

HOMOPHILY: MEASURES AND MEANING*

Barbara S. Lawrence
Anderson Graduate School of Management
University of California Los Angeles

Neha Parikh Shah
Rutgers Business School-Newark & New Brunswick
Rutgers University

August 10, 2017

Keywords: homophily, similarity, measurement, associations, meaning

* Corresponding author: Barbara S. Lawrence, Anderson Graduate School of Management, UCLA, Los Angeles, CA 90095, (310) 825-2881, barbara.lawrence@anderson.ucla.edu.

ABSTRACT

Homophily, the tendency to associate with similar others, is a fundamental pattern underlying human relationships. It acquires meaning to those involved and thus influences their social interaction. Although scholars largely agree on its theoretical definition, their empirical measures show considerable variation. We review the homophily literature and induce a typology that captures the similarities and differences of these empirical measures. Most importantly, we found that measures often diverge from scholars' theoretical intentions. We address three questions: How do empirical definitions address the interdependent effects of opportunity structures and individual preferences on homophily? How do scholars' measures of associations and similarity obscure the meaning of homophily? Finally, what impact might scholars' measures of multiple social contexts, multiple types of associations and multiple similarities exert on homophily theory? We explore these questions to advance theory and propose a closer articulation between the measures and meaning of homophily.

One of the fundamental patterns underlying human relationships is homophily, the tendency of individuals to associate with similar others, where similar others are those with common attributes or consistent values (Lazarsfeld and Merton 1954). Homophily structures the multiple social systems to which people belong. It influences how communities form and recreate themselves, how status is distributed, and how subgroups evolve in occupations and organizations. People reliably make friends with and get advice from others of the same sex, race, age, values and religion more frequently than would be expected by chance (McPherson, Smith-Lovin, and Cook 2001, Verbrugge 1977). These findings, generated from a wide variety of settings including geographic regions, firms, communities and schools, place homophily in an elite category of social science concepts whose existence remains uncontested.

Scholarly interest in homophily has increased dramatically since Lazarsfeld and Merton (1954) introduced the term. In 2001, McPherson, Smith-Lovin and Cook published an extensive review of the homophily literature demonstrating consistent evidence that homophily occurs across various types of associations and dimensions of similarity. This has become the most highly cited paper in the history of the *Annual Review of Sociology* (2016). A Google Scholar search for homophily identified a twenty-fold increase in results during the past fifteen years, rising from 240 in 2001 to over 4800 in 2015. Despite this proliferation, surprisingly little attention has been paid to vast differences in how homophily is measured, providing a somewhat shallow conception of the phenomenon itself.

To explore how scholars define and measure homophily, we examined 538 empirical articles with homophily in the title or abstract published between 1954 and 2015. Scholars seem to label almost any theoretical concept involving the tendency to interact, type of association and attribute similarity components that Lazarsfeld and Merton (1954) identified in their original

definition as “homophily” and assume their empirical measures define a common phenomenon. For instance, in an organizational example, homophily is “the *degree* to which *pairs of individuals* who interact are *similar in identity or organizational group affiliations*” (Ibarra 1993:61)¹ With regard to members of the New York State Assembly, homophily is “the *degree* to which *pairs of individuals* who interact are *similar in certain attributes*” (Ferber and Pugliese 2000:402). Even with reference to the social behavior of raccoons, homophily is “the *tendency* for individuals to *interact with similar individuals*” (Hirsch et al. 2013:463). Although the definitions differ in their specificity, their basic theoretical intent is common.

In contrast to this consistency, scholars’ empirical definitions of homophily show considerable variation. This raises the question, do measurement differences matter and, if so, why? We consider two possibilities. First, such variation may obscure our understanding of the mechanisms that produce homophily outcomes. The academic literature is replete with examples. For instance, Lazarsfeld and Merton (1954) originally suggested that homophily occurs when people share similar status attributes, whether acquired through experience, such as education, or ascribed to the individual, such as race. However, Alstott, Madnick and Velu (2014) found differences in their effects. Homophily on acquired attributes increased the speed at which individuals mobilized others through their social networks to complete tasks, whereas homophily on ascribed attributes produced mixed outcomes. Such results suggest our understanding of the mechanisms by which homophily produces effects depend on how scholars approach empirical definitions.

Second, these measurement differences may obscure the theoretical meaning of homophily, suggesting that the phenomenon under study may vary even while the homophily label does not. For instance, some scholars measure men’s tendency to associate with other men

in a work division as a function of the availability of men versus women in the division. Such demographic availability is often called the work division's opportunity structure (McPherson et al., 2001).² Other scholars study employees' tendency to associate with similar others using individual preferences, in this case, men's preferences for men versus their preferences for women. Are these scholars studying the same "homophily?"³ We argue that the opportunity and preference measures of tendency present two different lenses through which to examine homophily, but either alone provides an incomplete picture. That is, any way one measures tendency, association or similarity produces homophily effects, but what exactly 'homophily' *means* to individuals may differ depending on the measure. The resulting theoretical differences in the meaning of homophily suggest a need for refinement.

While refinement is not uncommon in an extensive literature covering so many years, scholars rarely review these distinctions to advance and develop homophily theory. We probed homophily's empirical differences by examining the interlocking roles of substantive theory and empirical method, described by Lazarsfeld and Merton (1954). This approach encourages the use of method to clarify and explicate theoretical concepts rather than test them. Empirical measures are indicators of the phenomena under study, and as such, they always guide the direction of theoretical inquiry, determining what can and cannot be concluded from the results. As Merton ([1949] 1968:171) notes "The process of empirical inquiry raises conceptual issues which may long go undetected in theoretic inquiry." By identifying empirical similarities and differences, we induced a homophily typology that captures the most commonly-used, but often latent theoretical dimensions underlying each component (See Table 1).

-Table 1 About Here-

Overall, this paper makes two contributions. First, we review the empirical choices scholars use to measure the three homophily components: tendency, association and similarity. The variation we observe suggests that while homophily scholars appear to be studying the same concept in theory, they may be studying different phenomena in reality. Second, we identify a need for closer connection between the measures scholars use to identify homophily and the meanings people ascribe to them. This connection is rooted in the understanding that people engage in homophilous interactions because they are meaningful: They matter to those involved. We conclude that homophily is an intricately textured rather than homogeneous concept. As it remains a central construct in social science research, the divergence between homophily's measures and meanings offer new opportunities for research on social interaction.

METHOD

We used a qualitative, inductive process to explore the nexus between scholars' theoretical definitions of homophily and their empirical measures (e.g., Miles and Huberman 1994). Our interest was to construct a broad, multi-disciplinary and theory-based sample. Thus, we did not limit our search by discipline, journal or methodological approach. Using the Web of Science, we searched for articles including "homophily" in the title or abstract.⁴ Because our interest was the correspondence between theoretical definition and empirical measurement, we included only empirical articles. These search criteria identified 538 articles published between 1954 and 2015.

During our initial review of each article, we extracted scholars' theoretical homophily definitions and their empirical measures using verbatim quotes. We found the theoretical definitions largely consistent with Lazarsfeld and Merton's (1954) original denotation, with most including all three components: tendency, association and similarity. In contrast, the empirical

measures presented widely divergent methods for capturing the phenomenon's conceptual validity. Moreover, while some measures easily mapped to all three components, others did not.

In order to make sense of this divergence, we next used pattern coding to identify the similarities and differences. We started by considering what it was about each measure that appeared similar or different. For example, in examining how scholars define similarity, we noticed that in some studies scholars *asked* respondents whether a given similarity attribute was salient, and in others the scholars *assumed* the similarity attribute was salient. As salience is important to explaining individuals' experiences and behaviors (Randel and Earley 2009), we identified attribute salience as a relevant theoretical dimension. We then reviewed the sample studies again, identifying whether attribute salience was explicitly measured or implicitly assumed. Each study was assigned to a category using the question, "Who decided this similarity attribute was salient to the individual: the respondent or the researchers?"

We continued this process until all empirical measures were placed in categories based on within-group similarities and between-group differences. Eventually, we agreed on a set of dimensions that best described and distinguished the key elements of the three components. This produced the typology shown in Table 1. These dimensions do not include every nuance of all empirical measures, but they cover the dimensions that seemed most theoretically important to the studies we examined.

A TYPOLOGY OF HOMOPHILY COMPONENTS AND DIMENSIONS

What is a Tendency?

McPherson, Smith-Lovin and Cook (2001:416) describe tendency as "a contact between similar people (that) occurs at a higher rate than among dissimilar people." This raises several questions: What is a higher rate of contact, what is it higher than and within what social context

is this higher rate observed? The answers identify two dimensions of tendency measures: rates and social context.

Rates are empirical measures that identify the frequency of associations with reference to a given social context. We use the term “association” rather than the term “contact” because contact implies physical or direct communication and some homophilous behaviors require neither. Homophily studies typically use either expected rates, relative rates or hybrid rates. Tendency is defined by expected rates when the likelihood of association is significantly higher than what would result if associations were randomly distributed within a give social context. This measure is also called its baseline homophily (McPherson, Smith-Lovin, and Cook 2001:419). For instance, if 62 percent of an organization’s employees are white, the expected rate for whites associating with other whites is 0.62. When the proportion of whites’ associations with other whites exceeds this baseline, they are exhibiting a tendency to associate with same-ethnicity others, also called inbreeding homophily (McPherson, Smith-Lovin, and Cook 2001:419). Since studies that employ expected rates account for the availability of homophilous associations within a given social context, these measures tend to be used with homophily theories predicated on opportunity structure.

Espelage, Holt, and Henkel (2003) use an expected rate to assess the tendency of students to engage in bullying and fighting within the social context of middle school peer groups. All else equal, the distribution of bullying behavior should be the same across peer groups. However, if students who bully show a tendency to associate with one another, then some peer groups will have a higher level of bullying behavior than others. Results from a hierarchical linear model of males’ peer group bullying show a significant intra-class correlation, indicating there is an

inclination for male bullies to associate with other male bullies in their peer groups more frequently than would be expected by chance if they were randomly distributed.

Tendency is defined by relative rates when an individual or group exhibits a higher number of associations with others than does another individual or group within a given social context. Unlike expected rates, relative rates assess comparison, but do not account for the population. For example, assume that 40 percent of all college students who attend parties believe that average daily beer consumption is five beers, while 20 percent of all non-party attendees share this belief. In this example, party attendees exhibit a stronger tendency towards homophilous beliefs about beer consumption than non-party attendees. Studies that employ relative rates tend to be used with homophily theories predicated on individual preferences.

Eyal and Rubin (2003) use relative rates to assess whether individuals' tendency to associate with television characters increases with the extent to which they perceive themselves as similar to those characters' attitudes and beliefs. In this example, the social context is a dyad including the individual and the television character. Subjects' homophily with the character is measured by their "strongly disagree—strongly agree" responses to a level of wishful identification scale. This scale includes items such as "I wish I could be more like the character," "I'd like to do the kinds of things the character or personality does on the program," and "When I watch the character on the program, I imagine myself in his/her place" (P. 86). Thus, subjects are more or less likely to identify with the character relative to one another. Here, the opportunity structure is neither identified nor measured.

Tendency is defined by hybrid rates when both the distribution in the population and the comparison between two groups (Behrman, Kohler, and Watkins 2002:716) are considered in a given social context. This is important because neither expected nor relative rates are useful,

alone, for comparing two groups whose available associations differ. A typical example involves homophilous ethnic associations. Take, for instance, a government agency whose employee population is 60 percent white and 40 percent Hispanic. The expected rate of associations with similar others for all subjects is 60 percent white and 40 percent Hispanic. A relative rate of associations with similar others is the comparison of within-ethnic associations of whites to those of Hispanics. Because there are more whites with whom to have associations, the results would likely show a higher proportion of within-ethnic associations for whites than Hispanics. This might lead to a possibly erroneous individual preferences interpretation that whites avoid associations with Hispanics.

Hybrid rates resolve this error. There are several ways of measuring hybrid rates (e.g., Kossinets and Watts 2009, Stark and Flache 2012). One possibility is to consider standardized values of two distributions within the same social context. In contrast to the government agency example above using relative rates, hybrid rates would compare the likelihood of within-ethnic associations for whites and that for Hispanics, accounting for expected rates in this agency. For instance, if white employees' average within-ethnic associations are 2.7 standard deviations *below* their expected rate and Hispanic employees' average within-ethnic associations are 2.7 standard deviations *above* their expected rate, we could infer that the tendency of whites to engage in homophilous within-ethnic associations is lower than that of Hispanics.

Another hybrid measure of homophily is the point correlation coefficient, often called an availability-adjusted measure. This measure ranges from -1 to 1, with "0" indicating no difference between the individual's actual associations and his or her expected associations given the numbers of available similar and dissimilar others. The properties of this measure are discussed in Gower and Legendre (1986, Kossinets and Watts 2009). Although scholars use it to

control for the opportunity structure, it empirically assumes that the effect of the opportunity structure and individual preferences are independent. The only variation uniquely attributed to individual preferences is that which is left over after accounting for the population distribution.

Ibarra (1992) uses this measure to assess men's and women's tendency towards same-sex associations in an advertising firm. Professional staff and account assistants in a small advertising firm (N=79) were asked to provide the names for five different associations, including names of those with whom they communicated in general, with whom they spoke to influence decision outcomes and from whom they received advice and support. Two measures of homophily were computed for each of the five associations. The first, a relative rate, identifies the percentage of same gender ties, a measure of individual preferences. If a woman provides names of eight friends of whom six were women, her homophily score for friends is six divided by eight or 75 percent. The second is a hybrid rate, which includes this information and adds available uncited ties. Available uncited ties include all the women and men she might have cited but did not. When appended to same gender ties, they account for the opportunity structure.

The relative and hybrid rate measures in Ibarra's (1992) article produce different interpretations. For instance, the relative rate measure of tendency shows that women select 33 percent same sex ties for their associations that influence decisions (1992:435). This might be interpreted as showing that women exhibit stronger preferences for men than women when they tap influence associations. However, the hybrid rate of -0.09 suggests that, given the low availability of women as influence associations, this tendency to prefer men is quite weak. Thus, using measures that account for both opportunity structure and individual preferences may produce different interpretations than using either one alone.

Social context. The fundamentally social component of homophilous behavior means that assessing a tendency requires defining the social context within which it occurs. The primary social context is the population within which individuals' tendency to associate with similar others is assessed. Examples include a dyad, group, organization or community. A secondary social context is another population theoretically relevant for evaluating individuals' tendency to associate. The primary social context for the bullying behavior in the Espelage et al. (2003) study is peer groups. Peer groups are primary because they define the population within which the tendency towards bullying is evaluated. A secondary social context is students' grade level. The authors find that bullying differs across sixth, seventh and eighth grade peer groups. In Eyal and Rubin (2003), the primary social context is the dyad of the subject and the television character. Although unused for interpreting the results of this study, a secondary social context is the sample of 219 college undergraduates enrolled at a Midwestern university from whom these data were collected.

The population is clear when using expected rates because they require a known social context, usually with formal boundaries, such as those designating cities or schools. The population for relative rates is often less clear because the boundaries may be informal, such as those surrounding individuals watching television. Here the population for each individual is identified, but creating meaningful formal boundaries for the set of all individuals' associations may be difficult.

A classic example of important but unused secondary social contexts involves the long history of laboratory research based almost exclusively on Western, Educated, Industrialized, Rich and Democratic (WEIRD) undergraduate students (Henrich, Heine, and Norenzayan 2010). The students in each of these studies are considered the population for generalizing the results.

Yet, it seems likely, for instance, that the tendency to select others of similar sexual orientations, differs dramatically for undergraduates in Western and Eastern cultures. Scholars should consider the theoretical implications of secondary social contexts as limitations to be studied in future research, even when this population is unmeasured.

In summary, identifying a tendency involves measuring the rate at which similar associations occur and identifying the social context within which they occur. Homophily studies utilize many different measures of rates, but the principle underlying each remains the same. If homophily is measured as the number of an individual's associations relative to an expected number within the social context, it is an expected rate. If homophily is measured as a comparison between the number of an individual's associations and the numbers of other individuals or groups within the social context, it is a relative rate. Hybrid measures include aspects of both. Tendencies are measured in primary social contexts. Secondary social contexts may suggest additional theoretical effects and should be identified and considered when interpreting the results.

What is an Association?

An association is a connection between one entity and another or one entity and multiple others. The connection might or might not involve communication and might occur at any level of analysis. For example, in Espelage et al. (2003), students who bully show a tendency to be friends with one another. Associations might be connections between two people, between one person and multiple others, between one person and a group, between two groups, between two organizations or between two countries. For simplicity, we use individuals as the level of analysis.

Although associations may include any characteristic of connections, we identify four that seem to appear particularly often. The association's *relationship type* indicates the content defining the association, for instance, communication or friendship; its *symmetry* signifies whether two individuals both see themselves as connected to one another; its *perspective* establishes whether the individual subjects or researcher decided the association exists and its *salience* identifies how important the association is to the individual.

Relationship type. Many of an individual's associations involve several *relationship types* simultaneously, but studies frequently focus on one of four: instrumental, expressive, identity and knowing. Instrumental associations involve a specific task-related activity or inferred function that circumscribes the interaction. For example, Fischer's (1982:36) Northern California Communities study examined aid one individual might provide another, such as watching one's home or getting advice on important decisions.

Expressive associations include affect. Affect may involve social or emotional support one individual gives another, developmental relationships, or friendships that bring individuals together. Behrman, Kohler & Watkins (2002) asked Kenyan women for the names of those with whom they had gossiped about the topic of family planning. Although these women ultimately employed the information they acquired to make decisions on contraceptive use, they selected associations defined by affective interactions described as informal or "chatting." Thus, the associations were with others relevant to their everyday, social support from friends and family.

Identity associations involve individuals' perceptions that their self-concept connects them with another individual or group. This view of identity associations draws on Tajfel (1978): "that part of an individual's self-concept which derives from his [sic] knowledge of his [sic] *membership of a social group (or groups)* together with the value and emotional significance

attached to that *membership*” (italics added). An individual might distinguish another as “like me,” such as in the Eyal and Rubin study (2003) above. The most commonly studied identity associations involve ascriptive characteristics such as gender, race and age. Other identity groups to which individuals might belong or be perceived to belong include occupations, such as carpenter, roles, such as motherhood, and organizational affiliations, such as alumni from a particular college.

Knowing associations involve knowledge or awareness. Knowing associations might involve recognition of a photograph or awareness of another person’s name on the cc: of an email. In order to identify a “knowing tie,” Lawrence (2006) asked managers in a large organization to list all the people they know within the firm. Brown and Reingen (1987) asked respondents to look through a word-of-mouth list of people involved in referrals to three piano teachers and “mark those persons that they knew” (p. 355).

Symmetry. An association’s *symmetry* denotes whether the associations in which the individual is involved are reciprocated by the others with whom he or she is involved. Although seldom mentioned as a theoretical component in homophily research, symmetry may play an important role. For instance, it seems probable that individuals in peer groups characterized by symmetric associations are more likely than those with asymmetric associations to view themselves as similar to one another and thus show a greater tendency to associate.

In some cases, symmetry is measured directly by asking the individuals on both sides of the association whether they perceive the association, or by examining relationships in which both parties are present. For example, in Martin et al.’s (2005) study of dyadic play between preschool children, the associations measured are symmetric because the children play together. Clean empirical measures of symmetry require information from both individuals. This is not

always available and researchers often infer their existence. This frequently occurs in studies of identity associations when researchers assume that individuals in the same similarity attribute category experience in-group symmetric associations, or by asking only one member of the association. Dellande, Gilly and Graham (2004) measure homophily by asking patients to evaluate themselves and their providers on a set of demographic attributes and attitudes. However, the providers are not asked to provide information on their patients. The potential errors involved in assuming symmetry without direct measurement may produce difficulty with theoretical inferences made from the results.

Asymmetric associations typically require direct measurement. One group of individuals is asked to select its associations with a group of others. When the others are similarly asked to select associations with the first, many do not reciprocate. Direct measurement is important because research suggests that individuals often misperceive the reciprocity of their associations with others. Almaatouq, Radaelli and Halford (2016), in a study of 84 undergraduates, found that over 50 percent of subjects who identified friendship associations they thought were reciprocated, were not.

Perspective. An association's perspective identifies whether the individual or researcher decides the connection exists. In some studies, subjects make these decisions and provide the information to the researcher. Stackman and Pinder's (1999:60) study of personal work networks asks respondents to identify individuals considered to be "useful in performing the tasks required of your current job" and those considered to be "friends of yours." In other studies, researchers infer subjects' associations from other available data. For instance, Appold, Siengthai and Kasarda (1998) assume that women in high-skilled jobs associate with one another even though they were not themselves asked whether they do associate. Given the likelihood that unmeasured

associations do not always exist as researchers assume, identifying who decides that the association exists is important for understanding the meaning of the association to those involved.

Salience. An association's salience identifies who decides whether the association is important to those involved. This is relevant because the association's perceived importance to the individual or individuals involved should influence the magnitude of their tendency to associate. An individual who sees a particular association as significant to his or her emotional health is more likely than someone who does not to associate with those others. In other cases, the association's perceived importance may matter less.

In summary, an association involves several characteristics. The association type identifies the content that defines the connection. Symmetry indicates the extent to which the association is reciprocated among those involved. Perspective clarifies whether the respondent observes or researcher infers the presence of an association. And finally, salience indicates the extent to which the association is important to those involved.

What is Similarity?

Similarity is perhaps the most basic component of homophily: the criterion on which individuals tend to associate. Individuals are similar to the extent that both hold some characteristic or attribute in common. Importantly, similarity does not mean an association exists between the individuals involved. An individual may be similar to others and, at the same time, have no associations with them. This distinction between similarity and association is important because it requires researchers to consider whether similarity plays more than an intra-individual role. For example, while two individuals may share an attribute, perhaps as alumni of the same university, that does not mean they necessarily have an association with one another. We discuss

three dimensions of associations: the *attribute* on which similarity is measured, the *perspective* from which similarity is defined, and the *salience* of the similarity to those involved.

Attribute. People use two attribute categories when assessing similarity (Lazersfeld & Merton, 1954; McPherson et al., 2001). The first is status, in which similarity is based on ascribed status such as gender, age and ethnicity, and acquired attributes, such as education, occupation, religion, network position, individual behaviors and social class. The second is values, in which similarity is based on values, attitudes, and beliefs. Although scholars disagree about the definition of values (Rohan 2000), we use values as defined by McPherson et al. (2001:419): “the wide variety of internal states presumed to shape our orientation toward future behavior.” Such values might include cognitions, emotions, practices or attitudes. Most homophily research focuses on status homophily, perhaps because ascribed and acquired statuses tend to be easier to identify and measure than values. Fischer, Gainer and Bristor (1997) provide an example of both status and value similarity. Their study examines how the consumers’ similarity to service providers influences their expectations of service quality. Similarity was studied using two attributes: 1) a status attribute—the sex of the consumer and the service provider, and 2) a value attribute—consumers’ beliefs of whether service providers in the occupation are stereotypically men or women.

Perspective. The perspective of a similarity attribute varies depending on who decides that two individuals are similar: the individual or the researcher. This criterion is important because the perspective from which similarities are identified exerts considerable impact on what can be inferred from the results. Researchers often assign individuals to attribute categories and assume that the individuals see themselves as similar to all others in the category.

Yet, people may categorize themselves differently than do researchers and this introduces error in the analysis. For example, in ethnically-diverse countries, people may not categorize themselves in the same way that researchers might. An Indian-American physician may be identified as similar to other Asians by a researcher, even when the physician does not see him or herself as similar. Diverse self-categorizations have become common in the United States; the last decennial census allowed subjects to specify multiple ethnic group memberships. Thus, specific measures of attitudes, abilities, beliefs or aspirations are necessary to evaluate similarity.

Saliency. A final criterion for measuring similarity is the *saliency* of the attribute to the individual(s) involved. Saliency is important because it seems likely that the more important a similarity attribute is to individuals, the more likely that homophily on that attribute influences their tendency to associate. Here the largest difference in empirical measures results from who decides the attribute is salient to the individual. In many homophily studies, the researcher selects a characteristic to study and assumes that characteristic is salient to the subject. For practical reasons, inferring rather than measuring attribute saliency tends to be a function of the increasing size of the sample and level of analysis. For instance, demographers who examine large, secondary data sets, such as the population of a city or country, have little control over whether the data include measures of saliency. The United States Department of Labor Statistics, for example, asks generalizable population samples many questions about work, but they do not measure whether their respondents find any of their attributes salient.

Mollica, Gray and Trevino (2003) provide an example of research in which the saliency of subjects' identity is measured. They examine the effect of social identity on the relationship between homophily and same-race friendship ties. They studied MBA students at the beginning of a two-year program and again at the beginning of the second semester. Prior to collecting

friendship ties, they had students complete a social identity salience survey on which they were asked to select their “social identity, or identification with groups” (2003:127) from a list including race, gender and religious affiliation. Subjects were then asked which of these identities was most important to them. This identified a subset of students for whom race was the most important social identity.

The authors then used the point correlation coefficient, a hybrid rate (Gower and Legendre 1986), finding the hypothesized relationship between racial homophily and friendship ties: Subjects had more friendship ties with those who were same-race than with those who were other-race. When they then controlled these relationships for students who selected race as their most important social identity, the relationship between homophily and friendship ties became significantly stronger. This relationship remained the same when data were collected at the beginning of the second semester. Thus, the salience of a similarity attribute influences students’ tendency to associate with similar others.

In summary, examining similarity involves identifying the basis of the similarity, the perspective from which the similarity is defined and the extent to which that similarity is important to those involved.

IMPLICATIONS FOR HOMOPHILY THEORY AND RESEARCH

Homophily matters because it acquires meaning to the people involved and therefore influences their social interaction. As we explored differences in empirical measures, questions about meaning kept reappearing. For instance, by definition, homophily requires that people associate with similar others. However, the meaning of associations in which people perceive similarity may differ from the meaning of associations in which researchers assume similarity. Thus, the perspective from which similarity is measured influences how we understand

homophily and, for this reason, it became a similarity dimension in our typology. Recognizing such differences in empirical measures helped us uncover several theoretical muddles.

We focus on three: First, how do empirical definitions address the interdependent effects of opportunity structures and individual preferences on homophily? Second, how do scholars' measures of associations and similarity obscure the meaning of homophily? Finally, what impact might scholars' empirical measures of multiple social contexts, multiple types of associations and multiple similarities exert on homophily theory? Each of these questions suggests that a thorough understanding of this phenomenon may be more complex than most current research acknowledges.

Q1. How Do Empirical Definitions Address the Interdependent Effects of Opportunity Structures and Individual Preferences on Homophily?

Many scholars note that both the opportunity structure and individual preferences facilitate homophilous behavior (e.g., McPherson, Smith-Lovin, and Cook 2001, Kleinbaum, Stuart, and Tushman 2013, Kossinets and Watts 2009, Feld 1981, Ibarra, Kilduff, and Tsai 2005), but few studies explicitly investigate both. In part, we attribute this separation to the increasing distance between social psychological and sociological measurement traditions. Extant homophily research typically resides within either the former, in which the social context involves dyads and small groups, or the latter, in which it includes larger social systems, such as communities or cities. Over time scholars have unintentionally reinforced this division through the types of data they collect, the empirical measures they use and the publication norms they have introduced and reinforced over time.

The earliest homophily research started somewhere between the two. Lazarsfeld and Merton (1954) were active scholars from a mid-twentieth century group of psychologists and

sociologists who explored social interaction using field studies. They explicitly established social interaction as a level of analysis where, even though they did not use the terms “opportunity structure” and “individual preferences,” these theoretical ideas were acknowledged and explored. For example, Jacob Moreno (2011b) an early sociometrist and sometime student of Sigmund Freud, disagreed with Freud’s intra-individual approach. Although known primarily for sociometry, Moreno (2011a) was an early proponent of connecting the contextual and psychological perspectives, maintaining that Freud ignored the “social” in social interaction:

Well, Dr. Freud, I start where you leave off. You meet people in the artificial setting of your office. I meet them on the street and in their natural surroundings. You analyze their dreams. I give them the courage to dream again. You analyze and tear them apart. I let them act out their conflicting roles and help them to put the parts back together again (P. 68).

Other scholars contributed to this view of individuals as embedded within field settings. Kandel (1966) examined how the socioeconomic homophily between physicians (residents) and patients in a mental health hospital influenced which patients received psychotherapy. She found both opportunity structure and individual preference effects on homophily. Residents were most likely to recommend psychotherapy for patients when the resident and patient had similarly high socioeconomic status. They were least likely to recommend psychotherapy when the resident was high in socioeconomic status and the patient was low. When the socioeconomic status of residents was low, the status of patients made little difference to the percentage who received psychotherapy. This pattern was even more pronounced in a structural condition when wards were examined separately. In the ward with the highest percentage of high status residents, homophily on socioeconomic status made the greatest difference. As Kandel (1966:648) notes “Whatever the specific factor at work, it seems that the composition of the social context can significantly affect the interpersonal processes taking place within it.”

During the 1970's and 1980's two factors split this mixed community into more distinct social psychological and sociological branches. First, the number of scholars conducting field studies declined. This made it easier to disregard the complex relationship between structure and preferences. Second, it enabled more discipline-focused research. Social psychologists started using laboratory methods with college students in conditions where the students were usually meeting for the first time (See reviews in Byrne and Griffitt 1973, Huston and Levinger 1978). Sociologists, with the advent of survey methods and computerized analysis, moved to increasingly large social contexts, studying communities, social movements and national populations. Kanter (1977) reflected on this split between the communities:

Most analyses to date locate male-female interaction issues either in broad cultural traditions and the sexual division of labor or in the psychology of men and women whether based on biology or socialization (Kanter 1976). In both macroscopic and microscopic analysis, sex and gender components are sometimes confounded by situational and structural effects (P. 968).

One consequence of this disciplinary split is that two different concepts emerged to explain organizational homophily, defined here as the increasing similarity of organizational members to one another over time. Social psychologists call it attraction, selection and attrition (Schneider, Goldstein, and Smith 1995, Schneider 1987): individuals' tendency to be attracted to, hire or promote similar others which, with attrition over time, increases the overall similarity of organizational members. Sociologists call it homosocial reproduction: the tendency for people to maintain hiring patterns that replicate organizations with "people like me" (Rivera 2013), a term attributed to Wilbert Moore (1962). Both concepts suggest almost identical mechanisms to explain organizational homophily but scholars have not questioned whether these concepts are theoretically or empirically interchangeable.

By and large, the data collection and measurement limitations that aided the split between

micro- and macro-level homophily studies persist. Studies of the opportunity structure generally require large populations and large scale databases, such as the General Social Survey (GSS), a national probability sample of the United States (1972 to present). Yet, the GSS is one of the few national samples that includes individual-level attitude, behavior and attribute information, which facilitates at least some study of homophily including both opportunity structure and individual preferences.

Most surveys of large social systems involve many restrictions. They are expensive and time consuming. Quick replications are impractical and questions that cannot be answered with data on hand remain unanswered. Successful funding generally requires that depth on any given topic be limited so that the data can be useful for a wide range of scholars interested in different topics. Moreover, restrictions on overall survey length mean that specialized questions, such as detail concerning social interactions, are carefully allocated. Adding a single name generation question to the 1985 GSS, which limited subjects to five names, warranted an entire *Social Networks* article debating its costs and benefits (Burt 1984). Thus, homophily research at this level of analysis tends to focus on opportunity structure as it presents the closest theoretical match between the phenomenon and available data.

Studies of individual preferences generally require data obtained through observation, laboratory studies, interviews or social network surveys in small to medium sized social settings. For example, Flynn, Reagans and Guillory (2010) present three studies on the psychological mechanisms underlying individual homophily. Each uses a different sample: a class of 53 MBA students, a group of 49 same-university college students and a group of 49 on-line participants.

Using experimental methods, the authors observe that subjects with high Need for Closure are more likely than those with low Need for Closure to assume that racially-similar

individuals communicate with one another, even in cases where they do not. Although all three studies shed light on the psychological mechanisms underlying individual-level homophily, only the first, with its classroom data, would permit hybrid inferences about the concurrent effect of social context.

More generally, then, the scarcity of research on the interdependence between the opportunity structure and individual preferences reflects two increasingly insular and distinct research traditions. Cialdini (2009) observed that while developments in social psychology, such as increasing publication standards that require multiple studies using mediational analyses, facilitate research quality, they discourage scholars from working with field data. Further, studies in naturally occurring settings, where both structural and preference mechanisms can be observed, take a long time and are difficult to publish. After submission, reviewers from both perspectives require a depth of research and inquiry from their own point of view. Doing this well requires more length than typically feasible in a journal article, so it necessitates a tradeoff between full depth of both perspectives and length cutoffs.

This disciplinary divergence is unfortunate because the nexus between the opportunity structure and individual preferences is where people construct much of the meaning in homophily. That is, these two are connected in how people construct and agree about their perceptions of their social context—their socially constructed reality. This is neither a structural nor an individual level phenomenon. Certainly the way attributes are distributed in the population constrains association probabilities. However, they also affect the ways in which people value similarity attributes. Lazarsfeld and Merton (1954) first connected this process to homophily by noting that attributes acquire meaning, which produces feelings of similarity and increases individuals' tendency to associate. The significance of age in homophily research, for

instance may result, not because age induces cognitive biases or recognizable distinctiveness, but because it acquires shared understandings within social systems (Eisenstadt [1956] 1962, Lawrence 1987, Neugarten, Moore, and Lowe 1965). Age indexes the socially-constructed schema that represent these associations in individuals' minds.

Thus, another mechanism influencing homophily is social-construction, the shared meanings that attributes acquire. This differs from the mechanisms of opportunity structure and individual preferences because shared meanings need not be observable in population distributions or individual preferences. Rather, they begin with individuals' interpretations of the perceptions and experiences that establish their definitions of associations and similarities. When these interpretations become common understandings, they are socially-constructed. Thus, even when the opportunity structure underlies interactions, socially-constructed homophily results because individuals develop shared understandings to explain everyday events. These understandings then influence their tendency to associate with similar others.

Q2. How Do Scholars' Measures of Associations and Similarity Obscure the Meaning of Homophily?

The association and similarity components of our homophily typology both include dimensions of perspective and salience. These dimensions flag two persistent theoretical concerns. The first is that scholars tend to confuse associations and similarities. They often assume that individuals who share similar attributes *necessarily* associate, when they may not. An example is individuals' professions. Individuals may identify with a particular profession because they are a member of it. However, they may not see themselves as similar to all others in that profession. Homophily studies of professions thus require two distinct measures: one of subjects' identity associations with same-profession others, and another of subjects' similarity to

same-profession others. Any study with only one is not homophily research. The second concern is assuming associations and similarities matter equally to everyone. This occurs when researchers assume, but do not measure, the association or similarity's salience to individuals.

Both problems result from inattention to the theoretical implications of measurement: From whose perspective are associations and similarity defined and to whom are these associations and similarities meaningful? Scholars typically make a choice between using measures defined by subjects and measures defined by others. A self-defined measure is one where subjects identify for themselves whether they have an association, whether the association is with a similar other and whether it is salient. An other-defined measure is one where someone besides the subject defines whether he or she has a given association, identifies whether the association is with a similar other and finds it salient. This empirical choice is significant because it connects associations and similarity directly to the meaning that emerges during social interaction. When associations and similarities are self-defined, we have greater confidence they are meaningful for the individuals involved.

These measures may also influence scholars' consideration of the interdependence between opportunity structures and individual preferences. Self-defined measures tend to divert attention away from opportunity structure explanations. When researchers collect individual level data, for instance from MTURK surveys, they typically learn about individuals' preferences but acquire no data about their social context. This produces studies where the results are theoretically assumed to be generalizable across populations, but their actual generalizability is unknown.

Conversely, other-defined measures tend to divert attention away from individual preference explanations. When researchers collect or use contextual data, such as secondary data

from the U.S. Bureau of Labor Statistics' Current Population Survey, they typically learn about the context but do not identify individuals' perceptions. Researchers may mistakenly assume that people associate, that the association is important to them, that a particular similarity is the reason for their association or that a particular similarity is important to participants. Such assumptions produce unknown error variance.

Indeed, in our review, we found homophily scholars frequently use other-defined measures of association and similarity when their theory calls for self-defined ones. This occurs for both the perspective and salience dimensions of these components. We elaborate on the role these assumptions play in homophily research by examining how scholars study the perspective and salience dimensions of the association and similarity components: association-perspective, association-salience, similarity-perspective and similarity-salience. These four are identified with notes in Table 1.

Association-perspective: When individuals' self-defined associations differ from those ascribed to them by researchers, researchers are assuming the associations exist when, in fact, they may not. Appold, Siengthai and Kasarda (1998) compare the efficacy of three theoretical perspectives that examine organizational forces and labor supply factor explanations for the number of women in high-skill jobs. One of these perspectives is employees' preferences for social homophily. The researchers use gender as an indicator of identity-based associations. Employees of the same gender are assumed to associate, even though they were not themselves asked whether they associate. This generates measurement error since we do not know if women actually identify with other women, with other genders or with others in comparable jobs. As a result, the authors' conclusion that social homophily plays a greater role than the other two theories depends on an unknown level of measurement error in gendered associations.

There may also be more nuance to understanding the association-perspective. For example, individuals may have little conscious awareness of their knowing associations, such as the reference group of others they use for social comparisons. However, these associations may well define the social context within which they make sense of all their perceptions.

Association-salience. Many homophily studies measure an association's salience indirectly: It is inferred or assumed. Social network researchers often use self-defined measures of association but other-defined measures of the association's salience. They may ask subjects to provide the names of instrumental and expressive associations, such as the individual's physicians for a critical illness or people who create trouble for him or her at work. This allows subjects to select their own associations. However, researchers might assume that all identified physicians are important to subjects rather than asking subjects how important each physician is. In such studies, any instrumental association is assumed to be highly salient to subjects. Other studies use tie strength to infer salience by assessing either the frequency or length of communications (Ibarra 1993:61), with increasing levels inferring increasing salience. In these cases, the degree to which self-defined associations are salient to subjects is often unknown.

The Appold et al. (1998) study above infers the salience of individuals' gendered-identity associations by assuming that the higher the proportion of women, the more positive their work attitudes. As they note "To the extent gender-based social homophily is salient to employees, *the effect of sex composition should be reflected in important indirect indicators*, such as (1) organizational commitment, (2) job satisfaction, and (3) the employees' relationship with their immediate supervisors" (P. 547, italics added). The problem with this measure is that it assumes that these attitudes indicate the salience of gendered-identity associations to respondents. To the extent that they do not, the study's results may hold little explanatory value for the authors'

social homophily hypothesis.

Similarity-perspective. If researchers' data establish that people are associating but then use other-defined measures to evaluate whether people see themselves as similar on an attribute, they may be incorrect. For example, a researcher may know the similarity attribute, such as whether subjects are men or women, and then assume that subjects see themselves as similar. Yet, these perceptions of similarity may have nothing to do with gender. They may result because of some other attribute. For example, some subjects may be members of the same athletic club and thus see themselves as similar. In this case, the similarity attribute is not necessarily attributable to gender.

Another example is using subjects' last names to identify their ethnicity. For instance, those with surnames of Santos or Lopez are categorized as Hispanic. However, researchers may inaccurately categorize subjects' ethnicity because they have unusual last names or because their last names are totally unrelated to their ethnicity. This might easily occur in inter-ethnic marriages where one partner takes the other's surname. This measurement error sounds similar to the error involved in using other-defined measures of association-perspective. However, it differs because here the important information is not whether these subjects associate, but whether they see themselves as holding the same similarity status or values.

Yamaguchi and Kandel (1993) provide an example of using other-defined measures of similarity when self-defined measures appear closer to their theoretical intent. This study examines whether drug use prior to marriage results from an increased tendency to make marital selections based on similarity in drug use or whether it results from the socialization to or influence of one partner on the drug use patterns of the other after marriage. The association perspective is self-defined because subjects identify themselves as married. The salience of the

marital association is also self-defined by asking subjects questions about the closeness of their relationship.

The measures of similarity are less clear. The similarity attribute, marital partners' drug use, is other-defined. Subjects did provide researchers with information about their own drug use. However, researchers did not ask subjects whether they, themselves, thought their drug use was similar to that of their partner's. Rather, researchers used an other-defined measure. They estimated the concordance over chance between the partners' self-defined drug categorization. The evidence suggests concordance is likely a good measure of similarity-perspective, making it a reasonable other-defined indicator of similarity. However, the theory suggests that individuals make these decisions and a self-defined measure would be preferable.

Similarity-salience. Problems with similarity-salience result when other-defined measures do not capture the salience the similarity attribute holds for subjects. A typical method of measuring salience is to assume that all social category members experience their membership as similarly salient. For instance, third-generation immigrants typically find their ethnic identity unimportant until they are placed in a social context that requires they see themselves as "different." A good example includes the many Japanese-American citizens resettled into American concentration camps after the Japanese bombed Pearl Harbor in 1941. Many of these U.S. citizens saw themselves as Americans, not as Japanese, but they were forced by the government to move and give up their belongings, businesses and friends.

In the Yamaguchi and Kandel (1993) example above, the similarity attribute, same drug use in marital partners, is other-defined. Since the authors did not ask subjects whether they perceived their partners to use drugs in similar ways, they also could not assess whether the subjects believed their similar drug-use was important to their decision to marry. That is, the

authors do not provide data on the salience of similar drug use. This matters theoretically because even though the marital partners' drug use is similar, it is unknown how important this similarity was to their marital decision. However, understanding the conditions under which tendencies to associate increase, including peoples' awareness of the reasons for their decisions to associate, would likely be a contribution to developing homophily theory on marital partner drug use.

The interaction between self- and other-defined measures. Thus far, we have presented examples where the researchers' selection of self- and other-defined measures is based more on available data than on theoretical interest in how these differences affect homophily theory. In some studies, researchers use a mix of self- and other-defined measures of perspective and salience because they are important to the theory being tested.

Simpson et al. (2000) provide an example. Their study examines the effect of racial and ethnic homophily on intent-to-purchase based on print advertisements. Their sample includes 98 undergraduate and graduate students identified as black by University records. The experimental design involves showing subjects photographs of either a white couple or a black couple endorsing a product. Here, the association-perspective is other-defined because the researchers decided whether subjects identified with the race of the couple they were shown. Researchers defined black subjects shown the black couple as racially congruent and black subjects shown the white couple as racially incongruent.

The researchers then collected a self-defined measure of similarity-perspective, subjects' perceptions of their similarity to the couple. This was assessed with McCroskey's perceived homophily scale (McCroskey, Richmond, and Daly 1975, McCroskey, McCroskey, and Richmond 2006), which includes questions about subjects' perceived similarity to the couple on

attitudes, values, appearance and background. Examples include: shares my values – doesn't share my values and looks different from me – looks similar to me.

The salience measures in this study are difficult to categorize. Subjects' association-salience and similarity-salience both appear measured using ethnic identity, a self-defined measure of subjects' personal association with being black and their sense of belonging to the black race. This was assessed with the 13-item Whittler ethnic identification scale (1989), which includes two factors: cross-race attraction and support of black causes.

Interestingly both self- and other-defined measures of subjects' ethnic associations play a role in these results. The results show an interaction between an other-defined measure of association-perspective, the researchers' definition of subjects' racial congruence with the couple, and a self-defined measure of association-salience and similarity-salience, their ethnic identity. The dependent variable is the self-defined measure of similarity-perspective, perceived homophily. Without using the self-defined measure of salience, ethnic identity, the authors would have found a simple positive effect of racial congruity on self-defined perceived homophily, suggesting that all subjects found their ethnic identity similarly salient. They would not have observed the impact that ethnic salience had on subjects' perceived homophily.

Summary. Meaning plays a crucial role in homophily theory. Without knowing whether associations or similarities matter to the individuals involved, we cannot understand the complexities of why individuals tend to associate with similar others. Researchers should acknowledge the theoretical importance of using and distinguishing between other- and self-defined measures of association and similarity, even if the data are unavailable.

Q3. What Impact Might Scholars' Measures of Multiple Social Contexts, Multiple Types of Associations and Multiple Similarities Exert On Homophily Theory?

Theories that involve many components, may require multiple social contexts, multiple types of associations and multiple similarities. Scholars generally agree that individuals' tendency to associate increases the more social contexts within which tendency is measured, the more types of associations individuals engage in and the more similarities individuals share. However, this notion, which assumes that more social contexts, more associations or more similarities produce more homophily, requires additional specification. We discuss three different approaches for studying the potential complexities of "multiples."

One approach is for scholars to include a theoretical rationale for selecting which types of social contexts, associations or similarities they study. Some are chosen based on previous research or because they are available, while others are chosen based on their relevance to the theoretical question. Aguilar-Gallegos et al. (2015) select a set of similarities by identifying attributes that together represent a global indicator of subjects' similarity. They studied oil palm growers in Mexico to identify which similarities in new technology and practice adoptions appeared positively related to economic value. The authors visited and collected data from 104 growers in Mexico. Eight categories of new technologies and practices emerged, including for instance, plant nutrition, plantation management and genetics. The authors compared each grower with each other grower and classified dyads by the number of common technologies and practices adopted. Grower dyads were then clustered based on their similarities and differences.

They found three clusters of oil palm growers with homophilous adoption patterns: basic adopters were primarily male, of intermediate age and received on average less than 50 percent of their income from oil palms; intermediate adopters included 25.7 percent female growers, of older age, and longer experience and who received on average over 82 percent of their income from oil palms; and advanced adopters who were mostly men, younger than basic adopters, but

with more experience and who received more than 78 percent of their income from oil palms.

The authors then studied communication across the three clusters by asking growers, “Who do you learn from and/or who do you consult when you encounter a problem in your plantation?” The answers suggest that growers using similar technologies and practices primarily communicate within but not across clusters. Basic adopters and advanced adopters do not communicate. A few intermediate adopters communicate with basic or advanced adopters. The results suggest that the diffusion of new technologies and practices in Mexican oil palm production is constrained by the interactions among growers with homophilous adoption patterns. This research is unusual in that the authors conducted a field study to identify all the relevant similarities before evaluating the impact of homophily on economic value.

A second approach is for scholars to consider that each additional context, association or similarity may exert different effects on homophily. McPherson, Smith-Lovin and Cook (2001:437) suggest that people are always involved in multiple, overlapping social contexts that may exert “different levels of homophily on different dimensions.” For instance, Rotolo and Wharton (2003) examine whether the gender composition of an individual’s occupation influences his or her tendency to join voluntary organizations with a similar composition. They find that the higher the proportion of women in a paid occupation, the higher the proportion of women in the voluntary organizations to which they belong. One explanation may be that the history of voluntary organizations indelibly marks them as gendered. Because few women worked in the early 20th century when many American voluntary organizations were established, membership was often gendered. Men belonged to voluntary organizations with high proportions of men and women belonged to those with high proportions of women. This dynamic continues today. Even as more women enter the work force, the composition of voluntary organizations

remains gendered.

Men's preferences for voluntary organizations have not changed as much as women's. For example, the Knights of Columbus is an old charitable organization of Catholic men whose working norms are well-established. Even if membership rules changed to allow female membership, being a female member of this voluntary organization would acquire a meaning that might make it more difficult for women to choose to be members. This historical effect on strong voluntary organization norms may be enhanced because some have local chapters, and so have a strong, embedded presence in civic life (Putnam 2000:48-53). This provides a good example of the different effects that additional social contexts may exert on individual preferences for voluntary organization membership.

A third approach is for scholars to consider how the multiples are related theoretically. For instance, researchers studying multiple associations may hypothesize that one type of association is completely independent of the other, that the two are separate and additive, in which case we could account for one and control for the others, or that the combined associations acquire meaning independent of each alone. Lazer et al.'s (2010:253) study of public-policy masters' degree students, for instance, asked them to identify other students from their class "if you discussed academics (outside of the classroom) with him/her this year," an instrumental relationship, and "if you got together with him/her for non-academic reasons this year," an expressive relationship. They assessed whether political attitude similarity could be attributed to three types of dyads, those that shared an expressive association, those that shared an instrumental association and those that shared a multiplex instrumental-expressive association. They found effects only for the expressive associations. This provides a good example of the distinct effects of different types of associations. Multiple associations do not always mean more

homophily.

CONCLUSION

Homophily holds meaning. Understanding how homophily affects the way people make sense of their relationships requires comprehending why it exerts so much influence on the decisions people make, the way they behave, the way they act and the way they feel. At its core, this phenomenon seems relatively simple. Homophily theory should answer the following questions: Who do we see as similar others? Why do we associate with them? And how do we associate with them? Yet, there is much complexity here, as evidenced by the extensive study of homophily by researchers in so many disciplines and in so many ways.

This complexity results in large part because there is little agreement on homophily's empirical measurement, despite considerable agreement on its theoretical definition. We induced a typology that captures differences in how scholars measure homophily's three components: tendency, association and similarity. This typology highlighted the fundamental constraints that empirical measures place on the inferences scholars make. Sometimes the constraints fit scholars' theoretical intentions, but frequently they do not. Certainly, including all the dimensions that we identified in our typology in one study would be quite difficult. Our goal is not to criticize this existing research, but to suggest how best to move forward.

The place to begin is the simplest explanation of homophily, which is the probabilistic effect of opportunity structures. Such effects are based on straightforward mathematical calculations that do not need to account for human interaction. However, as Blau (1977:6), a proponent of structural effects notes "Whether an attribute is inborn, like sex, or acquired, like occupation, what is crucial is that it has a discernible effect on social associations. If no such association can be discovered for an attribute, the initial assumption that it has relevance for

social life must be rejected...” Thus, while these probabilities influence associations and similarities, understanding homophily still requires examining how the associations and similarities matter to the people experiencing them.

To connect measures with meaning, scholars need more precision in how they examine homophily. If measures do not capture the theories scholars intend, theoretical inferences from this research lose subtlety or are simply wrong. Borrowing from psychometric test theory (Lord and Novick 1968), homophily is a phenomenon that has a true score, a hypothetical value with no measurement error. Our empirical measures consist of this true score plus error. If we cannot distinguish between theoretical differences and measurement errors, replicating and advancing research becomes difficult. Thus, it may prove useful to explore these shortcomings, rather than ignoring them. No situation allows perfect measures, but scholars must recognize and acknowledge what is missing in their measures and what implications this “missingness” exerts on theoretical inference. In other words, when studying homophily, scholars must demonstrate that they know what they do not know.

ENDNOTES

¹ Italics added to identify the three components in each definition.

² Other concept names for opportunity structure include baseline homophily (McPherson et al. 2001), induced homophily (Kleinbaum, Stuart, and Tushman 2013, McPherson and Smith-Lovin 1987) and macrostructural homophily (Blau, Ruan, and Ardel 1991:1038). Another term for individual preferences is choice homophily.

³ For instance, McPherson et al. (2001:419) define any homophilous associations “over and above the opportunity set as inbreeding homophily.” This includes “homophily induced by social structures below the population level..., to homophily induced by other dimensions with which the focal dimension is correlated..., and to homophily induced by personal preferences. Therefore, it does not in any sense indicate choice or agency purified of structural factors.” Thus, they do not distinguish between what we are calling homophily resulting from opportunity structures and homophily resulting from individual preferences.

⁴ Web of Science has since changed their search algorithm. “Abstract” is no longer a searchable option.

Running Head: Homophily: Measures and Meaning

Table 1
Typology of Homophily Components and Dimensions

Component	Dimension	Description
TENDENCY		
Rates	Expected	What comparison identifies whether the number of similar others represents a tendency? The probabilistic comparison of the observed number of associations to an expected number of associations, where the expected number is what would be observed if associations were random within a given population.
	Relative	The comparison of the number of associations observed in one group with the number of associations observed in another.
Social Context	Primary	Within what social context or foci of activity is the association identified? The social context that defines the association type within which homophily is evaluated. For instance, a study of homophily within dyads, organizations, communities, populations or occupations.
	Secondary	An attribute's secondary social context is one of many possible in which the primary social context is embedded.
ASSOCIATION		
Relationship Type	Instrumental	What is the basis of the association? Task-related activities with another or others
	Expressive	Affective responses to another or others
	Identity	Identification with another or others
	Knowing	Knowledge of another or others
Symmetry	Asymmetric	Is association being identified by the individual or by both the individual and another or others? Individual identifies associations
	Symmetric	Both individual and target or targets identify associations
Perspective ^a	Self	Who decides this was an association? Individual identifies the associations
	Other	Researcher identifies the associations
Salience ^a	Self	Who decides this relationship type is salient to the individual? Individual says relationship is salient
	Other	Researcher or other says relationship is salient
SIMILARITY		
Attribute ^b	Status	What is being compared? Ascriptive or acquired attributes, such as demographic characteristics
	Values	Attitudes, perceptions or beliefs
Perspective ^a	Self	Who is identifying the similarity? Individual identifies similarity i.e. this other person is the same as I am
	Other	Individual provides information used by researcher to identifies similarity
	Researcher	Researcher identifies similarity
Salience ^a	Self	Who decides this attribute is salient to the individual? Researcher asks individual whether attribute is salient, or at least assesses
	Other	Researcher assumes attribute is salient to individual

^a These dimensions discussed in Q2.

^b The status and values dimensions are taken from Lazarsfeld and Merton (1954). The authors do not define values, but refer to them as having "perceptual and attitudinal components" (P. 26). We use the concept broadly to include all non-status individual attributes used in homophily research.

References

- Aguilar-Gallegos, Norman, Manrubbio Muñoz-Rodríguez, Horacio Santoyo-Cortés, Jorge Aguilar-Ávila, and Laurens Klerkx. 2015. "Information networks that generate economic value: A study on clusters of adopters of new or improved technologies and practices among oil palm growers in Mexico." *Agricultural Systems* 135:122-132. doi: 10.1016/j.agsy.2015.01.003.
- Almaatouq, Abdullah, Laura Radaelli, Alex Pentland, and Erez Shmueli. 2016. "Are You Your Friends' Friend? Poor Perception of Friendship Ties Limits the Ability to Promote Behavioral Change." *PLOS ONE* 11 (3):e0151588. doi: 10.1371/journal.pone.0151588.
- Alstott, Jeff, Stuart Madnick, and Chander Velu. 2014. "Homophily and the speed of social mobilization: The effect of acquired and ascribed traits." *Plos One* 9 (4). doi: 10.1371/journal.pone.0095140.
- Annual Review of Sociology. 2016. "Annual Reviews, Most Cited." accessed February 28, 2016. <http://www.annualreviews.org/action/showMostCitedArticles?topArticlesType=sinceInception&journalCode=soc>.
- Appold, Stephen J., Sununta Siengthai, and John D. Kasarda. 1998. "The employment of women managers and professionals in an emerging economy: Gender inequality as organizational practice." *Administrative Science Quarterly* 43 (3):538-565.
- Behrman, Jere R, Hans-Peter Kohler, and Susan Cotts Watkins. 2002. "Social networks and changes in contraceptive use over time: Evidence from a longitudinal study in rural Kenya." *Demography* 39 (4):713-738.
- Blau, Peter M. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: The Free Press.
- Blau, Peter M., Danching Ruan, and Monika Ardel. 1991. "Interpersonal choice and networks in China." *Social Forces* 69 (4):1037-1062.
- Brown, Jacqueline Johnson., and Peter H. Reingen. 1987. "Social ties and work-of-mouth referral behavior." *Journal of Consumer Research* 14 (3):350-362.
- Burt, Ronald S. 1984. "Network items and the general social survey." *Social Networks* 6 (4):293-339. doi: [http://dx.doi.org/10.1016/0378-8733\(84\)90007-8](http://dx.doi.org/10.1016/0378-8733(84)90007-8).

Byrne, Donn E., and William Griffitt. 1973. "Interpersonal attraction." *Annual Review of Psychology* 24:317-336.

Cialdini, Robert B. 2009. "We have to break up." *Perspectives on Psychological Science* 4 (1):5-6.

Dellande, Stephanie, Mary C. Gilly, and John L. Graham. 2004. "Gaining compliance and losing weight: The role of the service provider in Health Care Services." *Journal of Marketing* 68 (3):78-91.

Eisenstadt, Shmuel N. [1956] 1962. *From Generation to Generation*. New York: Free Press.

Espelage, Dorothy L., Melissa K. Holt, and Rachel R. Henkel. 2003. "Examination of peer-group contextual effects on aggression during early adolescence." *Child Development* 74 (1):205-220.

Eyal, Keren, and Alan M. Rubin. 2003. "Viewer aggression and homophily, identification, and parasocial relationships with television characters." *Journal of Broadcasting and Electronic Media* 47:77-98.

Feld, Scott L. 1981. "The focused organization of social ties." *American Journal of Sociology* 86:1015-1035.

Ferber, P. H., and R. R. Pugliese. 2000. "Partisans, proximates, and poker players: The impact of homophily and proximity on communication patterns of state legislators." *Polity* 32:401-141.

Fischer, Claude S. 1982. *To Dwell Among Friends: Personal Networks in Town and City*. Chicago: University of Chicago.

Fischer, Eileen, Brenda Gainer, and Julia Bristor. 1997. "The sex of the service provider: Does it influence perceptions of service quality?" *Journal of Retailing* 1997 (73):3.

Flynn, Francis J., Ray E. Reagans, and Lucia Guillory. 2010. "Do you two know each other? Transitivity, homophily, and the need for (network) closure." *Journal of Personality & Social Psychology* 99 (5):855-869.

General Social Survey. 1972 to present. "The General Social Survey (GSS) is a project of the independent research organization NORC at the University of Chicago, with principal funding from the National Science Foundation."

<http://www.norc.org/Research/Projects/Pages/general-social-survey.aspx>.

Gower, J.C., and Pierre Legendre. 1986. "Metric and Euclidean properties of dissimilarity coefficients." *Journal of Classification* 3:5-48.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The weirdest people in the world?" *Behavioral and Brain Sciences* 33:61-135.

Hirsch, Ben T., Suzanne Prange, Stephanie A. Hauver, and Stanley D. Gehrt. 2013. "Genetic relatedness does not predict racoon social network structure." *Animal Behaviour* 85 (2):463-470. doi: 10.1016/j.anbehav.2012.12.011.

Huston, Ted L., and George Levinger. 1978. "Interpersonal attraction and relationships." *Annual Review of Psychology* 29:115-156.

Ibarra, Herminia. 1992. "Homophily and differential returns: Sex differences in network structure and access in an advertising firm." *Administrative Science Quarterly* 37 (3):422-447. doi: 10.2307/2393451.

Ibarra, Herminia. 1993. "Personal networks of women and minorities in management: A conceptual framework." *Academy of Management Review* 18 (1):56-87.

Ibarra, Herminia, Martin Kilduff, and Wenpin Tsai. 2005. "Zooming in and out: Connecting individuals and collectivities at the frontiers of organizational network research." *Organization Science* 16 (4):359-371.

Kandel, Denise Bystryn. 1966. "Status homophily, social context, and participation in psychotherapy." *American Journal of Sociology* 71:640-650.

Kanter, Rosabeth M. 1977. *Men and Women of the Corporation*. New York: Basic Books.

Kanter, Rosabeth Moss. 1976. "The policy issues: Presentation VI." *Signs: Journal of Women in Culture and Society* 1 (Spring, part 2):282-291, as cited in Kanter (1977).

- Kleinbaum, Adam M., Toby E. Stuart, and Michael L. Tushman. 2013. "Discretion within constraint: Homophily and structure in a formal organization." *Organization Science* 24 (5):1316-1336. doi: 10.1287/orsc.1120.0804.
- Kossinets, Gueorgi, and Duncan J. Watts. 2009. "Origins of homophily in an evolving social network." *American Journal of Sociology* 115 (2):405-450.
- Lawrence, Barbara S. 1987. "An organizational theory of age effects." *Research in the Sociology of Organizations* 5:35-71.
- Lawrence, Barbara S. 2006. "Organizational reference groups: A missing perspective on social context." *Organization Science* 17 (1):80-100.
- Lazarsfeld, Paul F., and Robert K. Merton. 1954. "Friendship as a social process: A substantive and methodological analysis." In *Freedom and Control in Modern Society*, edited by Morroe Berger, Theodore Abel and Charles H. Page, 18-66. New York: Van Nostrand.
- Lazer, David, Brian Rubineau, Carol Chetkovich, Nancy Katz, and Michael Neblo. 2010. "The coevolution of networks and political attitudes." *Political Communication* 27 (3):248-274.
- Lord, F.M., and M.R. Novick. 1968. *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Martin, Carol Lynn, Richard A. Fabes, Laura D. Hanish, and Tom Hollenstein. 2005. "Social dynamics in preschool." *Developmental Review* 25:299-327.
- McCroskey, James C., Virginia P. Richmond, and John A. Daly. 1975. "The development of a measure of perceived homophily in interpersonal communication." *Human Communication Research* 1 (4):323-332, as cited in Simpson et al. 2000, p. 882.
- McCroskey, Linda L., James C. McCroskey, and Virginia P. Richmond. 2006. "Analysis and improvement of the measurement of interpersonal attraction and homophily." *Communication Quarterly* 54 (1):1-31.
- 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 (3):370-379.

Running Head: Homophily: Measures and Meaning

- McPherson, J. Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual Review of Sociology* 27:415-444.
- Merton, Robert K. [1949] 1968. *Social Theory and Social Structure (1968 Enlarged Edition)*. 1968 Enlarged Edition ed. New York: The Free Press, Macmillan Publishing Co. Original edition, 1949.
- Miles, Matthew B., and A. Michael Huberman. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. Second ed. Thousand Oaks, CA: Sage Publications.
- Mollica, Kelly A., Barbara Gray, and Linda K. Trevino. 2003. "Racial homophily and its persistence in newcomers' social networks." *Organization Science* 14:123-136. doi: 10.1287/orsc.14.2.123.14994.
- Moore, Wilbert. 1962. *The Conduct of the Corporation*. New York: Random House, as cited in Rivera 2013.
- Moreno, Jacob L. 2011a. *The Autobiography of J.L. Moreno M.D. (Abridged)*. United Kingdom: The North-West Psychodrama Association.
- Moreno, Jacob L. 2011b. *Who Shall Survive? A New Approach to the Problem of Human Interrelations*. United Kingdom: Northwest Psychodrama Association. Original edition, 1934.
- Neugarten, Bernice L., J.W. Moore, and J.C. Lowe. 1965. "Age norms, age constraints, and adult socialization." *American Journal of Sociology* 770:710-717.
- Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.
- Randel, Amy E., and P. Christopher Earley. 2009. "Organizational Culture and Similarity Among Team Members' Salience of Multiple Diversity Characteristics." *Journal of Applied Social Psychology* 39 (4):804-833. doi: 10.1111/j.1559-1816.2009.00461.x.
- Rivera, Lauren A. 2013. "Homosocial reproduction." In *Sociology of Work: An Encyclopedia*, edited by Vicki V. Smith, 376-378.

Rohan, Meg J. 2000. "A rose by any name? The values construct." *Personality and Social Psychology Review* 4 (3):255-277. doi: 10.1207/S15327957PSPR0403_4.

Rotolo, Thomas, and Amy S. Wharton. 2003. "Living across institutions: Exploring sex-based homophily in occupations and voluntary groups." *Sociological Perspectives* 46:59-82.

Schneider, Benjamin. 1987. "The people make the place." *Personnel Psychology* 40 (437-453).

Schneider, Benjamin, Harold W. Goldstein, and D. Brent Smith. 1995. "The ASA framework: An update." *Personnel Psychology* 48:747-773.

Simpson, Eithel M., Thelma Snuggs, Tim Christiansen, and Kelli E. Simples. 2000. "Race, homophily, and purchase intentions and the black consumer." *Psychology and Marketing* 17:877-889.

Stackman, Richard W., and Craig C. Pinder. 1999. "Context and sex effects on personal work networks." *Journal of Social and Personal Relationships* 16 (1):39-64.

Stark, Tobias H., and Andreas Flache. 2012. "The Double Edge of Common Interest: Ethnic Segregation as an Unintended Byproduct of Opinion Homophily." *Sociology of Education* 85 (2):179-199. doi: 10.1177/0038040711427314.

Tajfel, Henri. 1978. *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations*. London: Academic Press, as cited in Ramarajan 2014, p. 593.

Verbrugge, Lois M. 1977. "The structure of adult friendship choices." *Social Forces* 56:576-597.

Whittler, Tommy E. 1989. "Viewers' processing of actor's race and message claims in advertising stimuli." *Psychology and Marketing* 6:287-315, as cited in Simpson et al. 2000, p. 882.

Yamaguchi, Kazuo, and Denise B. Kandel. 1993. "Marital homophily on illicit drug use among young adults: Assortative mating or marital influence?" *Social Forces* 72 (2):505-528.