A Note on the General Social Survey's Ersatz Network Density Item

Fall, 1985

Ronald S. Burt

Department of Sociology and Center for the Social Sciences
Columbia University
New York, NY 10027

GSS Technical Report No. 63 40

AUTHOR NOTE -- This technical note is a by-product of support from the National Science Foundation Sociology Program (SES-8208203) and the Measurement Methods and Data Improvement Program (SES-8513327) and has been produced as part of the Research Program in Structural Analysis housed at Columbia University's Center for the Social Sciences.

A Note on the General Social Survey's Ersatz Network Density Item

Data obtained with the GSS ersatz network density item are compared to density data obtained with the more traditional, more costly, GSS sociometric network items. The inexpensive ersatz density data are not independent of network density, but they are almost completely unreliable. The full range of possible densities occurs at each level of ersatz density and only 1% to 2% of variation in network density can be described with ersatz density. Hypotheses operationalized with the ersatz density variable specified as a predictor will be biased toward the null hypothesis. Given this GSS experiment, the reliability of conclusions from studies replacing sociometric network items with inexpensive items purporting to measure network structure directly should be viewed with suspicion.

In a sense, the 1985 General Social Survey (GSS) is a return to the tradition of social survey research established by Paul Lazarsfeld with various Columbia University colleagues through his voting studies (Lazarsfeld et al., 1944; Berelson et al., 1954), marketing studies (e.g., Katz and Lazarsfeld, 1955), and the more detailed substantive and methodological studies of his colleagues and students (e.g., Lipset et al., 1956; Coleman, 1958; 1961; Coleman et al., 1966; Rossi, 1966; Barton et al., 1973; see Barton, 1982, for historical review). In this tradition, respondent attitudes and behaviors are studied in the context of interpersonal environments, the social setting for respondent attitudes and behaviors. Taking advantage of recent developments in network analysis, the 1985 GSS brings this tradition to the national sampling frame. The usual rich diversity of data obtained in the GSS on American attitudes and behaviors is enhanced in the 1985 survey with network data on the interpersonal environments of respondents. Each respondent was asked: "Looking back over the last six months, who are the people with whom you discussed matters important to you?" Diverse data were then obtained on relations with and among the first five persons named.[1]

From these network data, measures can be constructed to represent the complexity of relationships comprising each respondent's interpersonal environment. Network density, the average strength of relations surrounding the respondent, is one of the most basic measures. Among the expected correlates of dense interpersonal environments are greater social support, poor access to diverse social resources, stereotypical opinions, behavior heavily influenced by normative factors, and poorly developed skills in interpersonal negotiations.

---

[1]Burt (1984) provides a detailed discussion of the network data and various issues taken into account by the GSS Board of Overseers in their deliberations over the network items.

Network size, density and a host of other network form measures can be specified in traditional survey research models, enhancing the models by taking into account the social context for respondent behavior and opinion.

The network items adopted for the GSS are sociometric in the sense of asking respondents to name specific people and describe relations with and between each pair of named individuals. Sociometric items were adopted on the assumption that they are more reliable and valid than the comparatively crude network items more typical of survey research.

However, no national probability data were available with which the assumption could be checked and some of the more traditional survey researchers on the GSS Board of Overseers expressed skepticism, preferring the simplicity and cost savings of asking respondents to report directly on qualities of their interpersonal environment.

An ersatz network item was hastily written as a methodological experiment to provide national probability sample data with which the assumed superiority of the sociometric items could be studied. My purpose in this brief note is to report the results of the experiment.

THE GSS ERSATZ NETWORK DENSITY ITEM

In contrast to the sociometric items in which data on specific relationships were recorded, the ersatz network item had respondents describe density directly in four broad response categories:[2]

> Some people have friends who know one another. Other people have friends who don't know one another. Would you say that all of your friends know one another, most of your friends know one another, only a few of your friends know one another, or none of your friends know one another?

Providing little data on the social structure or composition of the respondent's interpersonal environment, the ersatz network item is obviously no substitute for sociometric network items. Moreover, it asks a great deal of respondents. Respondents have to define their friends, inventory the relations among their friends, and then summarize those relations. The complex nature of the response task makes response reliability suspect. All of this notwithstanding, the ersatz network item requires very little time to administer and the

---

[2]Only 21 of the 1534 respondents had no answer; 15 could not answer the question on general grounds, 4 said that they had no friends and so could not answer the question, and 2 did not answer the question.

empirical question remains: How much of the precision provided by the sociometric items for measuring network density is lost when density is measured by these broad response categories?[3] If network density can be measured adequately with an ersatz network item, perhaps other significant aspects of interpersonal environments can also be measured with such items, eliminating the general need for the more expensive sociometric items.[4]

RELATIONS AMONG SOCIOMETRIC ALTERS

Network density means and ranges within each level of ersatz density are presented in table 1. In the first panel, density is the average strength of the relationship between any two pairs of a respondent's discussion partners. Density in the second panel is the proportion of those discussion partners tied by an especially close relationship. Density in the third panel is the proportion of a respondent's discussion partners who were total strangers in the sense that they would not have recognized one another if they "bumped into one another on the street."

There is a nonlinear association between ersatz density and the two strength of tie densities. Respondents who said that most of their friends know one another -- i.e., respondents high on ersatz density -- were more likely than average to cite discussion partners they perceived as strongly connected to one another (.60 network density and .53 strong tie density). Strong relations were less frequent among discussion partners cited by respondents lower on ersatz density. However, respondents claiming that none of their friends knew one another -- i.e., respondents low on ersatz density -- were once again more likely than average to perceive strong relations among their discussion partners (.51 network density and .46 strong tie density).

---

[3]Two issues should be noted in comparing the sociometric and ersatz network items. First, the two refer to slightly different kinds of relations and that might underlie differences between them. The sociometric items elicit relations with and among important discussion partners while the ersatz network item asks about relations among friends. Since the sociometric items only asked about the closeness between pairs of discussion partners, however, density from both measures is indexing the general strength of social relations among people close to the respondent. Second, the two might be confounded by being administered together in the last half of each interview. The ersatz and sociometric network items were not administered in an overlapping split ballot design, however, the ersatz network item (question number 98) was administered several topics ahead of the full GSS network items (questions 127 through 139).

[4]Comparing the ersatz network responses to the sociometric density data offers information on the reliability of the ersatz network item. Its validity would be revealed by studying how much correlations between criterion variables and sociometric density are attenuated when density is measured by the ersatz network item. This initial note is concerned with reliability because the validity question is moot if the ersatz network item offers unreliable data.

There is a monotonic association between ersatz density and the density of holes in a respondent's network. In the third panel of table 1, notice that the average proportion of completely absent relations among a respondent's discussion partners increases from .11 to .18 to .22 to .33 down the levels of ersatz network density. In other words, ersatz density better describes the density of acquaintance relations than the density of strong relations. This seems fitting in as much as the ersatz density item asked about friends knowing one another rather than friends being especially close to one another.

These differences across levels of ersatz density are nonrandom. The F ratio for the null hypothesis of equal means in table 1 is 5.99 for network density, 5.41 for strong tie density, and 7.73 for stranger density, each of which would have less than a .001 probability of occurring with 3 and 1,150 degrees of freedom if the null hypothesis were true.

But a reliable indicator ought to have more than a nonrandom association with the condition it purports to measure. The range results in table 1 show that precious little variation in network density is described by the ersatz density item. Notice that all three network density measures range from 0 to 1, the minimum possible to the maximum possible, within every level of ersatz network density. Further, notice that the standard deviation of network density across respondents within levels of ersatz density is often as large or larger than the standard deviation across all respondents.

The problem is illustrated in figure 1 with graphs of network item density across levels of ersatz density. The bold line in the top graph indicates the mean density of relationships between discussion partners and the bold line in the bottom graph indicates the mean density of strangers (network and stranger density respectively in table 1). The immediately higher and lower lines in each graph indicate the 99% confidence interval around means within each level of ersatz density. The lines farthest removed from the means in each graph indicate the upper and lower limits of the interquartile range (75^{th} and 25^{th} percentiles) within each level of ersatz density.

As a general indicator of network density, ersatz density is hopelessly unreliable. Note in the top graph of figure 1 that there is very little difference in the range of network densities observed at the maximum and minimum levels of ersatz density. The average within each level of ersatz density is well within the range of scores within any other level of ersatz density. The density

Table 1

Network Density by Levels of Ersatz Density

	MEAN	SD	Minimum	Maximum
NETWORK DENSITY				
All Know One Another	.600	.354	.00	1.00
Most Know One Another	.498	.326	.00	1.00
Few Know One Another	.453	.323	.00	1.00
None Know One Another	.505	.368	.00	1.00
All Respondents	.492	.331	.00	1.00
STRONG TIE DENSITY				
All Know One Another	.528	.415	.00	1.00
Most Know One Another	.417	.375	.00	1.00
Few Know One Another	.371	.364	.00	1.00
None Know One Another	.464	.396	.00	1.00
All Respondents	.413	.378	.00	1.00
STRANGER DENSITY				
All Know One Another	.112	.223	.00	1.00
Most Know One Another	.175	.270	.00	1.00
Few Know One Another	.218	.292	.00	1.00
None Know One Another	.328	.362	.00	1.00
All Respondents	.189	.279	.00	1.00

NOTE -- Network density is the average strength of relations among the specific people cited as important discussion partners (especially close = 1, acquainted = .2, strangers = 0). Strong tie density is the proportion of a respondent's discussion partner pairs tied by an especially close relationship. Stranger density is the proportion of a respondent's discussion partner pairs who were total strangers. Density is available for respondents naming two or more discussion partners. From maximum to minimum, the distribution of respondents in this table across levels of ersatz network density is 111, 603, 406, and 34.

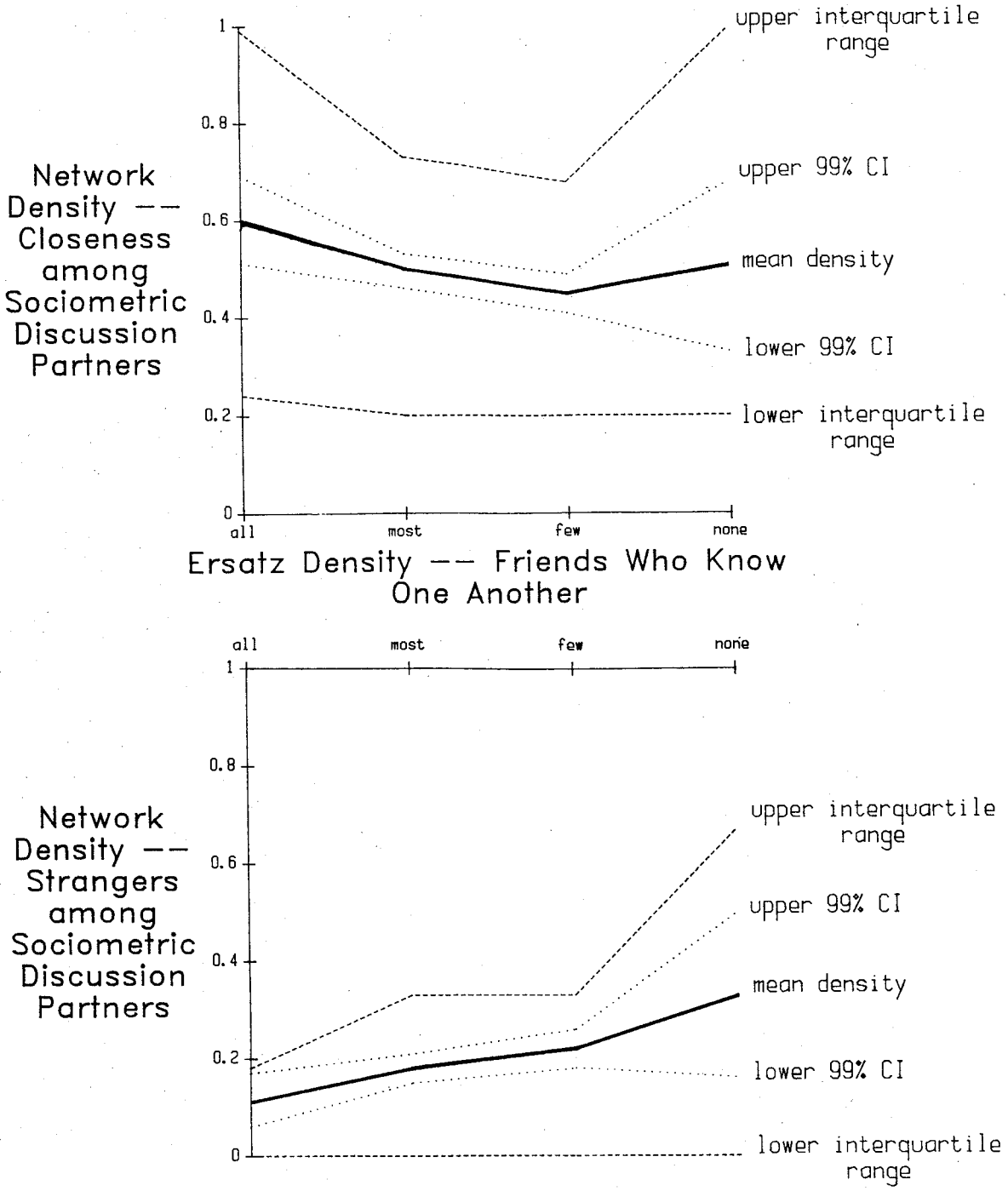

Figure 1

Variation in Network Item Density Across Levels of Ersatz Density

of close relations among respondent sociometric citations is very poorly predicted from ersatz density.

Ersatz density is a slightly better indicator of acquaintance. Note in the bottom graph of figure 1 that the discussion partners of respondents claiming that all their friends know one another do indeed tend to know one another and the distribution of stranger density is noticeably tighter around this average tendency than anywhere else in figure 1. At the other levels of ersatz density, however, stranger density is quite variable. Average stranger density within each level of ersatz density is within the range of scores at the "most," "few" and "none" levels of ersatz density. Putting aside maximally dense networks, the density of acquaintances among respondent sociometric citations is poorly predicted from ersatz density.

Summarizing variation within and between the ersatz density categories, the nonlinear correlation eta is .124 for network density, .118 for strong tie density, and .141 for stranger density. Even allowing for nonlinearity, in other words, ersatz density can describe no more than 2% of the variance in network density; 1.5% of the variance in network density, 1.4% of the variance in strong tie density, and 2.0% of the variance in stranger density.

RELATIONS WITH SOCIOMETRIC ALTERS

The same results are obtained in a comparison of ersatz density with the complexity of relations between respondents and their discussion partners. Respondents were asked if they felt equally close to all the people they cited as discussion partners or more close to some than others. One could argue that the ersatz density item is a subjective indicator of variation in a respondent's relations with friends as much as it is an indicator of variation in relations among friends. If a respondent feels equally close to his important discussion partners he is unlikely to perceive them as strangers to one another. Ceteris paribus, network density should decrease with increasing distinctions among discussion partners.[5]

Proportions of respondents making no distinctions in their relations to discussion partners are reported in table 2 across levels of ersatz density. About half the respondents felt equally close to their cited discussion partners

[5]There is some evidence of this in the GSS data. The average network density among respondents equally close to their discussion partners is .59 versus a .37 average density observed among respondents distinguishing especially close discussion partners. This simple dichotomy describes 11% of the variation in network density across respondents (.33 eta).

Table 2

Lack of Distinctions among Discussion Partners by Levels of Ersatz Density

	MEAN	SD	Minimum	Maximum
All Know One Another	.681	.467	0	1
Most Know One Another	.548	.498	0	1
Few Know One Another	.550	.498	0	1
None Know One Another	.635	.486	0	1
All Respondents	.568	.496	0	1

NOTE -- The distinctions among discussion partners variable is a sociometric dichotomy between respondents saying that they were closer to some discussion partners than to others (coded 0) versus respondents saying that they were equally close to all of their cited discussion partners (coded 1). From maximum to minimum, the distribution of respondents in this table across levels of ersatz network density is 182, 741, 535, and 52.

and half felt closer to some than others. As in table 1, the respondents most likely to feel equally close to their discussion partners are simultaneously at the high and low levels of ersatz density (.68 if "all friends know one another" and .64 if "none know one another"). Further replicating table 1, the differences across levels of ersatz density are nonrandom. The F ratio for the null hypothesis of no difference in the table 2 means is 4.16 (3,1506 df, $p < .01$) and the likelihood ratio chi-square statistic for the null hypothesis of independence between the equally-variably close dichotomy and categories of ersatz density is 12.73 (3 df, $p < .01$). Still further replicating table 1, these differences account for little of the variation in respondents making distinctions in their relations with discussion partners. Less than 1% of the variance in the equally-variably close dichotomy is accounted for by levels of ersatz density (eta = .091).

Continuing the search for network qualities measured by ersatz density, one could argue that ersatz density reflects network size. The fewer friends a respondent has, the less likely he is to make distinctions in his relations with them and the more likely they are to know one another. Ceteris paribus, network density should decrease with increasing network size. There is little evidence of this in the GSS data. Network size (ranging from 2 to 6) accounts for 2% of the variance in network density (.14 eta). Moreover, the association between network size and ersatz density is as unreliable as the associations described in tables 1 and 2.[6]

KINDS OF RESPONDENTS

The fact that ersatz density is unreliable for Americans as a whole does not imply that it is unreliable for every kind of American. Subpopulations might exist within which ersatz density adequately indicates network density.

---

[6]Network size, the number of people cited as important discussion partners in response to the sociometric name generator, varies from 0 to 6 with a mean of three discussion partners and a 1.71 standard deviation. Networks at the top and bottom categories of ersatz density are smaller than average. Respondents who said that most of their friends knew one another named about two discussion partners and respondents claiming that none of their friends knew one another named about two and half discussion partners. Once again, these differences are nonrandom. The F ratio for the null hypothesis of equal means in table 3 is 15.30 with 3 and 1506 degrees of freedom ($p < .001$). The likelihood ratio chi-square statistic for the null hypothesis of independence between categories of network size and ersatz density is 56.97 with 18 degrees of freedom ($p < .001$). At the same time, little variation in network size is described by ersatz density. There are almost no differences in the range of sizes observed at each level of ersatz density -- sizes ranging from 0 to 6 discussion partners at each level except the "none" level where size ranges from 0 to 5 -- and the standard deviations of size at each level are very nearly as large as the standard deviation across all respondents. A .172 eta summarizes variation within and between the ersatz density categories, indicating that ersatz network density describes about 3% of the variance in network size.

An analysis of kinds of relationships elicited by the GSS network items revealed several dimensions on which Americans varied in mixing contents within discussion relations. High socioeconomic status respondents differed from low. Elderly respondents differed from the young. White differed from nonwhite and urban-single differed from rural-married. These dimensions account for differences in respondent definitions of relations and so seem a logical starting point for distinguishing respondents likely to vary in the adequacy with which ersatz density indicates network density.

The results reported in table 1 and illustrated in figure 1 are repeated across these respondent subpopulations. Summarizing the variation in network density described by ersatz density with the squared nonlinear correlation eta: 2.8% of the variance in network density is described by ersatz density among respondents with only a primary school education and 2.5% is described among respondents educated beyond the Bachelor degree, 1.1% of the variance is described among respondents with annual personal incomes under $4,000 and 2.1% is described among respondents making $25,000 or more, .8% of the variance is described among retired respondents (67-87 years) and 3.8% is described among college age respondents (18-23 years), 1.8% of the variance is described among whites and 1.3% is described among nonwhites, 11.9% of the variance is described among respondents living in towns or villages and 1.4% is described among respondents living in large central cities, 2.1% of the variance is described among married respondents and 2.1% of the variance is described among singled, divorced or separated respondents. Allowing for the variable numbers of cases and the number of comparisons made, the consistently low proportions of variance in network density accounted for by ersatz density offer little promise of ersatz density being a reliable indicator within some subpopulation of Americans.

CONCLUSION

Reliability is the critical weakness of the ersatz density item. The problem is not that ersatz density is independent of network density. As reported in table 1 and illustrated in figure 1, the density of acquaintance within respondent networks increases monotonically with ersatz density and the density of strong ties for respondents claiming that their friends "all know one another" is higher than the density of such ties for respondents believing some or all of their friends to be unacquainted. The problem is that network density

varies so broadly within categories of ersatz density that nothing precise can be said about network density from a respondent's response to the ersatz density item. Density varies from zero to one within each level of ersatz density, its standard deviation within each level is often as large as its standard deviation across all respondents, and less than 2% of the variation in density is described by ersatz density.

This unreliability has dramatic implications for empirical research because network density is often used as an independent variable. Among the expected correlates of dense interpersonal environments are greater social support, poor access to diverse social resources, stereotypical opinions, behavior heavily influenced by normative factors, and poorly developed skills in interpersonal negotiations. It would be nearly impossible to detect density effects with an indicator such as the GSS ersatz density item because effects would be so biased toward the null hypothesis. In the simple model of density (x) predicting an outcome variable (y), for example, the s_{xy}/s_x^2 ordinary least squares estimate of b_{xy} would become $s_{xy}/49s_x^2$ so that a 1.5 regression coefficient from true density in the study population would have an expected value of .03 in empirical research with ersatz density. Moreover, sophisticated measurement models offer little help because it would be difficult to disaggregate the true 2% from the unreliable 98% of ersatz density variance and it is not clear that the dependent variables typically predicted by network density are measured with sufficient precision to reveal the effects of the 2% reliable variance in ersatz density.

In sum, there is no evidence from the GSS experiment -- the only test with data on a national probability sample of Americans -- to encourage the use of ersatz network items as a surrogate for sociometric name generator and name interpreter items. If one were to pursue the course of developing ersatz network items, the results of the GSS experiment indicate that focusing on the structure of absent relations (e.g., total strangers) would yield more reliable data than focusing on the structure of especially close relations. The general result here is that conclusions drawn from a study operationalizing network density with an item such as the GSS ersatz density item must be viewed with suspicion. Given the bias against detecting effects with such an unreliable indicator, effects reported as density effects in such a study are more likely to be the effects of correlated response errors.

REFERENCES

Barton, Allen H. (1982) "Paul Lazarsfeld and the invention of the university institute for applied social research," In *Organizing for Social Research*, edited by B. Holzner and J. Nehnevajsa. Cambridge, MA: Schenkman.

Barton, Allen H., Bogden Denitch and Charles Kadushin (1973, eds.) *Opinion-Making Elites in Yugoslavia.* New York: Praeger.

Berelson, Bernard R., Paul F. Lazarsfeld and William N. McPhee (1954) *Voting.* Chicago: University of Chicago Press.

Burt, Ronald S. (1984) "Network items and the General Social Survey," *Social Networks* 6:293-339.

Coleman, James S. (1958) "Relational analysis: the study of social organizations with survey methods," *Human Organization* 16:28-36.

Coleman, James S. (1961) *The Adolescent Society.* New York: Free Press.

Coleman, James S., Elihu Katz and Herbert Menzel (1966) *Medical Innovation.* New York: Bobbs-Merrill.

Katz, Elihu and Paul F. Lazarsfeld (1955) *Personal Influence.* New York: Free Press.

Lazarsfeld, Paul F., Bernard Berelson and Hazel Gaudet (1944) *The People's Choice.* New York: Columbia University Press.

Lipset, Seymour M., Martin A. Trow and James S. Coleman (1957) *Union Democracy.* New York: Free Press.

Rossi, Peter H. (1966) "Research strategies in measuring peer group influence," In *College Peer Groups*, edited by T. M Newcomb and E. K. Wilson. Chicago: Aldine.