JOURNAL OF CONSUMER PSYCHOLOGY, *10*(1&2), 101–108
Copyright © 2001, Lawrence Erlbaum Associates, Inc.

Other Multivariate Techniques

VII.A. HOW CAN I CREATE A CONTINUUM ON WHICH TO MEASURE RESPONDENTS AND STIMULI?

In my current research, I propose that an attitude can serve for the individual and for the object appraisal. Therefore, if we consider the self and an object as two poles, how can we place an individual's attitude toward the object on a continuum between these two poles?

Specifically, I would like to know a way to place an object on a continuum between two different poles. Thus, I need a way to compute a distance from both poles. Multidimensional scaling allows computing distances between objects. I would like to put several objects on a continuum. In my research, I assume that attitude toward an object can serve for an individual himself or herself (e.g., by preserving internal consistency) or for managing relationships with others (e.g., by preserving accordance). Therefore, I would like to get a self–others continuum on which several people's attitudes toward an object could be placed. For instance, attitude toward a "green" product can be more or less personal or social depending on the respondent. (In fact, my concern is more elaborate because I need to take three poles into account, but I think that the problem is the same, with three continua.)

Editor: It is always comforting to learn that you are thinking along the lines of great thinkers; the method you are seeking is one of 'unfolding' under the rubric of unidimensional scaling, á la Coombs, Thurstone, and so on, or preference modeling in multidimensional scaling (Carroll & Arabie, 1980; Coombs, 1964, pp. 80–180; Davison, 1983, pp. 152–188; DeSarbo, Young, & Rangaswamy, 1997; Green, Carmone, & Smith, 1989, pp. 90–99; McIver & Carmines, 1981, pp. 72–78; Torgerson, 1958, pp. 403–417; also see Andrich & Styles, 1998). If you have perceptual data that distinguish between stimuli on some quality, and you have preference data that indicate a participant's degree of attachment to each of those stimuli, you can place that individual on the same perceptual space that represents the similarity among the stimuli.

Imagine a horizontally drawn 9-point scale ranging from 1 (*liberal*) to 9 (*conservative*), along which there are statements about various political and current events issues. Perhaps their placement has been determined by obtaining the means from some sample that had been asked to evaluate the liberalness or conservativeness of each issue, with the resulting scale ranging from, for example, 1.7 (*pro-privacy*) to 8.2 (*anti-gun control*), and so on. In a variety of manners, you elicit information from your participant as to where he or she falls along the continuum (usually something like, "To what extent do you agree with this issue?" for each issue). Say the respondent is somewhat conservative and is placed around 6.3 on the scale. Then, you can imagine folding up this 1 through 9 scale at the 6.3 mark, until the left of the scale and the right of the scale look like a lopsided *V,* and keep folding until they are eventually both vertical. Then, the respondent is at the bottom of the vertically folded scale, and the anti-gun control statement is 1.9 (= 8.2 – 6.3) units above the respondent, and the pro-privacy statement is 4.6 (= 6.3 – 1.7) units above the respondent (or 2.7 = 4.6 – 1.9 units above the anti-gun statement). The distance from the respondent represents the extent to which he or she agrees with the political issue.

There is no particular magic in the metaphor of the folding or unfolding (we could have measured those distances while the data laid along the horizontal continuum), and indeed, there is not much to be leveraged from the image of folding or unfolding when generalizing to multiple dimensions. Even while striving toward goals like parsimony, we tend to believe that more dimensions will help us understand consumer preferences better than fewer dimensions, and so the implementation of multidimensional scaling (MDS) vastly dominates that of unidimensional scaling. In MDS, you are simply looking to fit the stimuli into a perceptual map and the respondents onto the same space as an overlay to represent their preferences, usually as ideal points. (In consumer behavior, we are often working with actual preferences, but these models work just as well in scenarios like that previously mentioned or that raised in the question regarding the alignment of a respondent with a set of attitudinal statements.) If you are comfortable with MDS, as the question suggests, and you could obtain such a solution, a unidimensional continuum could easily be derived, from projecting the stimuli onto any of the dimensions, or some other theoretically meaningful orientation in space.

REFERENCES

Andrich, David, & Styles, Irene M. (1998). The structural relationship between attitude and behavior statements from the unfolding perspective. *Psychological Methods, 3,* 454–469.

Carroll, J. Douglas, & Arabie, Phipps. (1980). Multidimensional scaling. *Annual Review of Psychology, 31,* 607–649.
Coombs, Clyde H. (1964). *A theory of data.* New York: Wiley.
Davison, Mark L. (1983). *Multidimensional scaling.* New York: Wiley.
DeSarbo, Wayne S., Young, Martin R., & Rangaswamy, Arvind. (1997). A parametric multidimensional unfolding procedure for incomplete nonmetric preference/choice set data in marketing research. *Journal of Marketing Research, 34,* 499–516.
Green, Paul E., Carmone, Frank J., Jr., & Smith, Scott M. (1989). *Multidimensional scaling: Concepts and applications.* Boston: Allyn & Bacon.
McIver, John P., & Carmines, Edward G. (1981). *Unidimensional scaling.* Newbury Park, CA: Sage.
Torgerson, Warren S. (1958). *Theory and methods of scaling.* New York: Wiley.

VII.B. HOW DO I CHOOSE THE OPTIMAL NUMBER OF CLUSTERS IN CLUSTER ANALYSIS?

I have used k-means clustering in an application of benefit segmentation. Although I am most satisfied with the technique, I still find it very hard to decide on the number of clusters to retain. I checked my clusters on external variables (sociodemographics) but still felt that this was not the ideal solution. Could you recommend a test or a technique that would be useful to determine the ideal number of clusters? (I use Statistical Procedures for the Social Sciences.)

The following is a related question from another researcher: When performing cluster analysis (e.g., in market segmentation), the marketing analyst is faced with the problem of deciding on an appropriate number of clusters to be discovered in a data set. To support this decision, several heuristics have been proposed so far. However, most of them are purely statistically motivated. Are there any information measures that combine statistics with soft facts like managerial appropriateness of cluster solutions?

Professor Phipps Arabie
Rutgers University

It is unlikely that a general, analytic solution will ever be found for the right number of clusters for most empirical data sets. The best analytic work on the topic is nicely summarized by Milligan (1996, pp. 361–365), along with a lot of other practical and useful information on clustering.

Concerning the specific use of clustering in market segmentation, Arabie and Hubert (1994) stated that

> An alternative, solution by fiat, is suggested by de Kluyver and Whitlark (1986, p. 280), who note that "To be managerially relevant, the number of clusters must be small enough to allow complete strategy development. At the same time, each cluster or segment should be large enough to warrant such strategic attention and to be reachable, and defensible against competitors." (p. 177)

REFERENCES

Arabie, Phipps, & Hubert, Lawrence. (1994). Cluster analysis in marketing research. In Richard P. Bagozzi (Ed.), *Advanced methods in marketing research* (pp. 160–189). Oxford, England: Blackwell.
de Kluyver, Cornelis A., & Whitlark, David B. (1986). Benefit segmentation for industrial products. *Industrial Marketing Management, 15,* 273–286.
Milligan, Glenn W. (1996). Clustering validation: Results and implications for applied analyses. In Phipps Arabie, Lawrence Hubert, & G. DeSoete (Eds.), *Clustering and classification* (pp. 341–375). River Edge, NJ: World Scientific.

Professor Laura Koehly
University of Iowa

Clustering methods attempt to divide a set of stimuli into categories such that those objects within a category are internally cohesive or similar to one another, and different clusters of objects are externally isolated or different from one another. There are two primary classes of clustering techniques: hierarchical methods and nonhierarchical methods. Hierarchical clustering methods investigate the stimulus structure at several different levels. Each cluster of objects displayed at some level is wholly contained, or nested, within a single cluster at a higher level. Nonhierarchical methods, however, do not impose a tree-like structure on the stimuli; rather, nonhierarchical methods partition the objects into a specified number of cohesive clusters. Theoretical considerations may suggest that one type of clustering technique is more appropriate for your research context than another. However, using both types of methods may help to clarify the solution. For an introduction to clustering methods and applications, refer to Gordon (1981), Hartigan (1975), Jain and Dubes (1988), or Sharma (1997).

For nonhierarchical clustering methods, two indexes can be used to compare cluster solutions: (a) the pooled within-cluster standard deviation and (b) the squared multiple correlation (R^2). These indexes are used as heuristics; inferential methods are lacking because the sampling distribution for these indexes are unknown. The pooled within-cluster standard deviation is an index of cluster cohesion. A cohesive set of clusters will provide small within-cluster standard deviations. Thus, in general, we aim for a cluster solution in which the pooled within-cluster standard deviation is relatively small, indicating cohesive, homogeneous clusters. The R^2 indicates the extent to which clusters are different from each, or externally isolated. If it is close to zero, it suggests that there is no difference between the clusters; however, an R^2 close to one indicates that a large proportion of the variation in our stimuli can be explained by cluster membership, suggesting a maximal difference between groups. The squared multiple correlation will necessarily increase as the number of clusters increases; similarly, the pooled within-cluster standard de-

viation will decrease as the number of clusters increases. Therefore, when comparing cluster solutions we need to evaluate whether the change in these indexes is significant. Unfortunately, there are no inferential procedures to guide us. Sharma (1997) suggested using plots, where the horizontal axis represents the number clusters in the solution and the vertical axis represents the index. We use this plot as we would a scree plot in factor analysis, where an elbow indicates the best solution.

Nonhierarchical methods can be sensitive to the initial configuration or starting clusters. To minimize this problem, a hierarchical clustering procedure can be employed prior to using a nonhierarchical method. Based on the hierarchical clustering solution, we can decide on a particular number of clusters and use the derived partition as the initial configuration in the nonhierarchical algorithm. Thus, we can view the hierarchical and nonhierarchical methods as complementary techniques that are used in conjunction with each other to arrive at a solution. With hierarchical procedures, in addition to the indexes mentioned earlier, the semipartial correlation coefficient and agglomeration schedule will provide some indication as to the best solution. The semipartial correlation can be viewed as an index of "loss of cohesion." As the number of clusters decreases, the variance in our data explained by cluster membership will decrease, indicating a loss of cohesion. The semipartial correlation will indicate whether there is a significant loss of information when two clusters are merged. The agglomeration schedule will also provide diagnostic information. The agglomeration schedule indicates the distance between mergers of clusters. Large distances in the agglomeration schedule suggest that the merged clusters are dissimilar, indicating a lack of cohesion between the two sets of stimuli.

Another option is to use "Isodata" (Ball & Hall, 1964), "Cluster" (Dubes & Jain, 1976), or "Groupals" (van Buuren & Heiser, 1989), three nonhierarchical clustering programs that find optimal solutions. Isodata will split a cluster if it has an unusually large variance, and it will merge two clusters if their cluster centroids are sufficiently close. Cluster finds and reports the best partitions in 1 through K clusters, where the clusters are not necessarily nested within each other. The advantage of Cluster is that it allows for easy comparisons across clustering solutions when the number of partitions has been changed. Groupals allocate stimuli into K clusters while simultaneously obtaining an optimal scaling of the original variables. Thus, the objects are optimally scaled into a lower dimensional space (e.g., homogeneity analysis; Gifi, 1990) and grouped into cohesive clusters based on their distances in this spatial representation. The Groupals program allows for measurements on the nominal, ordinal, or interval scales.

Ideally, theory will guide us as to the number of clusters underlying our stimuli and the location of each stimulus within a given cluster. Permutation tests can be used to compare a theoretical partitioning to an empirical cluster solution (Hubert & Schultz, 1976), and discriminant analyses can be used to validate a given cluster structure. However, if the analysis is more exploratory in nature, a hierarchical clustering method should be used in conjunction with nonhierarchical methods. Initially, a hierarchical solution is obtained, and the pooled within-cluster standard deviation, squared multiple correlation, semipartial correlation, and agglomerative distance are used to decide on the number of clusters needed to adequately explain the variability in the data. The K cluster hierarchical clustering solution can then be used as the initial configuration in a K-means clustering.

REFERENCES

Ball, G. H., & Hall, D. J. (1964, September). *Some fundamental concepts and synthesis procedures for pattern recognition preprocessors*. Paper presented at the International Conference on Microwaves, Circuit Theory, and Information Theory, Tokyo.

Dubes, Richard, & Jain, Anil K. (1976). Clustering techniques: The user's dilemma. *Pattern Recognition, 8,* 247–260.

Gifi, Albert. (1990). *Nonlinear multivariate analysis*. Chichester, England: Wiley.

Gordon, A. D. (1981). *Classification.* London: Chapman & Hall.

Hartigan, John A. (1975). *Clustering algorithms*. New York: Wiley.

Hubert, Lawrence J., & Schultz, James. (1976). Quadratic assignment as a general data-analysis strategy. *British Journal of Mathematical and Statistical Psychology, 29,* 190–241.

Jain, Anil K., & Dubes, Richard. (1988). *Algorithms for clustering data.* Englewood Cliffs, NJ: Prentice Hall.

Sharma, Subhash. (1997). *Applied multivariate techniques*. New York: Wiley.

van Buuren, Stef, & Heiser, Willem J. (1989). Clustering N objects into K groups under optimal scaling of variables. *Psychometrika, 54,* 699–706.

Professors Eric Bradlow and
Wes Hutchinson
University of Pennsylvania

Determining the right number of clusters for k-means clustering is a classic problem and continues to be an active area of research published in journals such as *Psychometrika* and *Journal of Classification.* Most statistical packages compute a variety of statistics for each k-means solution and at a minimum report the between-cluster (i.e., explained) and within-cluster (i.e., unexplained or error) variances. One of the oldest approaches is to compute a pseudo *F* statistic from the variances that compares a solution with k + 1 clusters to one with k clusters. In his classic book, Hartigan (1975) recommended, as a rule of thumb, that this pseudo *F* be at least 10 if an additional cluster is to be used. We believe that such rules of thumb are useful—particularly if there is no theoretical or substantive reason to search for (or believe in) the right number of clusters, but the researcher simply wants a standard criterion that prevents reviewers from questioning the objectivity of the decision. However, in most cases, k-means is used as an exploratory tool and not as a means of statistically testing hy-

potheses or for confirming specific predictions. As such, there is no universal rule for determining the number of clusters. It depends on what the researcher learns from each analysis. One noteworthy risk of using the pseudo *F* criterion is that k-means frequently identifies outliers in the data by defining clusters with only one or two observations. The incremental variance accounted for by such clusters may be small, but it can be very important to search for the cause of these anomalous observations and, in some cases, for deciding whether these observations should be removed from subsequent analyses. Milligan (1980) reported the results of a very nice Monte Carlo study that reveals the strengths and weaknesses of k-means clustering. A classic reference on marketing applications of cluster analysis is Punj and Stewart (1983).

REFERENCES

Hartigan, John A. (1975). *Clustering algorithms*. New York: Wiley.

Milligan, Glenn W. (1980). An examination of the effect of six types of error perturbations on fifteen clustering algorithms. *Psychometrika, 45,* 325–342.

Punj, Girish, & Stewart, David W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research, 20,* 134–148.

VII.C. META-ANALYSIS

For a meta-analysis, how should one decide if the same independent variable was used in all studies? When one is studying the effect of a drug, it is relatively easy (although, one could have variants in the same class of drug when one compares two classes of drugs). However, in consumer experiments, all the experimenters may state that they manipulated the same construct, but they actually manipulated the construct in two different ways (e.g., physician courtesy toward patients, print vs. video), and the type of manipulation produces differences in the results. (One could hypothesize that video provides nonverbal cues so that it is a different and more effective manipulation of the construct). Also, experimenters may state that they manipulated different constructs (e.g., "patient involvement with treatment decisions" and "patient participation with treatment decisions"; a distinction has been made between the two in the literature), but the manipulations may be quite similar, and the effects of the manipulations of the two constructs may also be quite similar. Finally, consider the case in which one has two distinct constructs ("patient participation with treatment decisions" and "physician courtesy toward patients") that could be aspects of a meta-construct ("interpersonal quality of care") and two distinct manipulations of each of these constructs. When would it be reasonable to aggregate the two sets of studies and claim that one is interested in examining the effects of the meta-construct?

Professor Donald Lehmann
Columbia University

One problem faced by meta-analysts is how tightly to define the design variables used to explain variation in the dependent variable of interest (e.g., see Farley & Lehmann, 1986; Farley, Lehmann, & Sawyer, 1995). For example, in an advertising meta-analysis, advertising has been treated as aggregate Gross Rating Points or at most by media type (e.g., TV, print), ignoring the impact of specific media vehicles.

At least three concerns enter into the decision to use broad definitions: (a) the limited number of observations (which leads to aggregation to minimize the number of variables and increase the observations for each level of the variable), (b) reporting standards that often make precise definitions impossible and force the use of general definitions, and (c) the basic principle behind meta-analysis (i.e., the belief that there are generalizations possible and that these occur across minor variations in variables). The effect of aggregation is to increase the error term in meta-analysis. Given that meta-analysis on aggregate (general) variables often explains 40% to 50% of the variance, however, this has not proven critical.

Of course, one can use more precise definitions, but then sample sizes for the included levels decrease and confounding (collinearity) among particular combinations tends to increase, making estimation difficult. Added to the fact that precise definitions tend to be correlated with a number of variables not included in the design matrix (e.g., author training and identity), interpretation of the variables' coefficients is problematic.

Essentially, the decision to use broad definitions rests on the prior belief about whether fine distinctions in variables (or estimation method, measurement, etc.) matter. If one takes the position that they do not (or at least that the fine distinctions have less impact than the major variables in the meta-analysis), then given the weak statistical power of most meta-analyses, it is unlikely the hypothesis of no difference will be rejected. On the other hand, if someone is convinced the distinctions matter, then they can code the variables in greater detail and include them in the analysis. Absent strong a priori reasons, however, it is probably better (and certainly consistent with the spirit of meta-analysis that searches for generalizations) to assume (implicitly) that fine distinctions are not crucial until and unless far greater numbers of studies appear in a given area than are currently available.

REFERENCES

Farley, John U., & Lehmann, Donald R. (1986). *Meta-analysis in marketing: Generalization of response models*. Lexington, MA: Lexington.

Farley, John U., Lehmann, Donald R., & Sawyer, Alan. (1995). Empirical marketing generalization using meta-analysis. *Marketing Science, 14,* G36–G46.

Professor Lawrence Hedges
University of Chicago

The question raised was, generally, how to decide if two different studies are actually manipulating the same independent variable. This fundamental question appears to be one of construct breadth—that is, it is equivalent to the question of how broad a construct is desired. This is a fundamental question of interpretation, to which there is no single right answer. It all depends on how broad the conclusions you wish to make may be (Hall, Rosenthal, Tickie-Degnen, & Mosteller, 1994; Matt & Cook, 1994).

Individual primary research studies typically operationally define constructs rather narrowly, however broadly they may be defined theoretically. Experiments using narrowly defined constructs are therefore subject to the threat to construct validity that Cook and Campbell (1979) called "mono-method bias." That is, because the operationalization of the construct in most primary studies is narrower than the intended construct, causal effects may be due to irrelevant aspects of the operationalization, rather than to the intended construct. Another way to put it is that experiments with narrow construct operations can only provide empirical evidence about rather narrow versions of constructs. The usual interpretations involve broader construct definitions, but there is, strictly speaking, no empirical warrant to generalize beyond the particular narrow construct definition.

To belabor the point slightly, consider the example of stimulus sets (e.g., word lists or context in which a social response is elicited) in experiments. Strictly speaking, the empirical results apply only to the stimulus sets used in the experiment. If you really want to generalize to other stimuli (and you think the stimulus set makes a difference), you should sample a set of stimulus sets and make them a random effect in the experiment. This issue has, of course, been discussed extensively in some areas of psychology (e.g., Herb Clarke's, 1973, article on the language as fixed-effects fallacy).

Meta-analyses have the potential to provide empirical evidence about broader construct definitions. To be sure, meta-analyses can function perfectly well using narrow construct definitions. One of the important benefits of meta-analysis, however, is that it makes it possible to obtain empirical evidence on broader construct definitions (by deciding that a broader array of operations are acceptable as exemplifying of the construct). By obtaining evidence on a broader construct, the conclusions based on the empirical evidence are less subject to mono-operational bias.

Therefore, it is possible, for example, to see if patient participation and involvement has an effect on outcomes. It is also possible to see how much variation in outcomes is likely to be due to differences in the particular operationalization of patient involvement used. If there is a great deal of variation of results across operationalizations of the construct, then generalizations must be stated conditionally. That is, one may be able to say that you get an effect of a particular type of patient involvement but not necessarily other types. If there is relatively little variation in outcome across different operations, there is less need to state generalizations conditionally (e.g., this particular type of patient involvement in decisions seems to help vs. many varieties of patient involvement seems to help and it does not matter which). By starting with an initially rather specific coding scheme to capture variation in operationalizations (ideally one based on theory), the meta-analyst preserves the option of using a narrower scheme or a broader one if that is empirically justifiable.

For example, Devine and Cook (1983) investigated the effects of various kinds of psychoeducational interventions on the post operative recovery of surgical patients. They coded specific kinds of interventions (operationalizations of the broad construct of psychoeducational care): breathing or coughing exercises, counseling that patients should expect to feel pain, and so on. They also coded who gave the care (doctors, nurses, others, etc.), the length of the treatment, and so on. All of these aspects of the construct could have mattered in the sense that some could have worked much better than others. The meta-analysis showed, however, that there was very little difference in treatment effect, regardless of which operationalization was used—the treatment had a substantial positive effect, shortening recovery time.

REFERENCES

Clarke, Herb H. (1973). The language-as-fixed-effec t fallacy: A critique of language statistics in psychological research. *Journal of Verbal and Learning Behavior, 12,* 335–359.

Cook, Thomas D., & Campbell, Donald T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Devine, Elizabeth C., & Cook, Thomas D. (1983). A meta-analytic analysis of effects of psycho-educationa l interventions on length of post-surgical hospital stay. *Nursing Research, 32,* 267–274.

Hall, Judith A., Rosenthal, Robert, Tickie-Degnen, Linda, & Mosteller, Frederick. (1994). Hypotheses and problems in research synthesis. In Harris Cooper & Lawrence V. Hedges (Eds.), *The handbook of research synthesis* (pp. 17–28). New York: Russell Sage Foundation.

Matt, George E., & Cook, Thomas D. (1994). Threats to the validity of research syntheses. In Harris Cooper & Lawrence V. Hedges (Eds.), *The handbook of research synthesis* (pp. 503–520). New York: Russell Sage Foundation.

Editor: Additional useful references on meta-analysis, including several by the modest Hedges and Lehmann follow.

REFERENCES

Farley, John U., Lehmann, Donald R., & Mann, Lane H. (1998). Designing the next study for maximum impact. *Journal of Marketing Research, 35,* 496–501.

Fern, Edward F., & Monroe, Kent B. (1996). Effect-size estimates: Issues and problems in interpretation. *Journal of Consumer Research, 23,* 89–105.

Hedges, Larry V. (1982). Statistical methodology in meta-analysis. Princeton, NJ: Educational Testing Service.
Hedges, Larry V. (1992). Meta-analysis. *Journal of Educational Statistics, 17,* 279–296.
Hedges, Larry V., & Olkin, Ingram. (1983). Clustering estimates of effect magnitude from independent studies. *Psychological Bulletin, 93,* 563–573.
Hedges, Larry V., & Olkin, Ingram. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic.
Hedges, Larry V., & Vevea, Jack L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.
Li, Zhaohai, & Begg, Colin B. (1994). Random effects models for combining results from controlled and uncontrolled studies in a meta-analysis. *Journal of the American Statistical Association, 89,* 1523–1527.
Schmidt, Frank L. (1992). What do data really mean? Research in findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist, 47,* 1173–1181.
Sorić, Branko. (1989). Statistical "discoveries" and effect-size estimation. *Journal of the American Statistical Association, 84,* 608–610.

VII.D. COMBINING DATA ON INDIVIDUALS AND GROUPS, NETWORKS AND DYADS

My methodological question concerns comparisons between group-level and individual-level observations in an experimental (repeated measures) design. To give a bit of context to the question, my research investigates (a) group influences on the alternatives considered by individual members for a group choice and (b) group influences on individual problem solving (e.g., sorting and categorization). Collecting repeated observations from each participant allows one to make both within-subjects comparisons and between-subjects comparisons. Although collecting both individual-level and group-level observations has precedent in the group polarization literature (e.g., Laughlin & Earley, 1982; Stasser & Titus, 1985), I have not been satisfied with the methods I have seen used to compare individual-level and group-level observations.

At least three comparisons seem relevant:

1. Comparison of two individual-level observations (for the same individual) in the absence of group interaction. For example, two questions might be posed to the same individual: (a) Which alternatives from this set would you consider if you were making a choice for yourself?, and (b) Which alternatives from this set would you consider if you were making a choice affecting the entire group of which you are a member? In this case, it seems fairly straightforward to compute difference scores or to use a traditional repeated measures design, because there is no intervening group interaction.

2. Comparison of two individual-level observations (for the same individual) before and after group interaction, as is common in the group polarization literature. For between-subjects comparisons, it has been suggested quite correctly by Myers, DiCecco, and Lorch (1981) that the performances of participants who have participated in a group interaction cannot be validly compared with the performances of individuals who have not participated in a group interaction. Their solution is to create 'pseudo groups' of individuals (averages of the individual scores) who have not participated in a group interaction. Should this procedure be generalized to the case in which repeated observations are collected from the same individuals, as in Stasser and Titus (1985)? One major drawback of this approach is the loss of power associated with aggregating the individual observations for each group.

3. Comparison between one individual-level observation and one group-level observation (in which the same individual contributes to both observations). For example, an individual might complete one problem-solving task individually and a second (counterbalanced) problem-solving task as part of a group. I would like to compare these observations using a repeated measures analysis of variance (ANOVA). Am I limited to using averages of individual scores for each group as one of the repeated measures (with the second being the group score), or can the ANOVA be structured to allow several individual member observations to be compared with a single observation for each group? Again, the loss of power resulting from aggregating individual observations is a crucial consideration.

REFERENCES

Laughlin, Patrick R., & Earley, P. Christopher. (1982). Social combination models, persuasive arguments theory, social comparison theory, and choice shift. *Journal of Personality and Social Psychology, 42,* 273–280.
Myers, Jerome L., DiCecco, Joseph V., & Lorch, Robert F. (1981). Group dynamics and individual performances: Pseudo-group and quasi-F analyses. *Journal of Personality and Social Psychology, 40,* 86–98.
Stasser, Garold, & Titus, William. (1985). Pooling of unshared information in group decision-making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48,* 1467–1478.

Editor: Research that compares group dynamics to individual processing is fascinating, but yes, a killer methodologically. Individual-level data can be compared cleanly, as per your question. Group-level indexes (e.g., performance) can also be compared cleanly, though the group serving as the unit of analysis requires numbers of participants multiplied by group size, which can be cumbersome (otherwise you confront the "reduced power" issue to which you refer when considering aggregating individual-level data to group-level data). Some researchers certainly proceed with aggregate functions (your pseudo groups), but this solution is not particularly satisfactory, given that it is not descriptive of the means by which the data were obtained. Furthermore, a mean computed to describe a pseudo group will be more stable (smaller standard error) than a single index obtained on a real group, so although superficially it may seem that one is comparing apples to apples, statistically it may be fruit salad. Researchers have also tried using dummy variables to represent group

membership—this practice may work reasonably well if the group sizes are fairly small and if much more information is included in the model that represents the individual-level data than the group-level data, otherwise one is simply creating large multicollinearity problems.

A modeling approach that may interest you is called hierarchical linear models (HLM; cf. Kreft & de Leeuw, 1998). These models are relatively new, but they are based on classic approaches. The purpose of this class of models is to integrate data at more than one level of observation or analysis. Much of this work has been developed to assess educational systems, where the data contain measures on pupils within schools and also schools within districts, for example (Bock, 1989; Bryk & Raudenbush, 1992; Raudenbush & Willms, 1991). Essentially, they are fitting a mixed regression model, with random and fixed effects.

For example, your individuals-within-groups-level data would be modeled as (Bryk & Raudenbush, 1992, pp. 208–209)

$$Y_i = X_i\beta_i + e_i$$

where Y_i is an n_i vector containing the scores on the dependent variable for the n_i individuals in Group i ($i = 1 \ldots k$). X_i is the $n_i \times p$ matrix containing the p explanatory variables on each of the n_i individuals; β_i is $p \times 1$, the usual vector of regression parameters to be estimated, for Group i; and e_i is the $n_i \times 1$ vector of random errors for the individuals in that Group i.

The groups-level data would be expressed by treating the betas in the previous model as a random variable to be modeled as a function of group-level characteristics:

$$\beta_i = X_i^*\beta_i^* + e_i^*,$$

where the matrices X_i^*, β_i^*, and e_i^* denote the analogs to the independent variables matrix and the parameter estimates and error vectors, respectively, for the group-level information. In this manner, both levels, group and individual, are modeled simultaneously.

REFERENCES

Bock, R. Darrell. (Ed.). (1989). *Multilevel analysis of educational data.* San Diego, CA: Academic.

Bryk, Anthony S., & Raudenbush, Stephen W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Kreft, Ita G. G., & Leeuw, Jan de. (1998). *Introducing multilevel modeling.* Thousand Oaks, CA: Sage.

Raudenbush, Stephen W., & J. Willms, Douglas. (Eds.). (1991). *Schools, classrooms, and pupils: International studies of schooling from a multilevel perspective.* San Diego, CA: Academic.

A related question follows: When researching a dyadic relationship, such as retailer–consumer, one sometimes wishes to interview both sides of the dyad and use the dyad as the unit of analysis. Using the dyad as the unit of analysis allows one to test hypotheses regarding, for example, the effect of employee training programs or other store policies on customer satisfaction. However, one will frequently have data in which there are several hundred dyads, but one side of the dyad is in common among many of the units. This can occur in contexts such as parent–child (parent with multiple children), as well.

This condition reduces variability in the measures of one side of the dyads and also may violate independence assumptions made in linear models. My question is, is this really a matter for concern in regression or structural equation modeling, where one may be using constructs on one side of the dyad as independent variables and constructs on the other side as dependent variables? What are the concerns, if any, and is there a solution?

Professor Laura Koehly
University of Iowa

The situation that you have described suggests a particular form of cluster sampling in which an individual, or ego, is sampled and then individuals, or alters, who are relationally tied in some way to the ego are also sampled; thus, the alters are nested within the ego. For example, one might gather data on managers in a large organization and each manager's clerical employees. Members of the clerical staff are tied to their supervisors, the managers, through their work relations. Generally, this type of sampling design originates from research questions that attempt to account for the social influence of relationally tied individuals on respondents' attitudes, beliefs, or behaviors.

There are several ways to think about these data. First, let us assume that the egos are independent. Therefore, there is no overlap among ego's alters, and the egos are independently randomly sampled. Therefore, the dependency between respondents exists only among the alters of a given ego. The issues that one should be sensitive to differ depending on which side of the dyad is treated as the independent variable. If the dependent variable is some attribute of the ego and the independent variables are attributes of the alters, then one's primary concerns are collinearity problems and model specification issues. If, however, the dependent variable is some attribute of the alters and the independent variable is an ego attribute, then the independence assumption underlying most statistical tests will be violated. This can lead to extremely erroneous results, where the parameters are significantly biased or the null hypothesis is rejected at rates significantly lower than the nominal level.

There are several approaches one could take with these type of data. First, the multilevel modeling or HLM methodology

would fit the framework defined in your question well. The HLM approach will allow the investigator to pose hypotheses about relations occurring at each level in a hierarchy of models (e.g., among alters, among egos) and across levels (between alters and egos). In this situation, the responses of the alters would represent the Level 1 model and the egos' responses would determine our Level 2 model. The egos may represent a random effect, depending on the sampling design. For a general introduction to HLM, see Bryk and Raudenbush (1992). An application of multilevel modeling to social relations data is discussed by Snijders, Spreen, and Zwaagstra (1995).

Another approach is the use of social influence models or network autocorrelation models; Marsden and Friedkin (1994) and Doreian (1989) provided a nice overview of social influence models. Ideally, we have some measure of relational ties between actors; examples of relationships include work-related communications, friendship relationships, and advice-seeking relationships. These relational ties are then converted into a weight matrix that represents a measure of interdependency among respondents. In the model specification, an individual's influence on his or her colleagues is weighted by the degree of dependency in the dyad as represented in the weight matrix. The advantage of social influence models is that (a) the strength of the dependency between respondents can differ across dyads, and (b) dependencies between egos can easily be incorporated into the model. For example, the managers and clerical staff sampling scenario presented earlier would suggest that the managers are not independent given that they work in the same organization. This dependency could easily be accounted for in a social influence model. Also, it is possible to model attributes of both egos and alters as both the dependent variable and independent variables in a regressive–autoregressive model specification. For example, a manager's efficiency is likely to be a function of both their supervisor's efficiency and the efficiency of the clerical staff with whom they work.

To summarize, yes, there are several concerns that we need to be sensitive to when there are interdependencies among our observations. These concerns vary depending on which side of the dyad is treated as the dependent variable. The HLM approach will account for the clustered nature of the data by modeling the alters as one level and the egos as another. Social influence models, however, provide a more general model, where attributes of both the egos and the alters can be represented as the dependent variable and as independent variables in a regression–autoregression equation. Furthermore, dependencies among the egos can easily be accommodated in a social influence model.

REFERENCES

Bryk, Anthony S., & Raudenbush, Stephen W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Doreian, Patrick. (1989). Network autocorrelation models: Problems and prospects. In D. Griffith (Ed.), *Spatial statistics: Past, present, future* (pp. 369–389). Ann Arbor, MI: Document Services.

Marsden, Peter V., & Friedkin, Noah E. (1994). Network studies of social influence. In Stanley Wasserman & Joseph Galaskiewicz (Eds.), *Advances in social network analysis* (pp. 3–25). Thousand Oaks, CA: Sage.

Snijders, Tom, Spreen, Marinus, & Zwaagstra, Ronald. (1995). The use of multilevel modeling for analysing personal networks: Networks of cocaine users in an urban area. *Journal of Quantitative Anthropology, 5,* 85–105.

Editor: Snijders's (University of Groningen) work is a nice blend of the egocentric networks and the HLM to which Professor Koehly refers (e.g., Van Duijn, van Busschbach, & Snijders, 1998). Snijders and Kenny (1998) used Kenny's social relation model to consider dyadic relations between individuals who are nested in groups, such as members of a family. The traditional HLM models integrate data at more than one level of observation and analysis. Snijders and Kenny noted that for studies of relational (e.g., dyadic) data, the relational connection is the point of focus, so that in a family setting, there are several possibly relevant units of analysis: the family, each family member, and the dyadic exchanges among the family members. Thus, egocentric networks are even more complicated than HLMs given their interdependencies. Snijders has also used these models to examine longitudinal data, to track structural and membership changes in personal networks as people experience such life events as the birth of the family's first child, relocation, and retirement.

You may wish to consider investigating network models and models for dyadic interactions (e.g., Iacobucci, 1996; Iacobucci & Hopkins, 1992; Iacobucci & Wasserman, 1988; Knoke & Kuklinski, 1982) and those for multiple informants (e.g., Anderson & Narus, 1990).

REFERENCES

Anderson, James C., & Narus, James A. (1990). A model of distributor firm and manufacturer firm working partnerships. *Journal of Marketing, 54,* 42–58.

Iacobucci, Dawn. (Ed.). (1996). *Networks in marketing.* Thousand Oaks, CA: Sage.

Iacobucci, Dawn, & Hopkins, Nigel. (1992). Modeling dyadic interactions and networks in marketing. *Journal of Marketing Research, 29,* 5–17.

Iacobucci, Dawn, & Wasserman, Stanley. (1988). A general framework for the statistical analysis of sequential dyadic interaction data. *Psychological Bulletin, 103,* 379–390.

Knoke, David, & Kuklinski, James H. (1982). *Network analysis.* Beverly Hills, CA: Sage.

Snijders, Tom A. B., & Kenny, David A. (1998). *Multilevel models for relational data.* Manuscript submitted for publication.

Van Duijn, Marijte A. J., van Busschbach, Jooske T., & Snijders, Tom A. B. (1998). *Multilevel analysis of personal networks as dependent variables.* Manuscript submitted for publication.

Copyright of Journal of Consumer Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.