, say, older friends should balance the tendency to nominate younger friends, subject to innocuous sampling and measurement errors. In that case, we would not have found a ubiquitous pattern of status asymmetry.

We now turn to comparisons across opportunity structures and types of selection. The estimated coefficients do not vary substantially by the type of selection. This suggests that students probably used similar criteria in selecting top ranked and less close friends. Future research should test this directly by interacting preference parameters with rank-order. In addition, estimates do not vary much by opportunity structure either—except for those pertaining to age. As expected, the inclusion of $\ln(o_{ij})$ as an offset term in G and D models has a huge impact on estimates of age homophily, with relative risk ratios increasing from 0.3~0.4 to 0.6~0.7 across the various models. Estimates of GPA and SES homophily are hardly affected by changes in the opportunity structure at all, due to their being uncorrelated with $\ln(o_{ij})$ in this data set. In comparison, correlations between age difference and $\ln(o_{ij})$ in the gradient and dichotomous opportunity structures are 0.5 and 0.7 respectively.

Table 4 displays estimated interracial attraction matrices for female and male dyads from model E1 (i.e., best friend nomination and the equal opportunity structure). Results estimated from the other models are similar and therefore not presented here. As with GPA and SES, estimates of interracial attractions are insensitive to opportunity structure based on grade levels, due to the orthogonality of these variables with the assumed opportunity structure. The "baseline model" in Table 4 contain race variables only, while "Model E1" also includes the covariates in Table 3. The coefficients are interpreted as the ratio of the probability of selecting a cross-race friend to that of selecting a same-race friend, everything else being equal. We see a very high level of racial homophily: with a couple of exceptions, estimated interracial attractions are in the neighborhood of 0.1~0.2. In addition, the strength of interracial attraction varies substantially by chooser's race and potential friend's race. Compared to minority groups, whites show less in-group bias. As Table 4 shows, the white-Hispanic and white-Asian attractions are about 0.5 and the white-black attraction is 0.15~0.2. Interracial attractions involving minority groups as choosers fall between 0.07 and 0.28, with the lowest observed between blacks and whites. In addition, a comparison of the baseline models to the E1 models indicates that estimates of interracial attractions are only slightly modified when age, GPA, and family SES are added. Thus, racial differences in GPA and SES contribute very little to the racial cleavage in friendship choice.

[Table 4 About Here]

In Table 5, we compare our interracial attraction estimates with results from two other studies that also used Add Health data, Quillian and Campbell (2003) and Mouw and Entwisle (2006)—hereafter Q&C and M&E, respectively.[19] Despite differences in research design, all

---

[19] The results in Table 5 are based on Table 1 of Quillian and Campbell (p. 551) and Table 6 of Mouw and Entwisle (p.416). In their analysis, Q&C separated Hispanics into three subgroups—white Hispanics, black Hispanics, and other Hispanics. For the purpose of comparison, we created weighted averages for a single Hispanic group, with weights determined by the sizes of

studies yielded estimates interpretable as interracial attractions. As Table 5 shows, the estimates are largely consistent except for those pertaining to Hispanics, which vary considerably across the three studies. For example, estimates of Hispanic-black attraction range from 0.153 to 0.803 and those of Hispanic-white attraction range from 0.153 to 0.611. We note several differences in sample selection and research design that may have contributed to some of the inconsistencies across studies. For example, M&E used the "in-home" subsample as opposed to the "in-school" full sample used by Q&C and us. The in-home survey was administered to a random sample of students, with oversamples of certain sociodemographic groups such as blacks with college-educated parents, Cubans, Puerto Ricans, Chinese, siblings, etc. The oversamples of Cubans and Puerto Ricans in M&E's analysis may have lead to a higher Hispanic-black attraction than what would be expected for the general Hispanic population. The research design of Q&C differs from the other two studies in that Q&C divided the target groups—but not the chooser groups—by immigration generational status and estimated intergroup attractions for race × (race by generation) combinations.[20] The results presented in Table 5 pertain to the third+ generation only

---

the Hispanic subgroups in the sample, which are 4%, 2% and 9% respectively. M&E showed that estimates of interracial attractions increase substantially when networks variables are introduced into the model, in support of their argument that network processes in friendship choice intensify racial homophily. Because neither of the other two studies used network variables as predictors, to facilitate comparison, we calculated interracial attractions for M&E's study based on their results from Model 4, which included an extensive list of controls, but not network variables. Estimates from our own study are weighted averages of interracial attractions for the female sample and the male sample, presented earlier under Model H1 in Table 4.

[20] Q&C's purpose for separating target groups by generational status was to test the hypothesis that racial homophily weakens across generations. However, since chooser groups were not

(i.e., those with U.S.-born parents); their intergroup attraction estimates for first and second generation immigrants are generally lower for white Hispanic, other Hispanic, and Asian choosers, but not necessarily for the other groups (note presented here). This could potentially explain why Q&C's interracial attraction estimates are somewhat higher than ours.

Finally, we examine gender differences in preference. As shown in both Table 3 and Table 4, girls exhibit a higher level of homophily than boys. The smaller coefficients for female dyads indicate that they have a smaller tendency than their male counterparts to select friends who are dissimilar to themselves with respect to age, GPA, family SES, and race. Additional analyses with combined two-sex samples and interactive terms confirmed that gender differences in homophily are statistically significant. This finding is consistent with the observation from adolescent peer group studies that female cliques tend to be smaller, closer, and more homogenous than male cliques (McPherson et al. 2001).

## IV. Conclusion

In the preceding sections, we proposed a framework for conceptualizing and analyzing preference and opportunity in discrete choice. We have limited ourselves to the problem of disentangling preference and opportunity as proximate determinants of choice outcomes. In sociological research, the more interesting question often lies in person-environmental interactions, that is, in the causal effects of preference and opportunity on one another. On the one hand, people seek out social groups they prefer and, in so doing, actively increase exposure to their preferred environment. On the other hand, social environment can alter underlying preferences. This set of causal relationships is illustrated by the following social processes involved in adolescent friendship choice: parents send their children to schools with what they consider the "right"

---

similarly broken down by generational status, it is doubtful that their test effectively evaluated the hypothesis.

sociodemographic compositions for them,[21] and the level of diversity in schools influences children's tendency to make interracial friends. While leaving endogenous causal processes to future research, this paper focuses on untying the knot between preference and opportunity as two—and the only two—proximate determinants of choice. Because preference is not directly observable, any attempt to sort out the dynamic process between preference and opportunity must begin with the inference of preference from choice. Only then do we have the pieces of the puzzle to work with to address more complicated research questions, such as how preference influences opportunity and vice versa.

Our approach to the separation of opportunity and preference is to view opportunity as characteristics of a choice context and preference as underlying dispositions of a person. Thus, opportunity and preference are conceptually distinguished by whether the source of variation in choice resides with the environment or in the psyche. Formally, we define them as choice probabilities under indifference and the deviation from those probabilities respectively. The empirical separation of opportunity and preference requires that the researcher have explicit knowledge about the choice context in the form of opportunity structure. In conclusion, we offer the following recommendations. For analysis of existing data sets, researchers should make explicit their assumptions about opportunity when inferring preference from choice, or better yet, conduct sensitivity analysis to explore the dependency of preference inference on specifications of opportunity structure. For new data collection on choice behavior, survey researchers should pay close attention to the context of choice, which is just as important as choice outcome itself. In addition, researchers should consider and develop methods for ascertaining the unobservable opportunity structure in different situations.

---

[21] In this case, parents exercise school choice on their children's behalf. Nonetheless, opportunity structure may be regarded as endogenous to preference.

Table 1a: Friendship Nominations

| Chooser's Race | Best Friend's Race: White (181) | Hispanic (106) | Black (572) | Asian (96) | No Nomination |
|---|---|---|---|---|---|
| White (181) | 73 | 24 | 34 | 9 | 41 |
| Hispanic (106) | 17 | 55 | 9 | 2 | 23 |
| Black (572) | 35 | 12 | 399 | 7 | 119 |
| Asian (96) | 17 | 7 | 4 | 46 | 22 |

Table 1b: Interracial Attractions

| Chooser's Race | Best Friend's Race: White | Hispanic | Black | Asian |
|---|---|---|---|---|
| White | 1.000 | 0.558 | 0.147 | 0.231 |
| Hispanic | 0.179 | 1.000 | 0.030 | 0.040 |
| Black | 0.277 | 0.162 | 1.000 | 0.104 |
| Asian | 0.194 | 0.136 | 0.014 | 1.000 |

Table 1c: Opportunity

| Chooser's Race | Best Friend's Race: White | Hispanic | Black | Asian |
|---|---|---|---|---|
| White | 32,580 | 19,186 | 103,532 | 17,376 |
| Hispanic | 19,186 | 11,130 | 60,632 | 10,176 |
| Black | 103,532 | 60,632 | 326,612 | 54,912 |
| Asian | 17,376 | 10,176 | 54,912 | 9,120 |

Table 2: A Comparison of Four Models for Studying Friendship Choice

| Unit of analysis | Unequal Opportunity Structure? | |
|---|---|---|
| | No | Yes |
| Aggregate data | Loglinear model | Log rate model |
| Individual-level data | Conditional logit model | Conditional logit model with opportunity |

Table 3: Effects of Age, Average Grades and Family Socioeconomic Backgrounds on Friendship Choice

| | *Best-friend nomination* | | | *Ordered nominations* | | | *Unordered nominations* | | |
|---|---|---|---|---|---|---|---|---|---|
| | E1 | G1 | D1 | E2 | G2 | D2 | E3 | G3 | D3 |
| Panel A: Female-Female nominations | | | | | | | | | |
| *Age difference* | | | | | | | | | |
| alter < ego | 0.339 | 0.626 | 0.585 | 0.342 | 0.637 | 0.594 | 0.338 | 0.629 | 0.587 |
| alter > ego | 0.347 | 0.66* | 0.621* | 0.351 | 0.67* | 0.628* | 0.347 | 0.662* | 0.622* |
| *GPA difference* | | | | | | | | | |
| alter < ego | 0.605 | 0.623 | 0.626 | 0.609 | 0.626 | 0.629 | 0.606 | 0.623 | 0.626 |
| alter > ego | 0.609 | 0.625 | 0.625 | 0.628 | 0.646 | 0.647 | 0.622 | 0.639 | 0.640 |
| *SES difference* | | | | | | | | | |
| alter < ego | 0.928 | 0.930 | 0.931 | 0.925 | 0.927 | 0.928 | 0.923 | 0.926 | 0.926 |
| alter > ego | 0.962* | 0.963* | 0.965* | 0.966* | 0.966* | 0.967* | 0.964* | 0.966* | 0.966* |
| *Observations* | 472215 | 472215 | 472215 | 480215 | 480215 | 480215 | 473095 | 473095 | 473095 |
| Panel B: Male-Male nominations | | | | | | | | | |
| *Age difference* | | | | | | | | | |
| alter < ego | 0.390 | 0.674 | 0.622 | 0.388 | 0.668 | 0.619 | 0.384 | 0.662 | 0.612 |
| alter > ego | 0.424* | 0.747* | 0.69* | 0.425* | 0.746* | 0.69* | 0.422* | 0.741* | 0.684* |
| *GPA difference* | | | | | | | | | |
| alter < ego | 0.648 | 0.664 | 0.667 | 0.642 | 0.658 | 0.661 | 0.639 | 0.653 | 0.656 |
| alter > ego | 0.742* | 0.757* | 0.756* | 0.75* | 0.767* | 0.765* | 0.747* | 0.761* | 0.76* |
| *SES difference* | | | | | | | | | |
| alter < ego | 0.952 | 0.956 | 0.957 | 0.954 | 0.959 | 0.960 | 0.953 | 0.957 | 0.958 |
| alter > ego | 0.973* | 0.977* | 0.978* | 0.981* | 0.985* | 0.986* | 0.98* | 0.984* | 0.985* |
| *Observations* | | 388,373 | | | 398,571 | | | 389,502 | |

*Note:*

a. Estimates are presented in exponential form. All coefficients are statistically significant at 0.001 level.

b. Asterisk indicates that asymmetry is statistically significant at 0.001. That is, the coefficient of "alter > ego" dyads is significantly different from that of the "alter < ego" dyads.

Table 4: Estimated Interracial Attractions in Best-Friend Selection

**Female-Female nominations**

*Baseline model*

| | White | Hispanic | Black | Asian |
|---|---|---|---|---|
| White | 1 | 0.516 | 0.153 | 0.454 |
| Hispanic | 0.092 | 1 | 0.092 | 0.153 |
| Black | 0.066 | 0.090 | 1 | 0.141 |
| Asian | 0.198 | 0.126 | 0.079 | 1 |

*Model E1*

| | White | Hispanic | Black | Asian |
|---|---|---|---|---|
| White | 1 | 0.587 | 0.164 | 0.490 |
| Hispanic | 0.095 | 1 | 0.095 | 0.165 |
| Black | 0.069 | 0.101 | 1 | 0.163 |
| Asian | 0.212 | 0.165 | 0.090 | 1 |

**Male-Male nominations**

*Baseline model*

| | White | Hispanic | Black | Asian |
|---|---|---|---|---|
| White | 1 | 0.513 | 0.191 | 0.570 |
| Hispanic | 0.213 | 1 | 0.213 | 0.267 |
| Black | 0.074 | 0.159 | 1 | 0.131 |
| Asian | 0.247 | 0.136 | 0.129 | 1 |

*Model E1*

| | White | Hispanic | Black | Asian |
|---|---|---|---|---|
| White | 1 | 0.573 | 0.205 | 0.598 |
| Hispanic | 0.217 | 1 | 0.217 | 0.274 |
| Black | 0.076 | 0.178 | 1 | 0.138 |
| Asian | 0.275 | 0.166 | 0.153 | 1 |

Table 5: Interracial Attractions from Three Studies Using Add Health Data

*Quillian and Campbell (2003)*

| | White | Hispanic | Black | Asian |
|---|---|---|---|---|
| White | 1 | 0.540 | 0.210 | 0.698 |
| Hispanic | 0.380 | 1 | 0.395 | 0.583 |
| Black | 0.115 | 0.255 | 1 | 0.159 |
| Asian | 0.415 | 0.213 | 0.181 | 1 |

*Mouw and Entwisle (2006)*

| | White | Hispanic | Black | Asian |
|---|---|---|---|---|
| White | 1 | 0.645 | 0.299 | 0.510 |
| Hispanic | 0.611 | 1 | 0.803 | 0.456 |
| Black | 0.170 | 0.432 | 1 | 0.332 |
| Asian | 0.200 | 0.260 | 0.314 | 1 |

*Our study*

| | White | Hispanic | Black | Asian |
|---|---|---|---|---|
| White | 1 | 0.580 | 0.184 | 0.541 |
| Hispanic | 0.153 | 1 | 0.153 | 0.217 |
| Black | 0.072 | 0.138 | 1 | 0.151 |
| Asian | 0.242 | 0.165 | 0.120 | 1 |

*Note:*

a. Quillian and Campbell separated Hispanics into three groups—white Hispanics, black Hispanics, and other Hispanics. For the purpose of comparison, we combined estimates for the three Hispanic subgroups using weighted averages. The results presented here are based on estimates for adolescents with U.S.-born parents only in their Table 1 (p. 551).

b. Results for Mouw and Entwisle are based on their Model 4 in Table 6 (p. 416), which controls for sex, grade, parental education and income, residential distance, and school diversity, but not network variables.

c. Estimates from our own study are weighted averages of interracial attractions for the female sample and the male sample, presented earlier under Model H1 in Table 4.

## References


Allison, P. D. and N. A. Christakis. 1994. "Logit-Models for Sets of Ranked Items." Pp. 199-228, *Sociological Methodology*.

Anderson, C. J., et al. 1999. "A P* Primer: Logit Models for Social Networks." *Social Networks* 21:37-66.

Ben-Akiva, M. E. and S. R. Lerman. 1985. *Discrete Choice Analysis : Theory and Application to Travel Demand*. Cambridge, Mass.: MIT Press.

Blau, P. M. 1977a. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*: Collier Macmillan.

—. 1977b. "A Macrosociological Theory of Social Structure." *The American Journal of Sociology* 83:26-54.

Breen, R. 1994. "Individual-Level Models for Mobility Tables and Other Cross- Classifications." *Sociological Methods & Research* 23:147-173.

Diprete, T. A. 1990. "Adding Covariates to Loglinear Models for the Study of Social-Mobility." *American Sociological Review* 55:757-773.

Feld, S. L. 1982. "Social Structural Determinants of Similarity among Associates." *American Sociological Review* 47:797-801.

Hallinan, M. T. and R. A. Teixeira. 1987. "Opportunities and Constraints - Black-White Differences in the Formation of Interracial Friendships." *Child Development* 58:1358-1371.

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

Harris, D. R. and H. Ono. 2005. "How Many Interracial Marriages Would There Be If All Groups Were of Equal Size in All Places? A New Look at National Estimates of Interracial Marriage." *Social Science Research* 34:236-251.

Hoff, P. D. 2003. "Random Effects Models for Network Data." *National Academy of Sciences: Symposium on Social Network Analysis for National Security*.

—. 2005. "Bilinear Mixed-Effects Models for Dyadic Data." *Journal of the American Statistical Association* 100:286-295.

Kubitschek, W. N. and M. T. Hallinan. 1998. "Tracking and Students' Friendships." *Social Psychology Quarterly* 61:1-15.

Logan, J. A. 1983. "A Multivariate Model for Mobility Tables." *American Journal of Sociology* 89:324-349.

—. 1996a. "Opportunity and Choice in Socially Structured Labor Markets." *American Journal of Sociology* 102:114-160.

—. 1996b. "Rules of Access and Shifts in Demand: A Comparison of Log- Linear and Two-Sided Logit Models." *Social Science Research* 25:174-199.

—. 1998. "Estimating Two-Sided Logit Models." Pp. 139-173, *Sociological Methodology*.

Logan, J. A., et al. 2001. "A Parametric Two-Sided Model of Marriage." *manuscript, University of Wisconsin, Department of Sociology*.

Louviere, J. J., et al. 2000. *Stated Choice Methods*: Cambridge University Press New York.

Luce, R. D. 1959. *Individual Choice Behavior a Theoretical Analysis*. New York: Wiley.

Mayhew, B. H., et al. 1995. "Sex and Race Homogeneity in Naturally-Occurring Groups." *Social Forces* 74:15-52.

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

—. 1978. "Modelling the Choice of Residential Location." Pp. 75-96 in *Spatial Interaction Theory and Planning Models*, edited by McFadden, D. and A. Karlqvist. Amsterdam: North-Holland.

McPherson, J. M. and L. Smith-Lovin. 1987. "Homophily in Voluntary Organizations - Status Distance and the Composition of Face-to-Face Groups." *American Sociological Review* 52:370-379.
McPherson, M., et al. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415-444.
Moody, J. 2001. "Race, School Integration, and Friendship Segregation in America." *American Journal of Sociology* 107:679-716.
Mouw, T. and B. Entwisle. 2006. "Residential Segregation and Interracial Friendship in Schools."
Parsons, G. R. and M. J. Kealy. 1992. "Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation." *Land Economics* 68:93-106.
Pattison, P. and S. Wasserman. 1999. "Logit Models and Logistic Regressions for Social Networks: Ii. Multivariate Relations." *British Journal of Mathematical & Statistical Psychology* 52:169-193.
Powers, D. A. and Y. Xie. 2000. *Statistical Methods for Categorical Data Analysis*. San Diego: Academic.
Pudney, S. 1989. *Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes*: Basil Blackwell.
Quillian, L. and M. E. Campbell. 2001. "Class, Race, and School Racial Composition in the Formation of Interracial Friendships." in *annual meeting of the American Sociological Association*. Anaheim, CA.
—. 2003. "Beyond Black and White: The Present and Future of Multiracial Friendship Segregation." *American Sociological Review* 68:540.
Raudenbush, S. W. and A. S. Bryk. 2002. *Hierarchical Linear Models Applications and Data Analysis Methods*. Thousand Oaks: Sage Publications.
Schoen, R. 1986. "A Methodological Analysis of Intergroup Marriage." *Sociological Methodology* 16:49-78.
Siddarth, S., et al. 1995. "Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data." *Journal of Marketing Research* 32:255-266.
Verbrugge, L. M. 1977. "Structure of Adult Friendship Choices." *Social Forces* 56:576-597.
Wasserman, S. and P. Pattison. 1996. "Logit Models and Logistic Regressions for Social Networks .1. An Introduction to Markov Graphs and P." *Psychometrika* 61:401-425.
Xie, Y. and K. A. Shauman. 1997. "Modeling the Sex-Typing of Occupational Choice - Influences of Occupational Structure." *Sociological Methods & Research* 26:233-261.
Yamaguchi, K. 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.
Yellott, J. I. 1977. "The Relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution." *Journal of Mathematical Psychology* 15:109-144.