A Preference-Opportunity-Choice Framework with Applications to Intergroup Friendship[1]

Zhen Zeng
University of Wisconsin—Madison

Yu Xie
University of Michigan

A long-standing 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. This study puts forward a general framework for discrete choice, where choice probability is specified as proportional to the product of preference and opportunity. To implement this framework, the authors propose a modification to the conditional logit model for estimating preference parameters free from the influence of opportunity structure and then compare this approach to several alternative methods for separating preference and opportunity used in the friendship choice literature. As an empirical example, the authors 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 specification of the opportunity structure.

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 friendship choice is governed by the laws of homophily and propinquity. *Homophily* is a preference principle referring to the tendency to seek out and bond with others who are like ourselves. *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; thus, people have a greater opportunity to make friends with similar others (Feld 1982). 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 (Blau 1977; McPherson and Smith-Lovin 1987; McPherson, Smith-Lovin, and Cook 2001; Quillian and Campbell 2003).[2]

The separation of the effects of preference and opportunity on friendship choice has been a long-standing concern in friendship research. There are at least two reasons for this interest. First, given the significance attached to intergroup relations for social integration, it is important to know whether the high level of homogeneous association in friendship is due mainly to people's psychological predispositions or to the constraints of social structure. Second, the separation of preference and opportunity allows researchers to compare patterns of preference across social contexts and to predict choice behavior under a new set of conditions.

Past studies have attempted to separate preference and opportunity in a number of ways. Recognizing that the likelihood of having an out-group friend is directly influenced by relative group size, some studies have used dyads (pairs of people) as the units of analysis (e.g., Hallinan and Teixeira 1987; Moody 2001; Quillian and Campbell 2003; Mouw and Entwisle 2006). A key advantage of dyad analysis is that it avoids the confounding effect of group size on friendship choice. Another approach is to statistically control for opportunity using variables that capture interpersonal exposure. For example, to account for individual-level variation in opportunity, Moody (2001) controls for the number of shared school activities in predicting friendship ties for dyads; Mouw and Entwisle (2006) control for residential distance between the pair of potential friends. At an aggregate level, researchers have used log-linear models to analyze frequencies of friendship ties cross-classified by respondents' and their friends' characteristics, such as race or age grouping (Yamaguchi 1990).

---

[2] McPherson and Smith-Lovin (1987) used the terms *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*.

A well-known advantage of log-linear models is that they estimate intergroup associations net of group size effect. Although these disparate methods have been found useful in practice, researchers have not yet systematically compared their properties and implications.

In this article, we argue that any statistical model purporting to separate preference and opportunity in discrete choice must meet two basic criteria: (1) for an indifferent decision maker, it should yield a set of choice probabilities corresponding to the opportunity structure, and (2) for an equal opportunity structure (where all alternatives have equal opportunity to be chosen), it should yield a set of choice probabilities corresponding to preference. Based on these criteria, we propose a choice framework with choice probability specified as proportional to the product of preference and opportunity. To implement this framework, we incorporate opportunity into the conditional logit model as an offset variable. We then 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 exogenous opportunity structure. In this article, we focus on the application to interracial friendship choice, for two reasons. First, the separation of preference and opportunity in interpersonal associations has always been of great interest to sociologists. However, the solution requires data on the social context of choice, which are usually not available in surveys. A great opportunity to study friendship choice in well-measured social context presented itself with the advent of the National Longitudinal Study of Adolescent Health (Add Health) in the late 1990s. Since then, there has been a spurt of research on adolescent friendship choice (e.g., Joyner and Kao 2000; Moody 2001; Quillian and Campbell 2001, 2003; Mouw and Entwisle 2006; Doyle and Kao 2007). Second, interracial friendship choice has been analyzed at both the individual level, with the choice set consisting of individuals, and at an aggregate level, with the choice set consisting of racial groups. This application 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 realistic model of friendship formation should take into account reciprocity of relationships, influence of common associates (i.e., transitivity), 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 behavioral model of friendship choice, but rather to

propose a method for decomposing friendship choice into preference and opportunity, which in turn can be incorporated into more sophisticated models of friendship formation in future research.

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 this unconstrained choice because in a hypothetical choice situation like this, it can be safely assumed that the decision maker can pick any item as he pleases from the choice set; the exercise of his preference is free from external influences such as product availability on the market.

Constrained choice describes the situation where a choice decision is influenced not only by the decision maker's intrinsic preference but also by external factors such as availability, abundance, and accessibility of the items in the choice set. In this article, 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 that surveys people 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.[3] 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 according to their own unique circumstances). As a result, people do not always end up with 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

---

[3] In order to treat accessibility of 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.

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) that 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 decision maker's underlying preference. That is to say, the fact that A is chosen between A and 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 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* is a survey-based methodology, 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, Hensher, and Swait 2000). *Revealed preference* analysis (e.g., a transportation survey) 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 themselves—in order to correctly deduce preference from choice. The same choice problem can be researched using 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 preference and revealed preference. This article deals with how to deduce revealed preference from real-world choices.

The POC Framework

As stated above, we use the term *opportunity* to broadly refer to the influence of all factors other than preference on choice. Depending on the choice problem, opportunity may encompass different factors. As an example, let us consider what would be relevant external factors in inter-

group 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 1 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 formally analyze these data 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 total 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.

To better understand the relationships of 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.[4] Correspondingly, the racial com-

---

[4] In this reasoning, we make the 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 article.

TABLE 1
FEMALE STUDENTS' NOMINATIONS OF SAME-SEX FRIENDS, BY RACE

	BEST FRIEND'S RACE				No	TOTAL
RESPONDENT'S RACE	White	Hispanic	Black	Asian	NOMINATION	RESPONDENTS
White	73	24	34	9	41	181
Hispanic	17	55	9	2	23	106
Black	35	12	399	7	119	572
Asian	17	7	4	46	22	96
Total nominations ...	142	98	446	64	. . .	. . .

position of all nominated friends should approximate the racial composition of the school. This scenario of indifference leads to our first expectation for the functional relationship of preference, opportunity, and choice:

POC 1 (state of indifference).—*If the decision maker 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 decision maker'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. This scenario of equal opportunity leads to the second expectation:

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

Criteria POC 1 and POC 2 specify the boundary behavior expected of any reasonable model for constrained discrete choice. How should the model behave for the general case, when the decision maker is not indifferent and alternatives do not have equal opportunity to be chosen? We propose the following multiplicative assumption, consistent with POC 1 and POC 2:

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 decision maker and J denote the set of alternatives.[5] Let P_{ij} be the

---

[5] 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.

621

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 nonnegative real values.[6] For a given decision maker i, $(P_{i1}, \ldots, P_{in})$ is 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, where n is the size of J. From now on, we denote these three vectors as $\mathbf{P}_i$, $\mathbf{O}_i$, and $\mathbf{A}_i$. 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

$$P_{ij} = \frac{O_{ij} A_{ij}}{\sum\limits_{k \in J} O_{ik} A_{ik}}. \tag{2}$$

It is easy to see that expressions (1) and (2) satisfy the aforementioned model criteria, POC 1 and POC 2. When the decision maker is indifferent to all alternatives in the choice set (i.e., $A_{ij} = A_{ik}, j \neq k$), 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 $P_J(j) = v(j) / \sum_{k \in J} v(k)$, where function $v(j)$ represents a response strength associated with response j, and choice probability is proportional to response strength.[7] In standard discrete choice models, v is interpreted as a utility function. In our framework, however, v is decomposed 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 occurs. By *circumstances*, we mean not only the social environment in which the

[6] Note that our conception of opportunity as unequal accessibility of choice alternatives diverges from the notion of opportunity as the *composition* of choice sets in standard discrete choice analysis. There, a distinction is made between the universal set of alternatives and a particular choice set for an individual, consisting of "alternatives that are both feasible to the decision maker and known during the decision process" (Ben-Akiva and Lerman 1985, p. 33). Opportunity thus refers to 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 of particular alternatives.

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

decision maker is situated, but also the particular position he or 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 schedule 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 decision maker 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 a *preference function*. Substantively, we are interested not in particular individuals' preferences for concrete choice alternatives but in the pattern of association between decision makers' characteristics and the characteristics of alternatives. In addition, for estimation purposes, we need to constrain the dimension of the parameter space to be smaller than that of the data space. Since only one binary outcome is observed for each combination of decision maker 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 A_{ij} as a function of the characteristics of the decision maker and those of the alternatives. Let $\mathbf{z}_{ij}$ be a vector of characteristics pertaining to i and j and $\boldsymbol{\beta}$ be a vector of parameters. We specify the following preference function:

$$A_{ij} = \exp(\mathbf{z}'_{ij}\boldsymbol{\beta}).$$

Substituting for A_{ij} in (2) leads to the following expression for choice probability P_{ij} in the POC framework:

$$P_{ij} = \frac{O_{ij}\exp(\mathbf{z}'_{ij}\boldsymbol{\beta})}{\sum\limits_{k \in J} O_{ik}\exp(\mathbf{z}'_{ik}\boldsymbol{\beta})}. \tag{3}$$

The exponential function for preference ensures that choice probabilities are nonnegative. We call (3) 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 sit-

623

uation of equal opportunity, (3) reduces to the standard conditional logit model:

$$P_{ij} = \frac{\exp(\mathbf{z}'_{ij}\boldsymbol{\beta})}{\sum\limits_{k \in J} \exp(\mathbf{z}'_{ik}\boldsymbol{\beta})}.$$

CLO can be estimated using computer programs written for the 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.

Operationalization of Opportunity

The form of (3) 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 decision maker—is not directly observable. Therefore, it is often necessary to derive it from knowledge and assumptions about the choice context. 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 more time with their friends). Equal opportunity structure is used in many 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, most studies of interethnic marriage that use U.S. census data implicitly assume an equal opportunity marriage market at the national or metropolitan level (Harris and Ono 2005). Usually, the lack of an explicit assumption about the structure is practically equivalent to assuming a homogeneous environment and hence equal opportunity—if parameters of the standard choice model are interpreted as effects of preference.

The school environment affords the researcher a more refined operationalization of an opportunity structure of friendship choice. As men-

tioned 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 them 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. Later, we will demonstrate how to construct opportunity structure as a function of grade-level difference between potential friends.

In the literature on interpersonal associations, the terms *exposure* and *opportunity* are often used interchangeably. We treat them as separate concepts, with exposure specifically as contact with choice alternatives and opportunity as an abstract quantity capturing the total effects of all environmental constraints on choice. In other words, opportunity reflects not only the amount of contact with choice alternatives, but also the effectiveness of contact on choice probabilities. While opportunity is proportional to choice probability in our framework, we expect increasing exposure to increase choice probability with diminishing effect.

It is often the case that researchers have some knowledge about the structure of a social environment but do not know the exact functional form of opportunity structure. For example, researchers may collect data on grade levels, classes, and activities, but it is not clear how opportunity varies by these measured structural conditions. One approach is to parameterize O_{ij} as an exponential function of the social positions of decision maker i and potential friend j, just like preference. This leads to the following expression for choice probability:

$$P_{ij} = \frac{\exp\left(\mathbf{w}_{ij}'\boldsymbol{\delta} + \mathbf{z}_{ij}'\boldsymbol{\beta}\right)}{\sum_{k \in J} \exp\left(\mathbf{w}_{ik}'\boldsymbol{\delta} + \mathbf{z}_{ik}'\boldsymbol{\beta}\right)}, \tag{4}$$

which is exactly the form of the standard conditional logit model. Equation (4) suggests that as long as the two sets of variables, $\mathbf{z}_{ij}$ 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, respectively.

If a clean separation of preference and opportunity cannot be obtained because the same variables influence both preference and opportunity,

the researcher may 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 assumptions about opportunity structures 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's selecting a same-race peer as friend is c times that of selecting a black peer, if the two potential friends are otherwise identical on observed personal attributes and social positions. The ratio c is thus the total effect of race on friendship choice (for white decision makers), consisting of a portion due to racial homophily and a portion due to unobserved 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 unobserved opportunity of same-race friendship choice is d times that of cross-race friendship, 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 exact knowledge of opportunity structure is always unattainable. Depending on the problem under investigation, there may be a number of ways to estimate opportunity empirically. For example, in a supermarket, whether a product is placed near the checkout 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 is done in two steps. First, place two competing brands, A and B, near checkout 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, $O_{premium}/O_{regular}$, is estimated by the square root of the cross-ratio $P_A P_B*/ P_B P_A*$.[8] Estimates of opportunity structure from such experiments can then be imported into other studies to estimate

---

[8] This estimation requires the assumption that consumers' relative preferences for A and B are constant before and after the two brands switch shelf locations. 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, $A_A/A_B = (P_A/O_{premium})/(P_B/O_{regular}) = (P_A^*/O_{regular})/(P_B^*/O_{premium})$. Rearranging the terms gives $O_{premium}/O_{regular} = \sqrt{(P_A P_B^*)/(P_B P_A^*)}$.

parameters of preference. This "borrowing" of opportunity structure is made possible by the conceptual separation of preference as traits of the decision maker and opportunity as properties of the choice context in the POC framework.

To recapitulate, in order to infer preference from observed choices, the researcher must operationalize and quantify the opportunity structure. Methods are available for empirically estimating opportunity structures. Nevertheless, researchers are often faced with a 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. Finally, we would like to emphasize that statistical control—in the traditional sense of adding independent variables to the right side of the regression equation—should not be taken as an automatic solution to the problem of separating preference and opportunity. In most situations, the exact form of the opportunity structure must be known in order to infer preference, and the correct adjustment is to add $\ln(o_{ij})$ as an offset variable, not as a regular control variable. In sum, the separation of preference and opportunity is not free knowledge and can only be obtained by supplying necessary information about the opportunity structure.

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, the measurement and estimation of intergroup preference, choice-set aggregation, and the comparison of CLO to alternative methods for separating preference and opportunity in friendship choice.

Models for Three Types of Friendship Choice Data

The ideal data for studying friendship choice are collected in social settings such as schools or workplaces through roster-based nomination. In Add Health, for example, school rosters were distributed to each student in sampled schools and students were asked to name their friends from the rosters. Roster-based nomination enables us to model friendship choice

from a well-defined choice set, and in that, it is superior to the free nomination method, where respondents report the characteristics of their friends but the context of choice is usually unknown.

There are three ways in which researchers may 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 nominate 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 ordered or unordered multiple friends can be handled as extensions of the basic selection problem.

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 the best friend from the remainder of set J. This process is repeated until finally the last friend is selected. Let M_i be the number of rank-ordered friends i selects from choice set J. We model the probability of rank-ordered selections conditional on $M_i = m$. This approach not only greatly simplifies the models for ordered and unordered selections of multiple friends, but also allows us to focus on preference, as opposed to friendliness (i.e., the number of friends respondents nominate). As we will explain later in the article, in roster-based friendship nomination data, M_i depends largely on choice set size and survey instrument and thus is not an aspect of intrinsic sociopsychological predispositions that we are interested in.

Let $P_i(j_1, j_2, \ldots, j_m | J)$ denote the probability that decision maker i selects j_1 as the best friend, j_2 as the second-best friend, and so on, out of choice set J. The probability of best-friend selection is represented as $P_i(j_1 | J)$. This notation refers to the same quantity as the previously used P_{ij}, except that we now make the choice set explicit. In modeling ordered selections, we assume that past choice outcomes do not affect later choice decisions. We call this assumption *irrelevancy of past choices*. Hence, $P_i(j_1, j_2, \ldots, j_m | J)$ can be written as the product of m choice probabilities, each modeled by CLO:

$$P_i(j_1, j_2, \ldots, j_m|J) = P_i(j_1|J)P_i(j_2|J - \{j_1\}), \ldots, P_i(j_m|J - \{j_1, \ldots, j_{m-1}\})$$

$$= \frac{O_{ij_1}\exp(\mathbf{z}'_{ij_1}\boldsymbol{\beta})}{\sum\limits_{h \in J} O_{ih}\exp(\mathbf{z}'_{ih}\boldsymbol{\beta})} \times \frac{O_{ij_2}\exp(\mathbf{z}'_{ij_2}\boldsymbol{\beta})}{\sum\limits_{h \in J-\{j_1\}} O_{ih}\exp(\mathbf{z}'_{ih}\boldsymbol{\beta})}$$

$$\times, \ldots, \times \frac{O_{ij_m}\exp(\mathbf{z}'_{ij_m}\boldsymbol{\beta})}{\sum\limits_{h \in J-\{j_1, \ldots, j_{m-1}\}} O_{ih}\exp(\mathbf{z}'_{ih}\boldsymbol{\beta})}. \tag{5}$$

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, and so on.

Under the assumption of equal opportunity, (5) simplifies to

$$P_i(j_1, j_2, \ldots, j_m|J) = \frac{\exp(\mathbf{z}'_{ij_1}\boldsymbol{\beta})}{\sum\limits_{h \in J} \exp(\mathbf{z}'_{ih}\boldsymbol{\beta})} \times \frac{\exp(\mathbf{z}'_{ij_2}\boldsymbol{\beta})}{\sum\limits_{h \in J-\{j_1\}} \exp(\mathbf{z}'_{ih}\boldsymbol{\beta})}$$

$$\times, \ldots, \times \frac{\exp(\mathbf{z}'_{ij_m}\boldsymbol{\beta})}{\sum\limits_{h \in J-\{j_1, \ldots, j_{m-1}\}} \exp(\mathbf{z}'_{ih}\boldsymbol{\beta})},$$

which is of the form of the *rank-ordered* or *exploded* logit model (for a detailed discussion of this model, see Allison and Christakis [1994]). In terms of estimation property, 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 and alternatives that are not selected may be regarded as censored cases. To estimate (5), 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 allow us to explore a wider range of research questions than best-friend data. We may investigate whether preferences vary by rank order or by the number of nominations respondents make. For example, using log-linear analysis, Yamaguchi (1990) found that 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 being generated by the same selection process as

ordered selections, except with missing rank order information. We use $P_i(\{j_1, \ldots, j_m\}|J)$ to denote the probability of i selecting $j_1, j_2, \ldots, j_m$ as an unordered set of friends from choice set J, distinct from the notation $P_i(j_1, \ldots, j_m|J)$ for ordered selections. As in ordered selections, we model this probability conditional on the number of nominations, m; $P_i(\{j_1, \ldots, j_m\}|J)$ 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; $P_i(\{a, b\}|J)$ may be expressed as the sum of two rank-ordered choice probabilities: $P_i(\{a, b\}|J) = P_i(a, b|J) + P_i(b, a|J)$. That is, either a is the best friend and b is the second-best friend, or vice versa. 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

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

(6)

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 reexpression of (5) for the probability of selecting an ordered set of m friends.[9] Equation (6) can be estimated as a special case of the conditional logit model, known as the conditional logit model with multiple positive outcomes, with $\ln(O_{ij})$ included as an offset variable.

Measuring and Estimating Intergroup Preference

As mentioned earlier, a major goal of friendship research is to infer preference patterns—who is attracted to whom—from choice behavior. In particular, sociologists are interested in intergroup (e.g., interracial, interfaith, and interclass) relations as reflected in friendship choice. We propose to measure intergroup attraction by the relative probability of choosing an out-group friend over an in-group friend under equal opportunity.

[9] The second summation in the denominator in (6) 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 elements in $\{j_1, \ldots, j_m\}$ as r increases from 1 to m.

Specifically, we denote the level of attraction that group w holds for group v by $\mathbf{A}_{vw}$ and $\mathbf{A}_{vw} \equiv (P_{ij}/O_{ij})/(P_{ik}/O_{ik})$, $i, k \in v, j \in w$. The symbol $\mathbf{A}$ without subscript is a square matrix, which we call the *intergroup attraction matrix*, with diagonals indicating in-group preferences and off-diagonals indicating intergroup preferences. Note that all in-group preferences are 1 by definition.

Defined as a ratio of choice probabilities with adjustment for opportunity, $\mathbf{A}_{vw}$ has two advantages as a measure of intergroup attraction. First, it is invariant to opportunity structure—which, in this case, is affected by relative group size and relative availability of alternatives in v and w. Second, the scaling by the inverse of in-group preference renders $\mathbf{A}_{vw}$ invariant to friendliness (i.e., the number of people nominated as friends). Friendliness is to a large extent determined by choice set size and survey instrument.[10] In general, while the probability of selection P_{ij} is approximately proportional to the number of nominations solicited and inversely proportional to choice set size, both the relative risk P_{ij}/P_{ik} and the adjusted ratio $(P_{ij}/O_{ij})/(P_{ik}/O_{ik})$ are insensitive to nomination size and choice set size. These invariance properties make $\mathbf{A}_{vw}$ a useful measure of intergroup attraction for comparison across choice contexts.

The perceptive reader may notice that $\mathbf{A}_{vw}$ is not a relative risk ratio (RRR), the usual measure of intergroup association used in the standard conditional logit model and in two other closely related models—the multinomial logit and log-linear models. A relative risk ratio is the cross-ratio of four choice probabilities or—if we are dealing with aggregate data—frequencies of ties. For example, let F_{vw} be the number of people in group w nominated by respondents in group v as friends; the RRR $F_{WB}F_{WA}/F_{HB}F_{HA}$ is a measure of intergroup preference, interpreted as the extent to which whites are more likely to select blacks over Asians as friends, compared to Hispanic decision makers. The RRR is invariant to row and column marginals, that is, the numbers of respondents and nominees of each group. In other words, the RRR is purged of group sizes, overall friendliness, and overall popularity of the groups. Unlike RRR, $\mathbf{A}_{vw}$ is purged of opportunity and friendliness but remains sensitive to overall popularity. Simply put, the difference is that $\mathbf{A}_{vw}$ measures group v's preference for group w, but an RRR measures group v's preference for group w relative to another group's preference. We prefer $\mathbf{A}_{vw}$ to RRR as a measure of intergroup preference because preference, rather than relative preference, is part and parcel of the interest in intergroup relations research.

Now consider the case of interracial friendship choice where students

---

[10] For example, researchers may ask respondents to make single nominations or unlimited nominations of friends, from a school roster or a class roster.

TABLE 2
INTERGROUP ATTRACTION MATRIX ESTIMATED FROM
INDIVIDUAL-LEVEL DATA

RESPONDENT'S	BEST FRIEND'S RACE			
RACE	White	Hispanic	Black	Asian
White	1	.558	.147	.231
Hispanic	.179	1	.030	.040
Black	.277	.162	1	.104
Asian	.194	.136	.014	1

fall into four groups: white, black, Hispanic, and Asian (W, B, H, and A). To estimate $\mathbf{A}$ using CLO (or its variants for multiple selections), we include a vector of 16 indicators, $\mathbf{x}_{ij} = (x_{ij}^{WW}, x_{ij}^{WB}, x_{ij}^{WH}, \ldots, x_{ij}^{AA})'$, as predictors to represent the race of decision maker i and that of potential friend j. For example, if i is black and j is white, then $x_{ij}^{BW} = 1$, and all 15 other indicators take on the value of 0. Let $\boldsymbol{\beta} = (\beta^{WW}, \beta^{WB}, \beta^{WH}, \ldots, \beta^{AA})'$ be the parameters for $\mathbf{x}$ 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. Additional independent variables may be included to simultaneously estimate preference parameters for other factors.

Table 2 displays the intergroup attraction matrix estimated from individual-level data, which were summarized in table 1. We used the simplest opportunity structure and assumed that all students are equally likely to be friends a priori. The estimated intergroup attraction parameters are easily interpretable. For example, $\mathbf{A}_{WH} = 0.558$ means that a white student's relative likelihood of choosing a Hispanic friend versus a white friend is 0.558 under equal opportunity. All intergroup attractions in table 2 are substantially smaller than 1, indicating a strong in-group bias.

Choice-Set Aggregation in Friendship Choice

As Ben-Akiva and Lerman (1985, pp. 31–32) point 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 analyze friendship choice at a group level, in which case the choice set consists not of individuals, but of groups (e.g., racial groups, age groups, and religious groups).[11] In our interracial friendship example, the individual-level approach presents the choice of a friend out of 955 students, while

[11] In this context, the group-level approach refers to aggregation of alternatives in the choice set, not aggregation of decision makers.

the group-level approach presents the "choice" of a racial group out of {white, black, Hispanic, Asian}. Will estimates of intergroup preference be different depending on the level of analysis? It turns out that a group-level CLO can be written as an aggregate form of the individual-level CLO. As long as model specifications are equivalent—we will explain these conditions shortly—the same estimates of intergroup attractions can be obtained using either approach. We see this unification of the individual-level and group-level modeling approaches as a major advantage of CLO and the POC framework. In contrast, if we were to apply the standard conditional logit model and the multinomial logit model to the same problem at the individual level and the group level, respectively, we would usually arrive at different estimates of intergroup attractions.

For two models to estimate the same parameters, they must utilize the same information. Because a group-level model cannot incorporate predictors that vary *within* groups of potential friends, the individual-level model must be similarly constrained. Furthermore, if multiple friends are selected, the group-level model will obviously be less informative than the individual-level model. Therefore, we will consider models for best-friend selection only.

Let P_{iw} be the probability that i chooses someone from group w as a friend, and let J' be the group-level choice set. Given that each respondent makes only one selection, P_{iw} can be written as the sum of choice probabilities for all alternatives in w: $P_{iw} = \sum_{j \in w} P_{ij}$. Furthermore, since $\mathbf{x}_{ij}$ does not vary within groups of potential friends, we may denote all $\mathbf{x}_{ij}, j \in w$, by $\mathbf{x}_{iw}$ without loss of information. Hence,

$$P_{iw} = \sum_{j \in w} P_{ij} = \frac{\sum\limits_{j \in w} O_{ij} A_{ij}}{\sum\limits_{v \in J'} \sum\limits_{k \in v} O_{ik} A_{ik}} = \frac{\sum\limits_{j \in w} O_{ij} \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})}{\sum\limits_{v \in J'} \sum\limits_{k \in v} O_{ik} \exp(\mathbf{x}'_{ik}\boldsymbol{\beta})}$$

$$= \frac{\sum\limits_{j \in w} O_{ij} \exp(\mathbf{x}'_{iw}\boldsymbol{\beta})}{\sum\limits_{v \in J'} \sum\limits_{k \in v} O_{ik} \exp(\mathbf{x}'_{iv}\boldsymbol{\beta})}.$$

Factoring out $\exp(\mathbf{x}'_{iw}\boldsymbol{\beta})$, we have the following:

$$P_{iw} = \frac{(\sum\limits_{j \in w} O_{ij}) \exp(\mathbf{x}'_{iw}\boldsymbol{\beta})}{\sum\limits_{v \in J'} [(\sum\limits_{k \in v} O_{ik}) \exp(\mathbf{x}'_{iv}\boldsymbol{\beta})]} = \frac{O_{iw} \exp(\mathbf{x}'_{iw}\boldsymbol{\beta})}{\sum\limits_{v \in J'} O_{iv} \exp(\mathbf{x}'_{iv}\boldsymbol{\beta})}, \tag{7}$$

where $O_{iw} = \sum_{j \in w} O_{ij}$.

Equation (7) reveals the connections between the group-level model and the individual-level model: (1) preference for group w, $\exp(\mathbf{x}'_{iw}\boldsymbol{\beta})$, has the same parameter vector as in the individual-level model; (2) opportunity to choose group w in the group-level model equals the sum of opportunities

to choose all alternatives in group w in the individual-level model; that is, $O_{iw} = \sum_{j \in w} O_{ij}$. Therefore, if the group-level model is estimated with the offset $\ln \sum_{j \in w} O_{ij}$, it should yield the same estimates as the individual-level model. This result holds under reasonable conditions, namely, that predictors are constant within groups of potential friends and each respondent chooses only one friend.

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 from group w. This reexpression decomposes group-level opportunity to two factors: group size and average dyad-level opportunity. If there is no systematic variation in average dyad-level opportunity across groups—that is, if $\bar{O}_{iw} = \bar{O}_{iv}$ for all $w, v \in J', w \neq v$—group-level opportunity O_{iw} reduces to group size n_w.[12]

Although individual-level models are generally superior to group-level models, in some situations it is more advantageous to model friendship choice at the group level. For example, one problem with individual-level models is that they often involve very large choice sets and therefore are computing intensive. Unless one wants to study variations within the grouping of choice alternatives, the group-level model can be used to estimate intergroup attractions. Another situation is where surveys collect data by asking respondents to name the social groups their best friends belong to, instead of using the roster method. This type of data is not amenable to modeling choice 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 as opportunity can be used in place of individual-level models.

Comparison of Models for Friendship Choice

Studies on intergroup friendship choice have applied a variety of methods, ranging from log-linear analysis for aggregate data (Yamaguchi 1990) to the conditional logit model (Quillian and Campbell 2001), the multinomial logit model (Doyle and Kao 2007), the ordinary logit model (Hallinan and Williams 1989; Moody 2001), and the p^* model (Mouw and Entwisle 2006) for individual-level data. The p^* model takes the form of an ordinary logit model, characterized by the use of network characteristics such as mutuality, transitivity, cyclicity, and others as predictors of friendships. We now turn to a brief comparison of these various methods. Given our

---

[12] Since respondents cannot nominate themselves as friends (i.e., $O_{ij} = 0$), O_{iw} reduces to $n_w - 1$ for $i \in w$.

focus on choice framework, we will first distinguish between two approaches—one that views friendship nomination data as generated by choice behavior and one that does not—and then compare how the separation of opportunity and preference is handled in various forms of discrete choice models.

The conditional logit model and the multinomial logit model are also known as McFadden's choice model. They can be derived from a behavioral model of discrete choice, where the decision maker follows a set of decision rules such as utility maximization or independence of irrelevant alternatives in making choices (Luce 1959; McFadden 1974; Ben-Akiva and Lerman 1985; Pudney 1989).[13] Although the log-linear model is usually not regarded as a choice model, previous studies have shown that it is closely related to the multinomial logit model and can be expressed in the form of the latter (Logan 1983; DiPrete 1990; Breen 1994). Statistically, the log-linear model belongs to the same family of models as the conditional logit model and the multinomial logit model but is used for aggregate data.

The main feature of the choice model vis-à-vis the ordinary logit model is that choice probability is specified as dependent on *all* alternatives in i's choice set: $P_{ij} = \exp(\mathbf{z}'_{ij}\boldsymbol{\beta})/\sum_{k \in J} \exp(\mathbf{z}'_{ik}\boldsymbol{\beta})$. The use of the scaling factor $1/\sum_{k \in J} \exp(\mathbf{z}'_{ik}\boldsymbol{\beta})$ ensures that the choice probabilities sum up to 1 for each decision maker. In contrast, probabilities in the ordinary (unconditional) logit model, $P_{ij} = \exp(\mathbf{z}'_{ij}\boldsymbol{\beta})/[1 + \exp(\mathbf{z}'_{ij}\boldsymbol{\beta})]$, are functions of the characteristics of i and j only and do not sum up to 1 for each decision maker. Thus, the ordinary logit model implies that people make independent decisions about each alternative without comparing it to other alternatives. Whether the conditional or the unconditional logit model should be used depends on the method of data collection. If the survey instrument sets a cap on the number of nominations, especially if only best friends are solicited, the conditional logit model is more appropriate. If, on the other hand, the instrument instructs respondents to nominate as many friends as they have or if the absolute level of selection probability is of interest, the unconditional approach is more appropriate.

In previous sociological studies on interpersonal associations, opportunity is conceptualized as the pattern of intergroup associations predicted under the assumption of random mixing, and preference is conceptualized as deviation from that pattern (see Verbrugge 1977; Mayhew et al. 1995; McPherson et al. 2001; Mouw and Entwisle 2006). Operationally, the key

---

[13] There are two fundamentally different decision rules in choice models: the probabilistic choice 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).

to separating preference from opportunity in the standard approach is the relative risk ratio. As explained earlier, the coefficients in the conditional logit model are interpreted as relative risk ratios, which measure the association between the traits of decision makers and those of the alternatives, and are conveniently invariant to decision maker's friendliness and alternatives' popularity. In log-linear analysis of grouped data, for example, opportunity structure is viewed as represented by the row and column marginals. It is customary for log-linear models to include ancillary parameters to account for the marginals so that the row-column association parameters measure *net* associations.

In the POC framework, opportunity does not equate to the marginals, as in log-linear models, or to alternatives' overall popularity, as in the standard conditional logit model. Instead, we define opportunity as choice probabilities of the indifferent chooser and leave it to the researcher to specify the state of indifference. To illustrate this difference using data in table 1, in the log-linear model, opportunity structure is represented by the predicted frequencies under the independence model with row marginals {140, 83, 453, 74} and column marginals {142, 98, 446, 64}; in the POC framework, under the assumption of a homogeneous environment, the opportunity structure for intergroup friendship choice can be represented by the number of dyads formed between {140, 83, 453, 74} decision makers and {181, 106, 572, 96} potential friends of various racial groups. The significant difference in this case is that students who were not nominated as friends contribute to the opportunity in our approach, but not in the conventional approach.

Using the CLO, the researcher specifies what she regards as the state of indifference in the offset term $\ln(O_{ij})$, and consequently the coefficients of a CLO may be interpreted as the effects of preference on choice. It may be worthwhile to think about how adding an offset term may affect the coefficients of a conditional logit model. After all, the attraction of the conditional logit model is that it yields estimates of chooser-alternative associations that are invariant to alternatives' overall likelihoods of selection. Indeed, in the special case that all respondents face the same opportunity structure—that is, if O_{ij} does not vary by i—the offset of opportunity affects only the intercepts, but not the coefficients, of the conditional logit model. In other words, in this special case, we will arrive at the same inferences about relative risk ratios, but the predicted relative risks (i.e., intergroup attractions as defined in **A**) will be different. In general, when decision makers face different opportunity structures, both the coefficients and the intercepts of a conditional logit model are sensitive to the offset of opportunity.

There are a few issues specific to friendship choice we do not attempt to resolve because they are beyond the central topic of this article. One

such issue is that individuals in a social network do not make decisions about friendship choices independent of one another, but rather their decisions are affected by network dynamics such as reciprocity (the tendency to reciprocate friendship nominations) and transitivity (the tendency to nominate a friend's friend as a friend), as well as others. The p^* model and the random-effects model offer promising solutions to the problem of interdependency. The p^* logit regression models directed or undirected friendship ties explicitly as functions of network characteristics. The random-effects model can be applied to dyadic analysis to account for correlations due to common actors (i.e., correlations among $P_i.$) and shared targets (i.e., correlations among $P._j$) via error covariance structures. Readers are referred to those two bodies of literature for more information (see Wasserman and Pattison [1996], Anderson, Wasserman, and Crouch [1999], and Pattison and Wasserman [1999] for p^* models; see Raudenbush and Bryk [2002] and Hoff [2003, Hoff 2005] for random-effects models).

AN EMPIRICAL EXAMPLE

Data

To illustrate the various methods discussed so far, we analyze data from Add Health, a school-based study of adolescents enrolled in grades 7–12 during 1994–95. 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 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 relationships are likely subject to different selection criteria 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, only 1% of which are friends. We took a random sample of nonfriend dyads while

[14] The roster lists all the students in the focal school (i.e., the respondent's 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 in this article limits the choice set to the focal school.

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 background on friendship choice. The first hypothesis is *homophily*: similarity in personal characteristics enhances the likelihood of friendship. The second hypothesis, which has not yet been formally tested in this literature, is *status asymmetry*: when distance in status is equal, there is a greater tendency to nominate a person with a higher status (in terms of age, GPA, or SES) than 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.

Related to status asymmetry is the issue of whether friendship choice should be treated as a one-sided or two-sided choice. A typical two-sided choice is marriage matching, where the relationship is symmetric by definition: i is married to j if and only if j 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 were a two-sided relationship by nature, then the 60% unreciprocated nominations would 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 relationship.

To test the homophily hypothesis, we include variables indicating the 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 the mutually exclusive categories of non-Hispanic white, Hispanic, non-Hispanic black, and non-Hispanic Asian. GPA is calculated as the average grade over 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.

---

[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.

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; for example, $\beta_1(1 - \delta_{ij})|\text{AgeDiff}_{ij}| + \beta_2\delta_{ij}|\text{AgeDiff}_{ij}|$, where $\delta_{ij} = 1$ if $\text{Age}_j > \text{Age}_i$, and $\delta_{ij} = 0$ 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; that is, $\beta_1 < \beta_2 < 0$.

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 nonfriends, along with those who were not nominated. In modeling unordered selections, we simply ignore the rank-order information on the multiple friends selected.

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 homogeneous environment, which leads to the *equal opportunity structure* (or E, for short) and hence to 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) specifies that opportunity is a continuous decreasing function of the distance in grade levels between a pair of students. Specifically, we assume $O_{ij} = 1/(d_{ij} + 1)^2$, where d_{ij} is the difference in grade levels between i and j. The *dichotomous opportunity structure* (or D) specifies that opportunity is greater between students in the same grade than between those in different grades by a factor of 6. That is, we assume the following opportunity structure:

$$O_{ij} = \begin{Bmatrix} 6 & \text{if } d_{ij} = 0 \\ 1 & \text{if } d_{ij} \neq 0 \end{Bmatrix}.$$

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

   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 twelfth-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 of 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 three types of nominations, three opportunity structures, and two genders. Within each gender, we label the models as E1, E2, E3, G1, G2, G3, and 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. Estimates in table 3 are relative risk ratios, 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 decision maker 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 one year greater than that between i and k, everything else being equal. The RRR of 0.339 applies only to alter < ego dyads; the value increases to 0.347 for alter > ego dyads. All coefficients in table 3 are statistically significant. Asterisks indicate whether the differences in the coefficients for alter > ego and alter < ego dyads are statistically significant.

   In table 3, we observe a strong pattern of homophily with respect to age and GPA. For female dyads, an age difference of one year reduces the relative risk of selection by approximately two-thirds in the equal opportunity models and by about 40% in the gradient and dichotomous opportunity models. One unit of GPA difference is associated with an RRR between 0.61 and 0.65 across the board for girls. With a coefficient of about 0.95, homophily based on SES appears much weaker, but it is

TABLE 3
Effects of Age, GPA, and Family Socioeconomic Background on Friendship Choice

	Best-Friend Selection			Ordered Nominations			Unordered Nominations		
	H1	G1	D1	H2	G2	D2	H3	G3	D3
Female-female nominations:									
Age difference:									
Alter < ego	.339	.626	.585	.342	.637	.594	.338	.629	.587
Alter > ego	.347	.660*	.621*	.351	.67*	.628*	.347	.662*	.622*
GPA difference:									
Alter < ego	.605	.623	.626	.609	.626	.629	.606	.623	.626
Alter > ego	.609	.625	.625	.628	.646	.647	.622	.639	.640
SES difference:									
Alter < ego	.928	.93	.931	.925	.927	.928	.923	.926	.926
Alter > ego	.962*	.963*	.965*	.966*	.966*	.967*	.964*	.966*	.966*
No. of observations	472,215			480,215			473,095		
Male-male nominations:									
Age difference:									
Alter < ego	.390	.674	.622	.388	.668	.619	.384	.662	.612
Alter > ego	.424*	.747*	.690*	.425*	.746*	.690*	.422*	.741*	.684*
GPA difference:									
Alter < ego	.648	.664	.667	.642	.658	.661	.639	.653	.656
Alter > ego	.742*	.757*	.756*	.750*	.767*	.765*	.747*	.761*	.760*
SES difference:									
Alter < ego	.952	.956	.957	.954	.959	.96	.953	.957	.958
Alter > ego	.973*	.977*	.978*	.981*	.985*	.986*	.980*	.984*	.985*
No. of observations	388,373			398,571			389,502		

Note.—Estimates are presented in exponential form. All coefficients are significantly different from 1 at the .001 level.
* Asymmetry is statistically significant at .001—ie., the coefficient for alter > ego dyads is significantly different from the alter < ego dyads.

actually much closer to homophily in GPA, as the standardized coefficients for GPA difference and SES difference are about 0.8 and 0.9, respectively. The estimates for boys are slightly higher than those for girls, indicating a somewhat lower-level homophily in male-male friendship.

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 examined in this analysis and the result is robust across models. 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. If that were true, we would not have found a ubiquitous pattern of status asymmetry.

We now turn to comparisons across opportunity structures and types of nomination. The estimated coefficients do not vary substantially by the type of nomination. 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 around 0.3–0.4 to around 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, because GPA and SES differences are uncorrelated with $\ln(O_{ij})$ in this data set. For 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 specification E1 (i.e., best-friend nomination and the equal opportunity structure). Results estimated from the other models are similar and are therefore not presented here. As with GPA and SES, estimates of interracial attractions are insensitive to opportunity structure based on grade levels, because of the orthogonality of these variables with

TABLE 4
ESTIMATED INTERRACIAL ATTRACTIONS IN BEST-FRIEND SELECTION

	FEMALE-FEMALE NOMINATIONS				MALE-MALE NOMINATIONS			
	White	Hispanic	Black	Asian	White	Hispanic	Black	Asian
Baseline model:								
White	1	.516	.153	.454	1	.513	.191	.570
Hispanic	.092	1	.092	.153	.213	1	.213	.267
Black	.066	.090	1	.141	.074	.159	1	.131
Asian	.198	.126	.079	1	.247	.136	.129	1
Multivariate model:								
White	1	.587	.164	.490	1	.573	.205	.598
Hispanic	.095	1	.095	.165	.217	1	.217	.274
Black	.069	.101	1	.163	.076	.178	1	.138
Asian	.212	.165	.090	1	.275	.166	.153	1

the assumed opportunity structure. The baseline model in table 4 contains race variables only, while the multivariate model 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 respondent'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 around 0.15–0.2. Interracial attractions involving minority groups as decision makers fall between 0.07 and 0.28, with the lowest observed attraction being between blacks and whites. In addition, a comparison of the baseline models to the multivariate models indicates that estimates of interracial attractions are only slightly modified when age, GPA, and family SES are controlled. Thus, racial differences in GPA and SES contribute very little to the racial cleavage in friendship choice.

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 studies of adolescent peer groups that female cliques tend to be smaller, closer, and more homogeneous than male cliques (McPherson et al. 2001).

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).[16] Despite differences in research design, all 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, Mouw and Entwisle used the in-home subsample, while Quillian and Campbell and we used the in-school full sample. 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, and siblings. The oversamples of Puerto Ricans and middle-class black students in Mouw and Entwisle's analysis may have contributed to a higher Hispanic-black attraction than would be expected for the general Hispanic and black population. The research design of Quillian and Campbell differs from the other two studies in that those authors divided potential friends— but not respondents—by generational status with respect to immigration and estimated intergroup attractions for respondent's race × (potential friend's race by generation) combinations.[17] The results presented in table 5 pertain to the third generation only (i.e., those with U.S.-born parents). Their intergroup attraction estimates involving first- and second-generation immigrants are substantially lower for white Hispanic, other Hispanic, and Asian decision makers, but not necessarily for the other groups

[16] The results in table 5 are based on table 1 of Quillian and Campbell (2003, p. 551) and table 6 of Mouw and Entwisle (2006, p. 416). In their analysis, Quillian and Campbell 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 the Hispanic subgroups in the sample, which are 4%, 2% and 9%, respectively. Mouw and Entwisle 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 Quillian and Campbell and we did not use network variables as predictors, to facilitate comparison, we calculated interracial attractions for Mouw and Entwisle's study based on their results in model 4, which included an extensive list of controls, but not network variables. Estimates from our own study are the weighted averages of interracial attractions for the female sample and the male sample, presented earlier under the multivariate model in table 4.

[17] Their purpose for separating target groups by generational status was to test the hypothesis that racial homophily weakens across generations.

TABLE 5
Interracial Attractions from Three Studies Using Add Health Data

	White	Hispanic	Black	Asian
Quillian and Campbell (2003):				
White	1	.540	.210	.698
Hispanic	.380	1	.395	.583
Black	.115	.255	1	.159
Asian	.415	.213	.181	1
Mouw and Entwisle (2006):				
White	1	.645	.299	.510
Hispanic	.611	1	.803	.456
Black	.170	.432	1	.332
Asian	.200	.260	.314	1
This study:				
White	1	.580	.184	.541
Hispanic	.153	1	.153	.217
Black	.072	.138	1	.151
Asian	.242	.165	.120	1

Note.—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 (Quillian and Campbell 2003, p. 551). Results for Mouw and Entwisle are based on model 4 in their table 6 (2006, p. 416), which controls for sex, grade, parental education and income, residential distance, and school diversity, 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 the multivariate model in table 4.

(not presented here). If Quillian and Campbell had collapsed the results over generational status, our results would be more similar.

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-environment interactions—that is, in the causal effects of preference and opportunity on one another. On the one hand, people seek out the social groups they prefer. On the other hand, social environment can alter sociopsychological predispositions. Leaving these endogenous processes to future research, this article 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 relationships between preference and opportunity must begin with the inference of preference and the assumptions for that inference.

The confounding of preference and opportunity should not be dealt with as though they were just another pair of variables in a regression. The conventional wisdom is that if two independent variables both exert an influence on variable Y, a multiple regression can be used to estimate their independent effects; the regression coefficient of an independent variable is then interpreted as its effect on Y, holding the other independent variable constant. However, the separation of preference and opportunity cannot be achieved in this fashion, because a variable may serve as an agent for both preference and opportunity. For example, a race effect in friendship choice may indicate in-group preference, social segregation along the racial line, or both. In such cases, it is impossible to apportion the estimated coefficients between preference and opportunity without additional assumptions. In a sense, the separation of preference and opportunity is not so much an issue of statistical estimation as a matter of interpretation.

Our approach is to view opportunity as comprising the characteristics of a choice context and preference as representing the underlying dispositions of a person. In the POC framework, we specify opportunity as the choice probabilities of an indifferent chooser and preference as the deviation of observed choice patterns from those expected under indifference. Thus, the problem of apportioning coefficients between preference and opportunity is converted into the problem of specifying the opportunity structure, which admits a very natural interpretation as the expected behavior of an indifferent chooser or a random mixing process. Hence, the empirical separation of opportunity and preference requires that researchers have explicit knowledge (or make explicit assumptions) about the opportunity structure.

In conclusion, we offer the following recommendations. For analysis of existing data sets, researchers should make explicit their assumptions about opportunity when drawing inferences about 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.

REFERENCES

Allison, Paul D., and Nicholas A. Christakis. 1994. "Logit Models for Sets of Ranked Items." *Sociological Methodology* 24:199–228.
Anderson, Carolyn J., Stanley Wasserman, and Bradley Crouch. 1999. "A p^* Primer: Logit Models for Social Networks." *Social Networks* 21:37–66.

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

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

Breen, Richard. 1994. "Individual-Level Models for Mobility Tables and Other Cross-classifications." *Sociological Methods and Research* 23:147–73.

DiPrete, Thomas A. 1990. "Adding Covariates to Loglinear Models for the Study of Social Mobility." *American Sociological Review* 55:757–73.

Doyle, Jamie Mihoko, and Grace Kao. 2007. "Friendship Choices of Multiracial Adolescents: Racial Homophily, Blending, or Amalgamation?" *Social Science Research* 36:633–53.

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

Hallinan, Maureen T., and Ruy A. Teixeira. 1987. "Opportunities and Constraints: Black-White Differences in the Formation of Interracial Friendships." *Child Development* 58:1358–71.

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

Harris, David R., and Hiromi 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* 34:236–51.

Hoff, Peter D. 2003. "Random Effects Models for Network Data." Pp. 303–12 in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, edited by Ronald Breiger, Kathleen Carley, and Philippa Pattison. Washington, D.C.: National Academies Press.

———. 2005. "Bilinear Mixed-Effects Models for Dyadic Data." *Journal of the American Statistical Association* 100:286–95.

Joyner, Kara, and Grace Kao. 2000. "School Racial Composition and Adolescent Racial Homophily." *Social Science Quarterly* 81:810–25.

Kubitschek, Warren N., and Maureen T. Hallinan. 1998. "Tracking and Students' Friendships." *Social Psychology Quarterly* 61:1–15.

Logan, John A. 1983. "A Multivariate Model for Mobility Tables." *American Journal of Sociology* 89:324–49.

Louviere, Jordan J., David A. Hensher, and Joffre D. Swait. 2000. *Stated Choice Methods*. New York: Cambridge University Press.

Luce, R. Duncan. 1959. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Mayhew, Bruce H., J. Miller McPherson, Thomas Rotolo, and Lynn Smith-Lovin. 1995. "Sex and Race Homogeneity in Naturally Occurring Groups." *Social Forces* 74:15–52.

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

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

McPherson, J. Miller, and Lynn Smith-Lovin. 1987. "Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups." *American Sociological Review* 52:370–79.

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

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

Mouw, Ted, and Barbara Entwisle. 2006. "Residential Segregation and Interracial Friendship in Schools." *American Journal of Sociology* 112 (2): 394–441.

Parsons, George R., and Mary Jo Kealy. 1992. "Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation." *Land Economics* 68:93–106.

Pattison, Philippa, and Stanley Wasserman. 1999. "Logit Models and Logistic Regressions for Social Networks: II. Multivariate Relations." *British Journal of Mathematical and Statistical Psychology* 52:169–93.

Pudney, Stephen. 1989. *Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes*. Oxford: Basil Blackwell.

Quillian, Lincoln, and Mary E. Campbell. 2001. "Class, Race, and School Racial Composition in the Formation of Interracial Friendships." Paper presented at the annual meeting of the American Sociological Association. Anaheim, Calif., August.

———. 2003. "Beyond Black and White: The Present and Future of Multiracial Friendship Segregation." *American Sociological Review* 68:540–66.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, Calif.: Sage Publications.

Verbrugge, Lois M. 1977. "Structure of Adult Friendship Choices." *Social Forces* 56: 576–97.

Wasserman, Stanley, and Philippa Pattison. 1996. "Logit Models and Logistic Regressions for Social Networks: 1. An Introduction to Markov Graphs and P." *Psychometrika* 61:401–25.

Yamaguchi, Kazuo. 1990. "Homophily and Social Distance in the Choice of Multiple Friends: An Analysis Based on Conditionally Symmetric Log-Bilinear Association Models." *Journal of the American Statistical Association* 85:356–66.

Yellott, John I. 1977. "The Relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution." *Journal of Mathematical Psychology* 15:109–44.