GROUP SOLIDARITY AS THE PRODUCT OF COLLECTIVE ACTION: CREATION OF SOLIDARITY IN A POPULATION OF INJECTION DRUG USERS

Douglas D. Heckathorn and Judith E. Rosenstein

ABSTRACT

This paper introduces a theory of group solidarity and a method for measuring it. Solidary groups are characterized by strong internal monitoring and sanctioning systems, strong intra-group ties, high exit costs, and lack of information about resources outside the group. This analysis suggests that all these attributes derive from the choice to invest differentially in social relationships within the group rather than forming cross-cutting ties. To explain variations in solidarity across groups thus requires an account of the conditions that favor intra-group ties. Drawing on a formal theory of collective action, the analysis shows how the return from investments in intra-group ties varies based on the shape of the production function for the collective goods produced by the group. The proposed measure of group solidarity is based on the degree to which the proportion of intra-group ties exceeds that which would be expected

Group Cohesion, Trust and Solidarity, Volume 19, pages 37–66.

ISBN: 0-7623-0898-2

were ties formed randomly. The theory of group solidarity and the measurement procedures are illustrated using data from a survey of 488 injection drug users in Connecticut.

INTRODUCTION

Solidarity is one of the venerable social science concepts for which there are almost as many definitions as theorists who have analyzed it. Following Homans (1950), Fararo and Doreian (1998, p. 16) distinguish four distinct ways to analyze solidarity. Analyses can focus on the affective bonds that unite members of solidary groups; the norms defining group obligations; the collectively oriented activity patterns characteristic of these groups, especially a preponderance of prosocial behavior; and the interaction patterns in which ties within the group are denser than ties across groups.

Diverse approaches to analyzing solidarity have flourished because any of these four elements can be taken as the fundamental characteristic of solidarity, and thereby be used to explain the other elements. For instance, Widegren (1997) views solidarity as deriving from altruistic sentiments directed toward those who reciprocate those sentiments. Altruism then is invoked to explain prosocial activity, strong group norms, and dense intra-group ties. In contrast, Lawler and Yoon (1998) view positive sentiments not as fundamental but as deriving from repeated interactions, especially mutually satisfactory interactions among status equals. Given that all four elements of solidarity generally occur together, analysts face a problem analogous to multicolinearity in regression analysis, when highly correlated terms make it difficult to disentangle causal linkages.

This paper has two purposes. The first is to propose an operational definition of solidarity that is applicable in contexts where corporate obligations lack clarity and the resources allocated on their behalf do not produce a documentary record. The second is to propose a theoretic account of the factors governing choice of intra-group vs. out-group ties. The aim is to extend Hechter's theoretic account of solidarity, which focuses on structural attributes of solidary groups, with an analysis of the conditions under which these attributes will be created as products of collective action.

Part I reviews Hechter's theory of group solidarity, extending his analysis of the structural determinants of solidarity to identify the form of collective action by which these structures are produced. Part II introduces a method for measuring solidarity, based on the argument that the strength of preference for intra-group ties, homophily, provides a quantitative way to measure solidarity that is applicable in a wide range of structured and non-structured settings.

Part III proposes a model to explain the relative preference for intra-group vs. out-group ties, which is based on an expansion of Heckathorn's (1988, 1989, 1990, 1993, 1996) theory of collective action. As in Hechter's analysis, the model focuses on the nature of the collective goods produced by solidary groups. The analysis differs by shifting the focus from whether the goods are fungible or imminent, to the forms of social dilemmas that arise when goods are produced within the group. The analyses are illustrated using a study of variations in group solidarity among a total of 488 injection drug users from three Connecticut cities, Meriden, Middletown, and New London. The conclusion discusses methods by which group solidarity can be increased and describes how those methods have been employed in the design of AIDS-prevention interventions.

I: HECHTER'S THEORY OF GROUP SOLIDARITY

An important step toward a coherent and comprehensive analysis of solidarity is Hechter's (1987) extension of dependence theory to solidarity. Noting that some analysts discuss solidarity without offering a definition, and others use definitions that cannot be operationalized, one of Hechter's aims was to define solidarity in a manner that would make the concept measurable. A second aim was to offer a general theoretic model that could be used to explain variations in solidarity across groups and over time.

Hechter (1987) defines group solidarity as the proportion of member resources employed to fulfill corporate obligations. Solidarity therefore depends on two factors, the rate of compliance with corporate obligations and the extensiveness of these obligations. His illustrations of how this definition makes solidarity measurable include a study of party solidarity in a legislature with roll-call votes. The corporate obligation involves legislators voting in the manner supported by party leaders. The extensiveness (E) of the group's corporate obligations is the proportion of votes on which party leaders take a position, and compliance (C) with those obligations involves party members voting consistently with that position. In this case, solidarity (S) refers to the proportion of votes consistent with party positions (i.e. $S = E * C$).[1] In a legislature, the essential resource is the ability to vote. In other groups, the relevant resource involves time spent on group rather than individual activities.

Hechter's substantive theory focuses on the factors that determine compliance with and extensiveness of corporate obligations. Compliance, he argues, depends on the sanctioning capacity of the group and the group's monitoring capacity. Thus, when violations of corporate obligations are detectable and can be effectively sanctioned, compliance is high. This is consistent with Hechter's view of behavior as incentive driven. However, it ignores research demonstrating the

paradoxical effects of incentives. For example, overly severe sanctions can backfire. For this reason, Ostrom (1990) emphasizes the importance of graduated sanctions, sanctions that are both proportionate to the offense and consistent with group norms, and Heckathorn (1990) shows that overly severe sanctions can provoke resistance rather than compliance. Similarly, research on the "hidden cost of reward" (Deci & Ryan, 1985) shows that rewards can weaken intrinsic motivation. The implication is that, paradoxically, strengthening sanctions or improving monitoring can reduce rather than increase compliance with corporate obligations.

Hechter's analysis of the extensiveness of corporate obligations seeks to identify the upper and lower bounds within which this may vary. The upper bound, he argues, is determined by the actor's dependence on the group. This determines the maximum demands the group can place upon the individual. Dependence is based on four factors: (1) lack of close substitutes outside the group for resources provided within the group; (2) lack of information about alternatives outside the group; (3) cost of exiting the group; and (4) commitments to personal ties within the group. Thus, the greater the individual's dependence upon the group, the greater may be the corporate demands on the individual. Corporate obligations may be less extensive than the upper limit established by dependence, but they cannot be greater, otherwise an individual would have incentives to exit the group, abandon within-group ties, and seek alternatives for that which the group provided. Therefore, dependence defines the upper bound of corporate obligations.

The lower bound for corporate obligation is determined by the cost of the joint activities undertaken by the group (Hechter, 1987). Given that solidarity refers to the extent to which individual resources are used in a manner consistent with corporate dictates, when solidarity increases, so too must the individual resources employed to fulfill corporate obligations. This follows directly from the definition of solidarity and from Hechter's assumption that the resources required to fulfill corporate obligations come from group members.

Hechter's theoretic contribution includes a measurable definition of solidarity and a consistent theoretic account of factors upon which solidarity depends. Two remaining problems provide the opportunity for further theorizing. One problem concerns the measurability of solidarity. When applied to a legislature, as in the above analysis, the measure is clear as long as one knows individual votes, party affiliations, and the positions of party leaders. In less structured contexts, measurement becomes less straightforward because corporate obligations lack clarity, and the resources allocated to fulfill them – such as time spent on alternative activities – do not leave a paper trail. This problem is apparent in occupational studies of time allocation. The variance in time spent

on occupational pursuits is highest for university faculty members, ranging from as little as a dozen to more than 60 hours per week. This variation results, in part, from ambiguity in the definition of academic work. If it is restricted to student contact time and structured research activities, investments of time are far lower than if the definition is expanded to include thinking about writing projects when in the shower, when preparing to sleep, and the like. Similarly, in family and peer groups, corporate obligations include mutual assistance, participating in joint recreational activities, and avoiding activities harmful to the group. For a compliant individual, satisfying the latter obligation could consume the entirety of available time. One aim of this paper is to propose an operational definition of solidarity that is applicable in contexts where corporate obligations lack clarity and the resources allocated on their behalf do not produce a documentary record.

The second problem – and the second area of potential further theoretic development – is a failure to specify how the structural features upon which solidarity depends are themselves created. According to Hechter, solidary groups are characterized by extensive systems for monitoring and sanctioning, which provide the tools by which the behavior of group members can be controlled. The question that remains unresolved is, Why would group members invest resources to create extensive systems of monitoring and sanctioning? The factors leading to this investment decision (the proximate causes of group solidarity) can then provide the basis for explaining the intervening causes of solidarity – that is, the enforcement resources that secure group control of members' behavior and restrict information about alternatives outside the group.

When the determinants of solidarity are examined, they can all be seen as deriving from a single proximate (i.e. originating) cause. Consider again the first determinant of solidarity, compliance with corporate obligations. In a solidary group, intra-group ties are denser than out-group ties, so group members necessarily have greater abilities to monitor one another's performance than do out-group members. Similarly, owing to the density of intra-group ties, group members have enhanced abilities to mediate rewards and punishments, thereby strengthening their sanctioning capacity. Therefore, greater compliance can be seen as resulting from a structural feature of solidarity systems, their greater density of intra-group ties. The question then becomes, under what conditions do group members invest differentially in intra-group ties?

Dependence can also be seen as a product of differential investments in intra-group ties. Consider, for example, the determinants of dependence identified by Hechter. In a solidary group, intra-group ties are denser than out-group ties, so individuals tend to have less information about substitutes for their intra-group ties and less information about alternative ties, and the

costs of exiting the group are greater because more ties would need to be reestablished. Furthermore, intra-group ties tend to be stronger than the out-group ties with which they could be replaced because newly formed ties tend to be weak and gain strength only over time if interactions prove rewarding. Therefore, all four determinants of dependence depend directly on the greater investment in intra-group ties in solidary groups.

Once again, the question arises, What motivates the preference for intra-group over out-group ties? Hechter provides a partial answer to this question in his discussion of the differences between groups producing fungible (salable) goods vs. imminent goods, such as participating in a ritual or enjoying the company of friends, which are consumed instantaneously with their production. The solidarity of groups producing fungible goods tends to be lower because the output can be accumulated and used for transactions outside the group, thus encouraging interactions with non-group members. In contrast, imminent goods can neither be hoarded nor used in extra-group exchanges. Yet imminent goods are produced in all social systems because whether solidarity is high or low, individuals enjoy one another's company and engage in the myriad other forms of interactions in which positive affect is exchanged. Thus, dense intra-group networks are not a prerequisite to the production of imminent goods, and some high-solidarity groups (e.g. charismatically organized religious sects) produce fungible goods.

In sum, monitoring and sanctioning capacities, and dependence on the group, are all products of collective action – products created when group members invest differentially in intra-group rather than out-group ties. Therefore, a fundamental issue in the explanation of solidarity is identification of the factors that motivate individuals to invest differentially in intra-group rather than out-group ties. Some of the factors relevant to this question are revealed in Burt's (1998) analysis of mobility patterns within a large investment banking firm. He found that the optimal network configuration varied based on duration of employment, ethnicity, and gender. For white males employed more than one year, advancement occurred more rapidly if they cultivated what Burt termed "entrepreneurial networks." These are networks that avoid transitive closure. That is, cultivating far-flung relationships rather than cultivating relationships with the friends of friends (an association that would produce transitive systems of social relationships, because A knowing B, and B knowing C, would lead to A knowing C). The optimal network structure for these men involved associations with many individuals who were socially distant from one another, many of whom did not even know one another. This structure optimized information gathering by filling what he termed "structural holes," in which poorly connected networks provide opportunities for brokers to gain from

facilitating the flow of information. In contrast, for women, minorities, and white men employed for less than one year, advancement was most rapid if they cultivated a dense network structure, a structure consistent with the relatively closed character of solidary groups. Burt interpreted this difference as resulting from variations in trust. Initially constructing a dense network ensured that those monitoring the employee's performance could readily share information. For white males who performed well, this led to acceptance as trusted organizational members, and they could then fashion entrepreneurial networks. In contrast, minorities and women were blocked from taking this additional step. What is relevant here is that a variety of situational factors may affect the relative preference for intra-group vs. out-group ties. This is consistent with the observation that groups vary greatly in solidarity, reflecting heavy investments in intra-group ties in some contexts and minor investments in others.

II: CONCEPTUALIZING SOLIDARITY

If, as was argued above, the mechanisms underlying solidarity derive from disproportionate investments in intra-group over out-group ties, then a conceptualization based on the distribution of ties may provide the basis for a quantitative measure of solidarity. Much prior research relevant to such an approach has been done. Building on Simmel's (1955) approach to analyzing social structure in terms of affiliation patterns, several formal models have been proposed (Fararo & Sunshine, 1964; Blau, 1977, 1994; Rapoport, 1979), all of which are based on the idea that if social system lacked structure, ties would be formed randomly. Each group's ties would therefore reflect an identical pattern, in which the number of ties directed to a specific group would be exactly proportional to that group's size. Of course, there are no social systems in which ties are formed without respect to gender, age, ethnicity, social status, and myriad other factors; hence the universality of social structure.

One form of departure from randomness has received special attention in the literature on affiliation patterns. As Galton recognized more than a century ago, affiliations tend to form among those who are similar in age, education, prestige, social class, and race and ethnicity (McPherson & Smith-Lovin, 1987). This bias toward self-affiliation is termed *homophily,* and it is a fundamental element of social structure. In Durkheim's (1956) terms, this corresponds to *mechanical solidarity*, in which solidarity is based on similarity. To the extent that solidarity derives from differential investments in intra-group rather than out-group ties, then homophily provides the network-structural basis for solidarity and implies that homophily can provide a structurally based indicator of solidarity.

Fararo and Sunshine's (1964) "inbreeding bias" parameter provides the basis for a quantitative measure of homophily (Heckathorn, 1997). Perfect homophily, in which all ties are formed within the group, is assigned the value + 1; and no homophily, in which ties are formed without regard to group membership, is assigned the value zero. When the level of homophily is intermediate, such as 1/3, ties are formed one third of the time within the group, and two thirds of the time randomly, without regard to group membership.

Tie formation need not favor the in-group. It can also favor the out-group, as in social systems with exogamous marriage norms, competitive career tracks in which upward mobility depends on the ability to cultivate ties to those of higher status, and musical groups in which different types of instruments are represented. Therefore, negative homophily, termed heterophily, is also possible. This corresponds to Durkheim's (1956) concept of *organic solidarity*, in which solidarity is based on interdependence. Encompassing this possibility requires a straightforward expansion of the definition of homophily (Heckathorn, 2002). Perfect heterophily exists when no intra-group ties are formed, so homophily is –1. Intermediate levels of negative homophily can be defined in a similar way; for example, if homophily is –1/4, then 25% of the time ties are formed with out-group members, and the other 75% of the time ties are formed randomly, irrespective of intra- or out-group membership. As defined in this way, homophily is negative if and only if the proportion of ties that are internal to the group is less than the proportional size of the group. More formally, where P_a is the proportional size of group A, and S_{aa} is the proportion of A's ties that are within the group, the homophily of group A homophily, H_a, can be defined as:

$$H_a = \frac{P_a - S_{aa}}{(P_a - 1)} \text{ if } S_{aa} > P_a$$
$$H_a = \frac{S_{aa} - P_a}{P_a} \text{ if } P_a > S_{aa} \qquad (1)$$

For example, surveys of 190 injection drug users in Meriden, Connecticut (see Table IA) showed that the proportion of intra-group ties among Hispanic injectors was substantially larger (0.452) than this group's estimated size (0.198) so homophily was substantial $(0.198 - 0.452)/(0.198 - 1) = 0.317$. Thus, these injectors created networks as though 31.7% of the time, they formed ties to other Hispanic injectors, and the other 68.3% of the time, they formed ties

randomly, irrespective of ethnicity. Homophily was also substantial for non-Hispanic black injectors (35.7%), and non-Hispanic whites (36.2%). Only the small group in the "other" category, consisting of Native and Asian Americans, was not homophilous. When affiliation by gender was examined a different pattern was found (see Table IB). The estimated 41% of female injectors formed within-gender ties 42% of the time, so homophily was near zero (1.8%). In contrast, the estimated 59% of male injectors formed within-gender ties 73.6% of the time, so homophily was substantial (35.5%). This difference may reflect the relative status of males in injector populations.

As thus conceived, positive and negative homophily can coexist in the same system. This can occur when a division of labor emergences in a group of otherwise similar individuals. For example, jazz musicians perform in groups that bring together performers with similar musical preferences, such as a preference for classical jazz, contemporary jazz, or fusion. But groups also include performers specializing in diverse types of instruments, so homophily coexists with heterophily (Heckathorn & Jeffri, 2001). The implication is that the measure for homophily and heterophily are not rivals, because in complex systems patterns of differential association generally include affiliation based on both similarity and on difference. Therefore, the appropriate indicator for solidarity is the absolute value of the homophily measure, because positive and negative values reflect solidarity based on similarity vs. solidarity based on interdependence.

This approach to measuring homophily in terms of network structure resembles Markovsky and Lawler's (1994) proposal that solidarity be conceptualized in terms of the network property of reachability, which is defined as the number of links required to go from any group member to any other group member. Reachability increases, as does homophily, when intra-group ties are substituted for out-group ties, so the two measures are positively related. An advantage of the latter is that it facilitates comparisons among groups of different sizes because homophily depends only on the proportion of in-group vs. out-group ties.

Homophily provides only an indirect measure when solidarity is conceived in terms of resource allocation. In this respect, it resembles other measures of solidarity. For example, in the legislature analyzed by Hechter (1987, p. 79), solidarity was measured by counting votes consistent with party dictates. This provides an indirect measure of resource allocation, because from the standpoint of party leaders, not all issues are of equal value: votes on bills viewed as crucial by party leaders and votes that provide the party with the margin needed for victory, for example, have high value. When that margin has been fulfilled, party members whose votes are not needed may be released

Table 1. Recruitment by Race/Ethnicity and Gender.

Table 1A: Race and Ethnicity (Recruitment Count; Selection Proportion, S)

Race and Ethnicity of Person who Recruited	Race and Ethnicity of Recruit				
	White	Hispanic	Black	Other	Total
Non-Hispanic White	**102.000**	**10.000**	**8.000**	**6.000**	126.000
	0.810	**0.079**	**0.063**	**0.048**	1.000
Hispanic	**18.000**	**19.000**	**4.000**	**1.000**	42.000
	0.429	**0.452**	**0.095**	**0.024**	1.000
Non-Hispanic Black	**7.000**	**2.000**	**5.000**	**0.000**	14.000
	0.500	**0.143**	**0.357**	**0.000**	1.000
Other	**3.000**	**5.000**	**0.000**	**0.000**	8.000
	0.375	**0.625**	**0.000**	**0.000**	1.000
Total Distribution of Recruits,	130.000	36.000	17.000	7.000	190.000
Sample Distribution, SD	0.684	0.189	0.089	0.037	1.000
Equilibrium, E	0.700	0.169	0.094	0.037	
Mean Network Size, N	55.200	38.400	63.300	76.700	
Homophily, H	0.372	0.317	0.301	−1.000	
Population Estimate, P	0.702	0.198	0.080	0.020	

Table 1B: Gender (Recruitment Count; Selection Proportion, S)

Gender of Recruiter	Gender of Recruit		
	Female	Male	Total
Female	**21.000**	**29.000**	50.000
	0.420	**0.580**	1.000
Male	**37.000**	**103.000**	140.000
	0.264	**0.736**	1.000
Total Distribution of Recruits	58.000	132.000	190.000
Sample Distribution, SD	0.305	0.695	1.000
Equilibrium, E	0.313	0.687	
Mean Network Size, N	37.500	57.100	
Homophily, H	0.018	0.355	
Population Estimate, P	0.410	0.590	

from party obligations. When supporting the party line would entail a political cost, legislators who are facing strong electoral opposition may be asked to support the party position only when that is essential for the party to prevail. For this reason, controversial bills, such as tax increases, generally pass by only small margins. Because of these complexities in the voting process, merely counting the proportion of a legislator's votes that are consistent with party

dictates provides an imperfect measure of the extent to which the resources possessed by the legislator fulfill corporate obligations. However, because the right to vote is a legislator's principal resource, counting votes remains the most feasible measure of resource allocation. Similarly, the ability to create and dissolve network ties in less structured groups is not the only resource available to individuals; however, these choices about ties determine an individual's dependence upon the group and the enforcement resources available to the group. Thus, these choices serve as a reasonable measure of solidarity, since each time an individual creates or dissolves a network tie, he or she thereby votes to increase or decrease the group's solidarity.

Another advantage of conceptualizing solidarity in terms of homophily is that analysis can be extended beyond formally constituted groups to any category by which individuals can be differentiated. It thereby provides a way of determining what attributes are socially significant. An example of a socially irrelevant differentiation is whether one is born in an odd or even month. No group of which we are aware considers this socially significant. Given that this factor has no effect on tie formation, social systems lack structure with respect to this factor. That is, homophily is zero. In contrast, a host of other factors affect tie formation, such as the demographic factors upon which much social science analysis depends, including age, gender, race, ethnicity, and social status.

A consideration in using homophily as a measure of solidarity is that homophily can result from multiple processes. For example, McPherson and Smith-Lovin (1987) distinguish between choice homophily and induced homophily; the former arises when choices among alternative ties are based on similarity, and latter results when opportunities to form out-group ties are limited, as when those living in ethnically homogeneous neighborhoods have no choice but to form intra-ethnic ties. Similarly, Yamaguchi and Kandel (1993) distinguish among homophily based on selection (i.e. choice homophily), homophily based on socialization in which affiliated persons influence one another in ways that make them become more similar, and homophily resulting from common responses to shared social influences. As these examples illustrate, homophily can result from multiple processes. Irrespective of its origin, greater homophily reflects greater investment of ties among persons of the same status, so dependence and enforcement resources are increased. Consider again the analogy with solidarity as measured by counting votes. Many processes can affect voting decisions. A principled legislator may vote his or her conscience, and conformity with party dictates depends on the extent to which the two converge. Alternatively, a pragmatic legislator may respond to compliance incentives from the party when deciding upon a vote. Campaign

promises and past voting practices also can affect voting decisions. Therefore, a vote consistent with party obligations can result from many considerations, some involving the control capabilities of the party, and others for which party power is irrelevant. Nonetheless, counting votes remains a plausible measure of solidarity. But irrespective of the reasons why intra-group ties are formed, they nonetheless can validly be seen as contributing to group solidarity, just as votes consistent with party obligations strengthen the party's position, irrespective of the complex and frequently conflicting motivations that govern the votes of party members.

The solidarity literature has focused on groups, which, in Hechter's definition as in most standard definitions, include clear boundaries defined by criteria for membership. Yet, solidary processes can occur in collectivities that are more diffuse. For example, in a study of attitudes and affiliation patterns among university undergraduates in three large Midwestern public universities, preferred forms of recreation structured interactions. Three forms were prominent, one oriented toward heavy alcohol use, as defined by an orientation toward "getting drunk at parties"; a second oriented toward soft drugs, particularly marijuana; and a third oriented toward non-drug related forms of recreation including studying together (Heckathorn & Lucas, 1982). The boundaries among these sets of students were too diffuse to include membership criteria, yet each group was in many respects distinct. Thus, solidary processes can occur among sets of individuals even though they do not reach the level of organization required to count as a group in the standard restrictive sense. Racial and ethnic groups provide another example. Although these categories are frequently important determinants of affiliation patterns, they generally do not constitute groups in the restrictive sense. Similarly, the greater the extent to which intra-ethnic affiliations predominate over cross-ethnic affiliations, the greater is the indication that some form of solidary process is reflected in these affiliation patterns, such as the presence of a cohesive ethnic subculture.

Measuring Homophily Using Respondent-Driven Sampling

As defined by Eq. 1, measuring homophily requires information on affiliation patterns and information on group size. Both types of information are readily available for a small number of variables, such as census data on inter-marriage by race and ethnicity (e.g. see Blau, 1977). Gathering such data in other settings, such as McPherson and Smith-Lovin's (1987) study of voluntary associations, requires substantial resources. Therefore, studies of homophily have been characterized by McPherson (1998) as "a theory starved for data." This changed with the advent of a new means for gathering information on

affiliation patterns and group size, termed respondent-driven sampling (RDS) (Heckathorn, 1997, 2002).

RDS, a form of chain-referral sampling, was developed to improve methods for sampling hard-to-reach populations, such as injection drug users, by overcoming the biases traditionally associated with chain-referral methods, including volunteerism, masking, and oversampling of groups with large networks (see Heckathorn, 1997). RDS has also served as a mechanism for recruiting active injection drug users for HIV prevention education and services (Broadhead et al., 1998; Heckathorn et al., 1999). When it plays that role, RDS is termed *participant-driven recruitment* (PDR). This article suggests that RDS can also serve as a means for studying affiliation patterns.

A comprehensive discussion of RDS would exceed the scope of this paper (see Heckathorn, 2002); however, essential features of the method can be summarized. RDS employs recruitment incentives to produce a highly robust system of chain referrals, such that over the course of more than a dozen recruitment waves, a single initial subject can yield hundreds of additional subjects (see Fig. 1). Long referral chains ensure that the sample will have considerable sociodemographic depth and that initial subjects will constitute only a minor portion of the total sample. A counterintuitive feature of this sampling procedure is the demonstration that as the sample expands from wave to wave, the effects of the choice of initial subjects become progressively weaker and the sample composition ultimately attains an equilibrium that is independent of the starting point.

The manner in which this occurs is illustrated in Fig. 2, which depicts the results of two simulations showing how the composition of each wave would have changed had recruitment begun from either one or more Hispanic intravenous drug users Since this comes up infrequently, you can dispense with the abbreviation if you like. (Fig. 2A) or one or more non-Hispanic white intravenous drug users (Fig. 2B), based on projections from Table IA's recruitment patterns. The vertical axes represent the percentage of each type, and the horizontal axes represent the number of recruitment waves, where wave 0 refers to the seed or seeds, which in this exercise were assumed to be ethnically homogeneous. Wave 1 refers to the seeds' recruits, wave 2 refers to the recruits' recruits, and so forth. If recruitment had begun with only Hispanic seeds, the percentage of Hispanics in each wave would have decreased from the initial value of 100%, to 45% in the first wave and 27% in the second wave, and eventually stabilized at 17% after five waves. This equilibrium does not change with later waves. The implication is that when referral chains are long enough for equilibrium to be attained, RDS is a reliable sampling method.

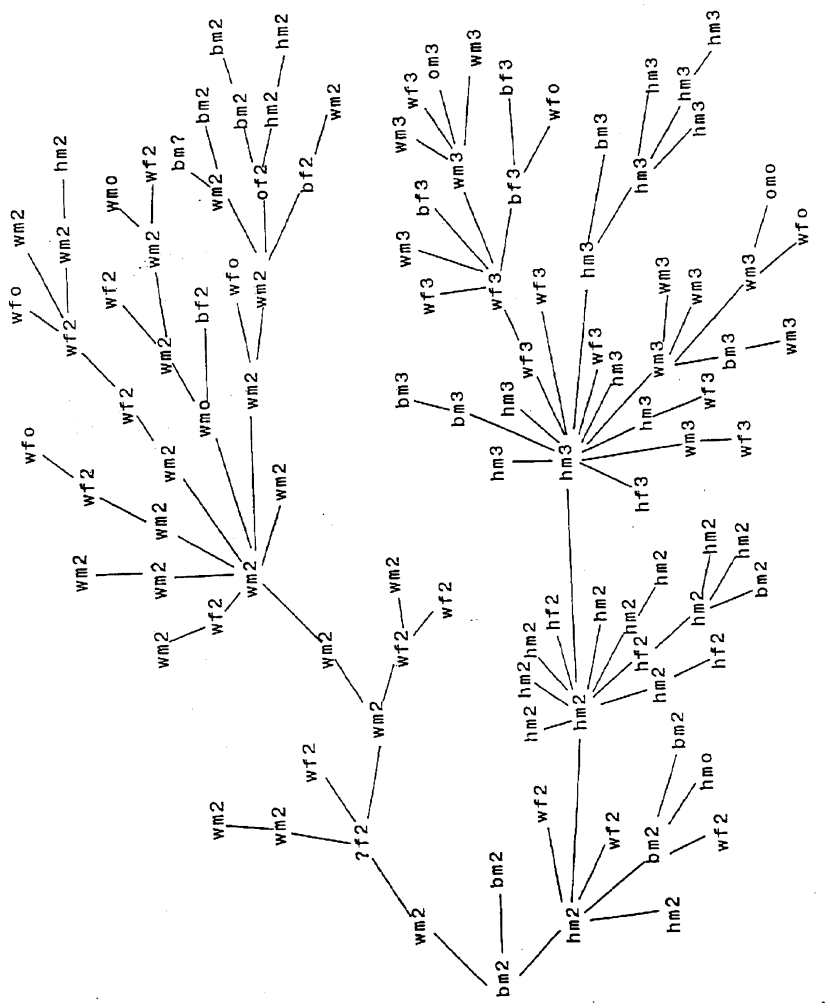

Fig. 1. Peer Recruitment Network Beginning from a Single Seed.

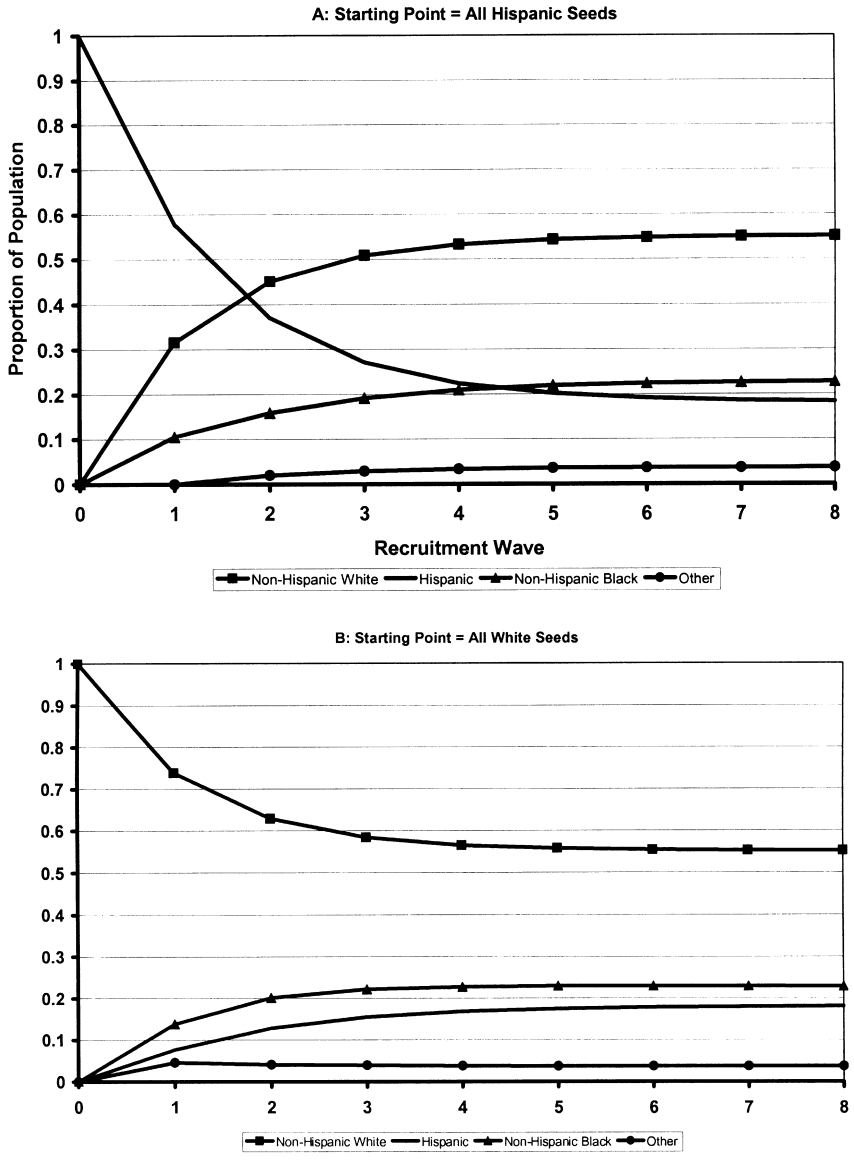

Fig. 2. Race and Ethnicity of Recruits in a Respondent-Driven Sample, Beginning with Only White or Hispanic Sets.

In RDS, information on affiliation patterns is derived from a recruitment process in which respondents recruit those with whom they have a pre-existing relationship, and the link between recruiter and recruit is established using recruitment coupons, each of which has a unique serial number. This provides a behavioral measure of affiliation patterns, a measure that correlates strongly with self-reported network composition (see Heckathorn et al., in press). Using a behavioral measure of network affiliation has an advantage, for when self-reports are used as the measure of affiliation, analysis is limited to matters of public knowledge, such as race, ethnicity, and gender. In contrast, when the measure of affiliation is behavioral, patterns can be identified, even those of which the respondents are unaware. For example, even though injectors generally do not share information about their HIV status, affiliation by HIV status was studied using the RDS method. In this study, HIV status was determined through an optional HIV test offered to respondents following their interviews. Based on the test results, affiliations were determined behaviorally, by who recruited whom (Heckathorn et al., 1999). The results showed that homophily by HIV status was near zero, indicating that the HIV infected individuals are thoroughly dissolved into the larger population of injection drug users.

III: SOLIDARITY AS THE PRODUCT OF COLLECTIVE ACTION

Insightful analyses of alternative types of homophily have been proposed (McPherson & Smith-Lovin, 1987; Yamaguchi & Kandel, 1993), yet they have not led to an explanatory theory of homophily. A step toward a theory of homophily derives from the recognition that homophily is a form of collective action. The formation of a reciprocal relationship requires mutual consent, so it requires action by at least two individuals. For example, choice and selection homophily are based on the decision of with whom to build ties, whereas socialization and common response homophily depend on the choice to maintain ties. Thus, homophily is ultimately based on choices to create or dissolve ties. An apparent exception is induced homophily, in which ties reflect opportunities for contact; however, recall that homophily refers not to a preponderance of intra-group ties but rather to a greater proportion of such ties than would be produced by random affiliation. Therefore, large groups will possess many intra-group affiliations even in the absence of homophily (i.e. a number in exact proportion to the group's size). The situation with induced homophily is more complex in systems with multiple groups, as in McPherson and Smith-Lovin's (1987) study of voluntary association. They found that similar people tended

to affiliate, not because they sought one another out in associations with diverse membership, but because voluntary associations tended to be homogeneous. From this they concluded that induced homophily was more prevalent than choice homophily. This distinction is clear so long as one focuses on affiliations within organizations; however, the decision to join a voluntary association is also a choice. In such cases, the line between choice and induced homophily depends on the level of decision. For instance, if an individual chooses to join a homogeneous voluntary association, such as a racially exclusionary country club rather than an association that is more diverse, choice homophily at the organizational level leads to induced homophily at the level of individual ties. This suggests that choice homophily has priority over induced homophily, because induced homophily is a consequence of earlier choice homophily.

At whichever level it occurs, homophily involves the pulling together of similar people through the joint work of individuals to build and maintain relationships. In this way, homophily is a form of collective action. Heterophily involves pulling together people who are complementary, such as persons from different clans or of different genders, so it too is a form of collective action. The implication is that theories of collective action may contribute to an understanding of homophily.

A number of formal theories of collective action have been proposed (Heckathorn, 1988, 1989, 1990, 1993, 1996; Marwell & Oliver, 1993; Macy, 1990), any of which could provide a basis for a theory of homophily. One approach would be to detail the implications of each for an explanation of homophily, compare those conclusions with what is known about homophily, and make a principled decision regarding which theory or combination of theories could best explain homophily. That would exceed the scope of this paper.

An important insight from recent studies of collective action is that all collective action produces some form of collective good (i.e. some joint benefit), and that the shape of the production function importantly affects the emergence, stability, and dynamics of collective action (Marwell & Oliver, 1993; Macy, 1990; Heckathorn, 1996). Production functions for collective goods, like production functions for many other goods, often reflect startup costs when production begins, and diminishing marginal returns above an upper level of production. When viewed graphically, the result is an S-shaped curve linking contribution and production levels (Fig. 3). This is the standard function assumed in much economic analysis, and it fits other cases as well, such as social movements (Heckathorn, 1996). Thus, when production begins, the production function is accelerating, it then becomes approximately linear, and it eventually decelerates. Given this production function, the *marginal gains* from each

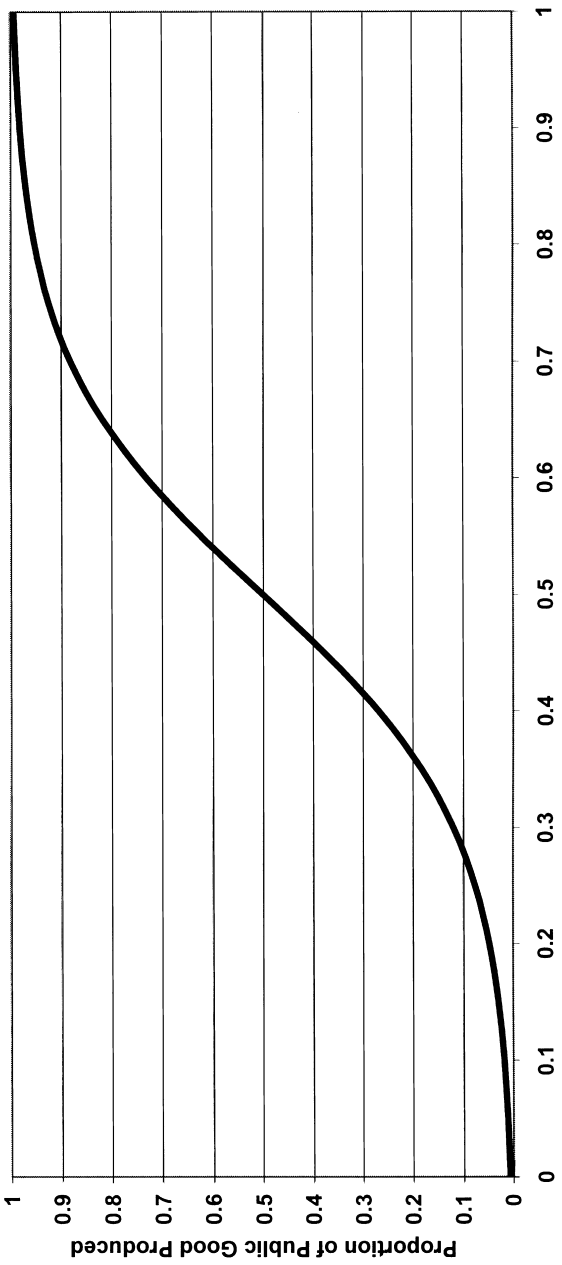

Fig. 3. Generalized Production Function.

additional unit of contribution (e.g. from the addition of each contributor) are initially low, but eventually they increase and then decline, thereby defining a bell-shaped curve (Fig. 4). The curve defines the net gains attainable from participating in a collective endeavor, including collective action.

To see how this is related to homophily, consider the hypothetical case of someone who wants to participate in a joint endeavor, such as playing volleyball, protesting government policy, or participating in a drug-oriented subculture. If the number of suitably oriented persons is too small, the endeavor is not feasible, so homophily based on the endeavor is low and collective action fails. If the number of suitably oriented persons is larger, then by banding together they make the endeavor possible, and homophily is substantial. Finally, if the number of suitably oriented persons is so large that no special effort needs to be made to reach them, homophily is again low. Thus, a curvilinear relationship can theoretically be expected between the abundance of persons suitable for the collective endeavor and homophily.

One of us observed the emergence of a solidary group that illustrated some of the above-described processes. Two first-year graduate students learned that they shared a love of opera. They searched the graduate student community in other departments and located several other opera buffs, and then established a routine. They car-pooled to New York City, attended performances, and returned very late the same night. By sharing transportation, they reduced the burden of driving and the cost of gas; by returning the same night they avoided hotel bills; and by attending jointly, they shared the performances with others who appreciated them. When Wagner's *Ring* cycle was presented, the group showed remarkable stamina by commuting to New York each day of the series and returning early the next morning. Had the number of opera fans been smaller, such a group could never have formed, and had it been far larger, devotees would not have had to seek out one another, so the impact on affiliation patterns would have been minimal.

Support for the hypothesis that homophily depends on group size is provided by an analysis of homophily by age. Figure 5 shows the relationship between homophily and the proportion of injectors aged 18 to 25 in RDS samples from New London, Middletown, and Meriden, Connecticut. When the percentage of younger injectors is small (9.6%), so too is homophily (7.2%). When the percentage of younger injectors is greater (18.7%), homophily becomes very substantial (56%). Finally, when the percentage of younger injectors is large (48.5%), homophily falls dramatically, to 13.6%.

This pattern becomes comprehensible if associations among younger injectors are seen as a form of collective good. If there are not enough younger injectors, forming a youth subculture is impossible, so younger injectors are dissolved

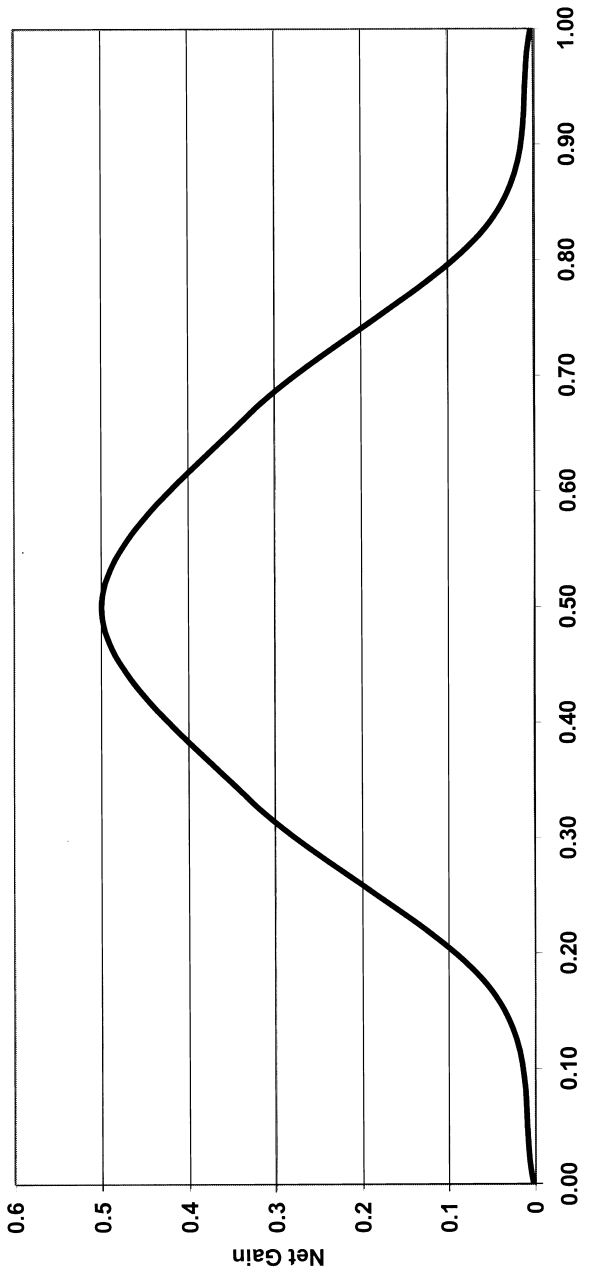

Fig. 4. Net Gain from Homophily.

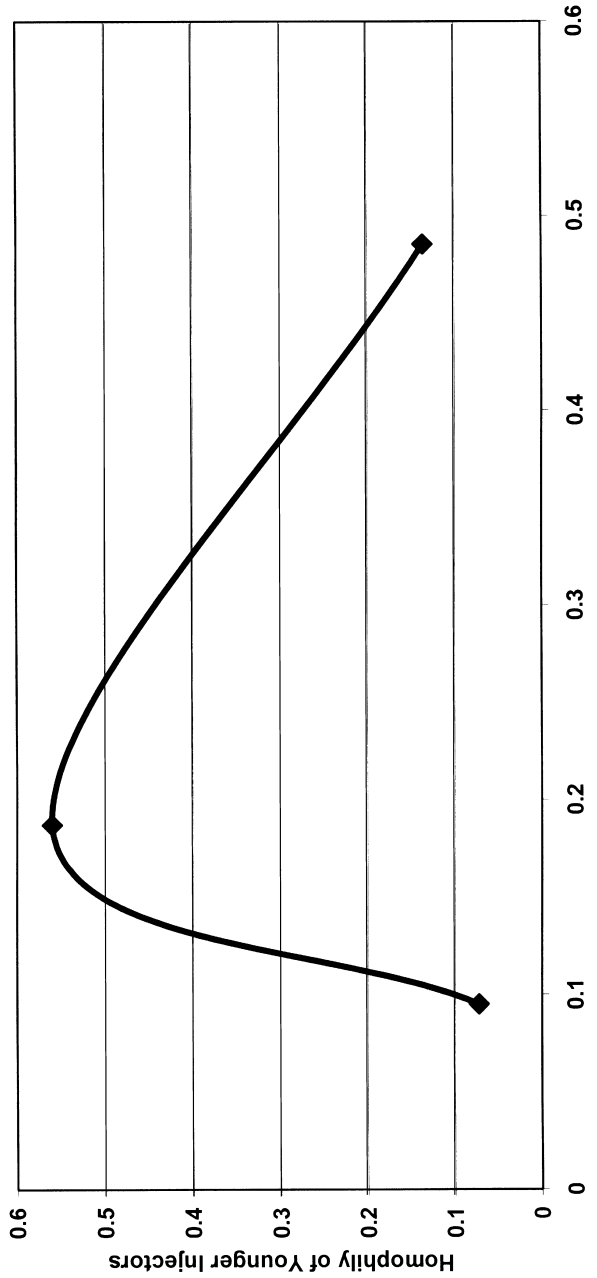

Fig. 5. Population and Homophily: Age of Injection Drug Users.

into the general injector population, and homophily is low. When younger injectors are more common, a critical mass exists for creation of a youth subculture, so homophily is substantial. Finally, when younger injectors are abundant, establishing a youth subculture requires no special selectivity in forming ties. In this instance, younger injectors' taste for associating with other younger injectors is amply satisfied without the need to seek out other younger injectors, resulting in low homophily.

In addition to illustrating how changes in homophily are theoretically linked to collective action, this analysis has substantive implications. Life course studies have routinely emphasized the changes in social positions and relations as individuals age. Understanding this process is complicated by the recognition that these changes are context-dependent; even within a given society, growing older has quite different implications depending on socioeconomic status and other factors. The analysis of homophily among younger injectors illustrates the way in which RDS can be used to study context dependence. The finding reported above shows that whether a younger intravenous drug user becomes part of a homophilous subgroup of younger injectors depends on the proportion of younger injectors in the system. This has theoretic implications for understanding the contexts under which age-specific peer norms will emerge, and practical implications for interventions targeting this group for HIV-prevention services, as illustrated by the relationship between group size and homophily. The bell curve also characterizes syringe sharing. When the proportion of those who do not share syringes is moderate (50% in Meriden) or large (75% in Middletown), their homophily is near zero (0% and 6%, respectively), but when the proportion is intermediate (64% in New London), homophily is substantial (50%). This indicates both that HIV-risk behavior can be salient enough to affect affiliation patterns, and that whether this occurs is context dependent.

Though the production function for interactions by age and syringe sharing may be consistent with the standard S shape, not all production functions can be theoretically expected to take this form, for two reasons. First, because of technological or other constraints, the curve may be truncated, so only a segment of the curve is empirically realized. Collective sanction systems (Heckathorn, 1988) provide an example. These are systems in which responsibility is joint, in the sense that everyone is responsible for one another's behavior. The group may be promised a collective reward or threatened with a collective punishment based on each individual's compliance. Such systems are characterized by an accelerating production function (Heckathorn, 1996), so they reflect only the left part of the S curve. In contrast, in the brokerage systems analyzed by Burt's studies of structural holes, exclusive access to information confers benefits. The value of information degrades as it becomes more broadly disseminated, so the

production function is decelerating. Such systems therefore reflect only the right part of the S curve.

More complex production functions are identified by contemporary theories of collective action. For example, Heckathorn (1996) analyzed the essential features of both collective action and social cooperation as described in a wide range of literatures and found that: (1) social cooperation must yield some form of benefit – a collective good, whose valuation can be low or high or even negative; (2) contribution to production of the collective good entails a cost that varies across individuals and systems; and (3) contributions to production of the collective good are linked to the ultimate amount of production by a production function. The aim of the analysis was to identify the range of social dilemmas that can arise given these three conditions, where a social dilemma is defined as a situation in which individual actions can aggregate into a collectively irrational outcome. By this definition, collective action is concerned with resolving social dilemmas.

Based on an analysis of these three essential features of collective action systems, a theorem was derived showing that collective action systems can give rise to exactly five possible forms of social dilemmas (Heckathorn, 1996). Distinct issues are problematic in each of the five dilemmas. Although much of the formal-theory literature has focused on the prisoner's dilemma, in which trust is problematic, other dilemmas were identified. These include coordination problems, which arise when production functions are accelerating and hence interdependence is high; bargaining problems, which arise when production functions are decelerating and hence interdependence is low; invisible-hand systems when the gains from cooperation are high; and the altruist's dilemma in which autonomous action yields greater benefits than does collective action. Given that each of these five dilemmas is characterized by a distinct production function, and that RDS can help to identify the shape of production functions empirically, RDS potentially provides a means for identifying the type of dilemma underlying collective action systems.

Author: Should these be numbered?

Power Homophily and Social Inequality

An important form of homophily results from a tendency for ties to form among status equals. One means by which this can occur is illustrated by Axelrod's (1984) account of the emergence of cooperation in very large groups. He argued that a prerequisite for the emergence of cooperation under these circumstances is that those who are disposed toward cooperation are able to identify one another. By selectively interacting, cooperators form homophilous clusters. The non-cooperators are thereby left only with the opportunity to interact with one

another. The emergence of cooperation thereby entails the emergence of a stratification system. This is consistent with the deviance literature, in which the principal social sanction involves allocation of status. Norm adherence increases status; deviance decreases it. Norm adherents therefore form homophilous clusters. This may be termed *power homophily* because homophily reflects this group's superior social status. The non-elite group is also homophilous because of members' inability to form ties with the elite. This may be termed *exclusion homophily* because homophily results from exclusion from groups of higher status. This differentiation between power and exclusion homophily is illustrated by studies of the network structure of U.S. schools (Farmer and Farmer 1996). When each student was asked to list the other students he or she most liked, the most popular students were often named by both the most and the least popular. However, this does not mean that the less popular students are isolates, for they associate with one another.

Thus homophily may reflect not only social differentiation (i.e. separate but equal) but also social inequality (i.e. separate but unequal). A process of this type was found in the study of Connecticut drug injectors. Heavy injectors, defined as injectors whose injection frequency exceeds 20 during the last 30 days, appeared to have differing status depending on their numbers (see Fig. 6). The relationship has a sharp negative slope. In New London the percentage of heavy injectors was small (22%), and their homophily was substantial (42%). In Middletown the percentage of heavy injectors was greater (31%), and their homophily was lower (36%). In Meriden the percentage of heavy injectors was still larger (47%), and homophily fell to near zero (−1%). To make sense of this finding, it is useful to consider the resources relevant to an injector population. One resource is a reliable source of drugs. This requires ties to multiple sources, because any single source could run dry or be shut down by police, thereby producing an interruption in drug access. A second major concern is police activity. Knowledge of recent arrests is valuable, as those who are arrested are frequently pressured to inform on their associates. Similarly valuable is knowledge of where enforcement efforts are concentrated, because enforcement intensity varies by area. These resources are relevant to understanding the relationship between homophily and injection frequency because heavy injectors buy more product and hence tend to have better contacts with suppliers and purchase their drugs at lower prices. Through these contacts, they are also likely to have better information about police. When heavy injectors are modest in number, the information they possess is scarce and therefore valuable. To obtain further knowledge, heavy injectors may tend to seek out one another, thereby exhibiting power homophily. They occupy a privileged status. However, as in Burt's (1998) brokerage systems, contacts with

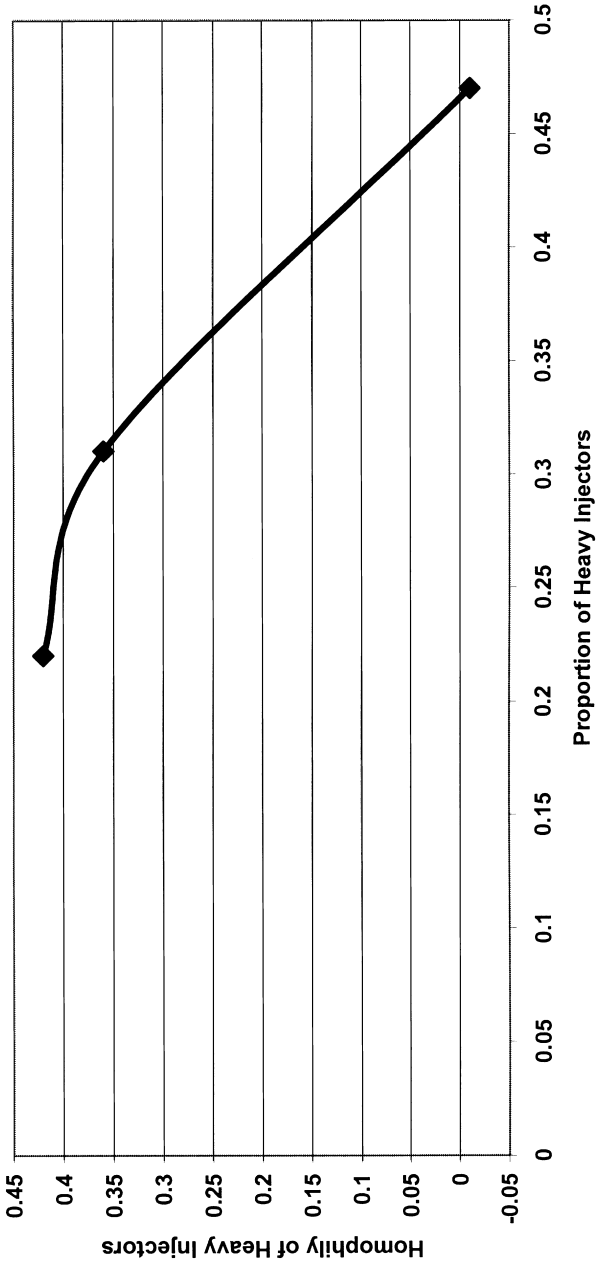

Fig. 6. Population and Homophily: Injection Frequency.

additional heavy injectors have diminishing marginal value. At some point, additional information about dealers and police becomes redundant, so when heavy injectors are abundant, an individual's knowledge is not in high demand and the homophily of heavy injectors approaches zero. They are neither differentially sought out by themselves (power homophily) nor do they thereby induce homophily in others (exclusion homophily). The system thereby ceases to be stratified.

CONCLUSION

Solidary groups are united by a shared attribute that shapes affiliation patterns. The proposed means for conceptualizing and measuring homophily provides an approach for studying these patterns. Socially irrelevant factors, such as whether one's surname has an odd or an even number of letters, do not affect affiliation, so homophily is zero. By contrast, socially salient factors, such as race and ethnicity, affect affiliation; hence, they have non-zero homophily, as do factors associated with them, such as in parts of the United States where it matters whether one's surname ends in a vowel or a consonant.

The essential hypothesis is that solidary groups emerge through collective action to produce a collective good or goods. In the case of injection drug-using communities, these collective goods involve information about enforcement activities, such as where police surveillance is most intense; and information about sources of drugs, including assessments of the quality of drugs available from alternative dealers, and the pricing and reliability of those dealers. In addition, the group provides other goods, such as companionship, a sense of common purpose derived from collective participation in a non-conventional lifestyle, and access to potential sexual partners. A high level of cooperation therefore characterizes groups of injection drug users. Thus, lonely people may join not just for the drugs but to become a part of a supportive community. This also illustrates one of the reasons it is important to foster alternative focal activities around which solidary groups can congeal.

The creation of solidary groups can be facilitated through bringing together people who share common interests or problems. This possibility is illustrated by interventions implemented to combat AIDS that were developed and implemented by Broadhead and Heckathorn (1994). The interventions are based on what are termed secondary sanctions – rewards given not for the individual's behavior but for the positive behaviors the individual elicits from peers. These incentives provide a means for harnessing peer pressure. They also provide a means for initiating the development of prosocial norms, by strengthening the incentives of individuals to regulate one another's behavior. The original

intervention of this type (Broadhead et al., 1998) targeted active injection drug users for HIV prevention education and services. The aim was to create and strengthen prevention norms and thereby reduce HIV-associated risk behavior. The design was highly participatory, in that participants in the intervention carried out two of the most important tasks, education of their peers and recruitment using a form of RDS. The intervention met with sufficient success so as to be described in a National Academy of Sciences report as the "state of the art of preventive intervention" (Institute of Medicine, 1995). To emphasize the extent to which the intervention design is based on participants providing services for their peers, the intervention is termed a *peer-driven*, or equivalently, a *participant-driven intervention* (PDI).

An extension of the PDI design was intended to increase adherence to antiretroviral therapy among HIV-infected active injection drug users (Broadhead et al., in press). Through a combination of standard medical screening and participant-driven recruitment, a set of subjects were identified. These subjects were then provided the opportunity to participate in the intervention, in which each would play the roles of Health Advocate and Peer. As Health Advocate, a subject regularly monitored the adherence of his or her Peer. Very modest financial rewards were provided based on the ability to elicit positive responses from the Peer. The aim was to harness peer pressure on behalf of the intervention's public health objective (increasing adherence to therapy), as well as reducing HIV risk behavior. What is notable in this context is that the subjects bonded to the intervention, welcoming the opportunity to serve as a vital part of the intervention and strengthening their ties to other intervention participants. In essence, the intervention created a structure in which prosocial behavior flourished, a structure that promoted the development of a solidary system. Its aims were to harness peer pressure, create pro-health norms, and increase adherence to the exacting demands of primary care for HIV disease. Given that the alternatives to those seeking solidary associations in some communities include joining the injection drug-using community and or a gang, the development of a means for fostering safer and healthier forms of solidary association is a worthwhile objective.

NOTE

1. Extensiveness and compliance are therefore not the causes of solidarity, rather they are the terms by which solidarity is defined and measured. As Markovsky and Lawler (1994) have noted, this distinction is not consistently maintained by Hechter (1987), who at various points discusses extensiveness and compliance as both the causes and definitions of solidarity, but an examination of the internal logic of the model makes clear that these serve as definitional terms.

ACKNOWLEDGMENTS

An earlier version of this paper was presented at the conference Solidarity and Prosocial Dispositions held in Schiermonnikoog, The Netherlands, March 27–29, 2001. This research was made possible by grants from the Centers for Disease Control and Prevention (U62/CCU114816–01) and the National Institute on Drug Abuse (RO1 DA08014).

REFERENCES

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Blau, P. M. (1977). *Inequality and Heterogeneity*. New York: Free Press.

Blau, P. M. (1994). *Structural Context of Opportunities*. Chicago: University of Chicago Press.

Broadhead, R. S., & Heckathorn, D. D. (1994). AIDS Prevention Outreach among Injection Drug Users: Agency Problems and New Approaches. *Social Problems*, *41*, 473–495.

Broadhead, R. S., Heckathorn, D. D., Weakliem, D., Anthony, D., Madray, H., Mills, R., & Hughes, J. (1998). Harnessing Peer Networks as an Instrument for AIDS Prevention: Results from a Peer-Driven Intervention. *Public Health Reports*, *113* (Supplement 1), 42–57.

Broadhead, R. S., Heckathorn, D. D., Altice, F. L., van Hulst, Y., Carbone, M., Friedland, G., O'Connor, P. G., & Selwyn, P. A. (in press). Increasing Drug Users' Adherence to HIV Therapeutics: Feasibility Results of a Peer Driven Intervention Study. *Social Science and Medicine*.

Burt, R. (1998). The Gender of Social Capital. *Rationality and Society*, *10*, 5–46.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.

Durkheim, E. (1956). *The Division of Labor in Society*. New York: Free Press.

Fararo, T. J., & Doreian, P. (1998). The Theory of Solidarity: An Agenda of Problems. In: P. Doreian & T. Fararo (Eds), *The Problem of Solidarity: Theories and Models* (pp. 1–33). Amsterdam: Gordon and Breach Publishers.

Fararo, T. J., & Sunshine, M. H. (1964). *A Study of a Biased Friendship Net*. Syracuse NY: Syracuse University Press.

Farmer, T. W., & Farmer, E. M. Z. (1996). Social Relationships of Students with Exceptionalities in Mainstream Classrooms: Social Networks and Homophily. *Exceptional Children* (March/April).

Hechter, M. (1987). *Principles of Group Solidarity*. Berkeley, CA: University of California Press.

Heckathorn, D. D. (1988). Collective Sanctions and the Emergence of Prisoner's Dilemma Norms. *American Journal of Sociology*, *94*, 535–562.

Heckathorn, D. D. (1989). Collective Action and the Second-Order Free-Rider Problem. *Rationality and Society*, *1*, 78–100.

Heckathorn, D. D. (1990). Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control. *American Sociological Review*, *55* (June), 366–384.

Heckathorn, D. D. (1993). Collective Action and Group Heterogeneity: Voluntary Provision vs. Selective Incentives. *American Sociological Review*, *58*, 329–350.

Heckathorn, D. D. (1996). Dynamics and Dilemmas of Collective Action. *American Sociological Review*, *61*, 250–277.

Heckathorn, D. D. (1997). Respondent Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems*, *44*, 174–199.

Heckathorn, D. D. (2002). Respondent-Driven Sampling II: Deriving Statistically Valid Population Estimates from Samples of Hidden Populations. *Social Problems*, *49*, 11–34.

Heckathorn, D. D., & Jeffri, J. (2001). Finding the Beat: Using Respondent-Driven Sampling to Study Jazz Musicians. *Poetics*, *28*, 307–329.

Heckathorn, D. D., Broadhead, R. S., Anthony, D. L., & Weakliem, D. L. (1999). AIDS and Social Networks: Prevention through Network Mobilization. *Sociological Focus*, *32*, 159–179.

Heckathorn, D. D., & Lucas, W. L. (1982). Bridging Consensus and Conflict Theories in Drug Use Research: A 'Unified' Theoretic Perspective. *Journal of Drug Issues*, *12*, 443–455.

Heckathorn, D. D., Semaan, S., Broadhead, R. S., & Hughes, J. J. (in press). Extensions of Respondent-Driven Sampling: A New Approach to the Study of Injection Drug Users Aged 18–25. *AIDS and Behavior*.

Homans, G. C. (1950). *The Human Group*. New York: Harcourt, Brace and World.

Institute of Medicine (1995). *Assessing the Social and Behavioral Science Base for HIV/AIDS Prevention and Intervention*. Washington, D.C.: National Academy Press.

Lawler, E. J., & Yoon, J. (1998). Network Structure and Emotion in Exchange Relations. *American Sociological Review*, *63*, 871–894 (need page numbers).

Macy, M. W. (1990). Learning Theory and the Logic of Critical Mass. *American Sociological Review*, *55*, 809–826.

Markovsky, B., & Lawler, E. J. (1994) A New Theory of Group Solidarity. *Advances in Group Processes*, *11*, 113–137.

Marwell, G. & Oliver, P. (1993). *The Critical Mass in Collective Action: A Micro-Social Theory*. Cambridge U.K.: Cambridge University Press.

McPherson, J. M. (1998). *Personal communication*. San Francisco.

McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups. *American Sociological Review*, *52*, 370–379.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.

Rapoport, A. (1979). A Probabilistic Approach to Networks. *Social Networks*, *2*, 1–18.

Simmel, G. (1955). *Conflict: The Web of Group Affiliations*. K. H. Wolff & R. Bendix (Trans.). New York: Free Press.

Widegren, O. (1997). Social Solidarity and Social Exchange. *Journal of the British Sociological Association*, *31*, 755–773.

Yamaguchi, K., & Kandel, D. (1993). Marital Homophily on Illicit Drug Use among Young Adults: Assortative Mating or Marital Influence? *Social Forces*, *72*, 505–528.

APPENDIX: COMPUTING HOMOPHILY

This appendix describes the procedures for computing homophily vased on RDS data. A challenge is that whereas information on affiliation patterns is available through recording patterns of recruitment, computing homophily also requires information on group size. This is problematic when studying populations such as injection drug users (Heckathorn et al., in press) and jazz musicians (Heckathorn & Jeffri, 2001), for which no sampling frame exists. For example, there are no membership records for these groups. Therefore, methods were

developed for estimating subgroup sizes based on available data, specifically affiliation patterns and network sizes. This appendix summarizes these procedures for the two-category case. For amore detailed presentation that includes analysis of the general n-category case, see Heckathorn (2002).

Estimates of group size using RDS are based on ana analysis of the structure of networks of reciprocal ties. In such networks, relative group sizes can be computed based on two types of information, affiliation patterns and network sizes (Heckathorn, 2002). This method for estimating population size is suitable for analysis of RDS data because respondents recruit almost exclusively – that is, more than 97% of the time – from those with whom they have preexisting ties, generally friends, acquaintances, or relatives. Such ties are reciprocal because a link from any individual x to y implies that a link also exists from y to x. Hence foe two groups a and b, the number of llinks from a to b (T_{ab}) will equal the number from b to a (T_{ba}), so $T_{ab} = T_{ba}$. Furthermore, the number of ties from any group x to y is the production of three terms, the size of the group (P_x), the mean network size of group members (N_x), and the proportion of ties from that go and from x to y (S_{xy}), so, $T_{xy} = P_x N_x S_{xy}$. Hence, for two groups, a and b, $P_a N_a S_{ab} = P_b N_b S_{ba}$. Without loss of generality group size can be expressed as a proportion, so $1-P_a$ can be substituted for P_b, and this expression can be solved for group a's size, P_a as follows:

$$P_a = \frac{S_{ba} N_b}{S_{ba} N_b + S_{ab} N_a}$$

This is the estimate of group size based on the reciprocity model (Heckathorn, 2002), and it provides the means for estimating subgroup sizes based on both affiliation data, and on self-reported network size. For example, consider again the case of gender. From Fig. 1B, where females are group a and males are group b, the proportion of males recruited by females was $S_{ab} = 0.58$, and the proportion of females recruited by males was $S_{ba} = 0.264$. The mean network size for females was $N_a = 37.5$, and $N_b = 57.1$ for males. Substituting these values into the above expression yields the estimated proportion of females in the population as:

$$\frac{0.264 \quad 57.1}{0.264 \quad 57.1 \; + 0.58 \quad 37.5} = 0.41$$

Finally, homophily is computed based on the population estimate derived from the reciprocity model, and from the recruitment selection proportions in the manner described above.