

CAM JOURNAL

Cultural Anthropology Methods

Volume 9, Issue 1 • February 1997

Welcome to CAM

CAM is where social scientists discuss how to *use* and *teach* qualitative and quantitative research methods.

CAM costs \$18 USC per year, including postage, for three issues sent anywhere in the U.S. For Canada and Mexico, add \$5.00 postage. For all other countries, add \$10.00. Purchase orders, add \$3.00. Florida residents, please add 6% sales tax. Send your check to CAM, 5246 NW 47th Lane, Gainesville, FL 32606. For credit card orders, visit: www.analytictech.com

Submitting articles: Send hard copy and disk to CAM, 5246 NW 47th Lane, Gainesville, FL 32606. For more information, or for enquiries about submitting an article, contact any editor of CAM:

Editors

H. Russell Bernard (U Florida)
ufruss@nervm.nerdc.ufl.edu
Stephen Borgatti (Boston C)
steve_borgatti@msn.com
Pertti Pelto
efbp.baroda@smr.sprintrpg.ems.vsnl.net.in
Gery Ryan (UCLA)
gryan@ucla.edu
Oswald Werner (Northwestern U)
ossy@aol.com

Associate Editors

Candice Bradley (Lawrence U)
candice.bradley@lawrence.edu
W. Penn Handwerker (U Connecticut)
handwerk@uconnvm.uconn.edu
A. Michael Huberman
huberman@uni2a.unige.ch
Thomas Schweizer (U Cologne)
thomas_schweizer@uni-koeln.de
Eben Weitzman (UMass-Boston)
weitzmane@umb.umb.edu
Barry Wellman (U Toronto)
wellman@epas.utoronto.ca

Managing Editor

Carole M. Bernard (ECS)

CAM is a product of Editorial Consulting Services (ECS) of Gainesville, FL. CAM is issued three times a year, in February, June, and October. ISSN 1064-8631.

Developing Categories for Interview Data: Consequences of Different Coding and Analysis Strategies in Understanding Text: Part 2

Karen A. Jehn and Lorna Doucet
The Wharton School of the
University of Pennsylvania
jehn@wharton.upenn.edu

Introduction

We are interested in conflicts in U.S.-foreign joint ventures from the point of view of the people who live them. In our current study, we compare two inductive techniques for categorizing two kinds of conflict situations. We use factor analysis and multidimensional scaling to categorize same-culture and different-culture conflict situations experienced by American managers abroad.

Our design compares differences in measuring and coding for similarity across conflict situations (scenarios). Using the same data (scenarios), we converted the data to two different types of matrices (similarity-MDS, profile-factor) using different types of judgment tasks. Then we used different analytic techniques (factor analysis and MDS) on the matrices to identify potential patterns.

The Data

We interviewed 76 American managers who had recently worked in or were cur-

rently working in Sino-American joint ventures. We asked them to describe two conflict situations in which they were involved; one with a same-culture manager (intracultural process) and one with a different-culture manager (intercultural process). The interviews were done in person in the People's Republic of China (n = 45) or over the phone (n = 31). Hence, 76 same-culture conflict scenarios and 76 different-culture conflict scenarios were collected (n = 152 total scenarios).

In providing each conflict scenario description, the interviewer made sure that the manager covered the following five questions: (1) "What is the nature of your relationship with this person (e.g., peer, superior, or subordinate; same or different department)?" (2) "Who else was involved?" (3) "What was the conflict about?" (4) "What caused the conflict?" and (5) "Was the conflict resolved? If so, how? If not, what is its current status?" The conflict scenarios were transcribed for analysis and presented in Part 1 of this article (Jehn and Doucet 1996), where we showed how simple frequency counts can

tell us something of the salience of words and themes in a text.

While frequency counts provide a clear indication of the prominence of certain themes in the conflict scenarios, they do not take advantage of contextual information to clarify how the terms are being used. For example, frequency counts may indicate that a particular aspect of conflict (i.e., aggressiveness, shouting, negotiating) is quite salient, without revealing whether it is typical or not, good or bad, intense or mild, etc. To answer these kinds of questions, we used factor analysis and multidimensional scaling for a broader perspective on the categories of conflict.

Factor Analysis—Item Ratings

Each scenario was rated on a wide range of aspects about conflict. Once these ratings were made, factor analysis of the ratings were used to search for meaningful groups or categories of conflict in the two conditions (i.e., same-culture conflict and different-culture conflict). We had two raters rate the conflict scenarios.

Here are the actual instructions to the raters: "The following questions relate to the informant's description of the conflict. Please answer the following questions in terms of what the informant says about the conflict without trying to make judgments of whether the informant is right or wrong."

And here is an example of the aspects of conflict we asked the raters to answer on a 5-point scale (1 = "not at all" to 5 = "to a great extent") about the informant, the informant's co-worker, and the general situation:

To what extent did the informant describe the conflict as:

- _____ a. a contest, battle, clash struggle, fight, competition
- _____ b. about work
- _____ c. about rules, policies, authority
- _____ d. scheduling problem
- _____ e. emotions, feelings

The aspects of conflict used for rating were based on text analysis of the scenarios from

both conditions (Jehn and Doucet 1996). Note that the categories are not mutually exclusive. The raters evaluated 27 different aspects of conflict for each of the 152 scenarios. The raters achieved a high level of agreement and mutual understanding of terms, referring to a thesaurus or dictionary whenever they disagreed on the meaning of a term.

High interrater reliability was ensured by having the two raters periodically evaluate common scenarios (35 total) independently and compare their ratings. The raters rated four or five scenarios on their own and then 1 of the 35 that they rated in common. They would then compare ratings to make sure they still agreed how to rate the 27 questions about each scenario.

The final rating used for these commonly rated scenarios was the agreement reached after the raters met together and discussed discrepancies. We helped only when the raters could not reach agreement within one point on the scale (2.6% of all ratings). Rater 1 individually rated 70 scenarios; Rater 2 individually rated 47 scenarios for a total of 117 (117 + the 35 commonly rated = 152 total scenarios).

Using Factor Analysis

Factor analysis of the two sets of scenario ratings (two sets of 76 scenarios; 27 characteristics for each scenario) were then conducted and provided a way to examine the categories of conflict present in the intra- and intercultural scenarios. Scree figures indicate three-factor solutions for both the intra- and intercultural scenarios. For each factor analysis, the three factors had eigenvalues above 1.00. The three factors for the intracultural scenarios account for 36% of the variance: factor 1 = 21%, factor 2 = 11.5%, factor 3 = 3.5%. The three factors for the intercultural scenario account for 32% of the variance: factor 1 = 15.4%, factor 2 = 9.8%, factor 3 = 6.8%.

We can interpret the factors qualitatively and quantitatively. Qualitatively, we examined the factor loadings and determined differences between the intra and inter factor results. Items with loadings above .40 (a commonly used cutoff) on the first

intercultural factor reflect the amount of interpersonal animosity and hostility that is felt in the conflict situation.

Phrases included in scenarios that had high scores on this factor were: "There is a lot of hate involved in this situation," "The dislike is overwhelming," and "I was very angry." The second factor represents the aggravation present: "I was very frustrated with my co-worker" and "Their inconsistencies really aggravated me." The third factor indicates the volatile nature of conflict in intercultural situations: "She's a bitch," "We are constantly shouting and screaming," and "There are so many volatile outbursts when we get together."

Items with loadings above .40 on the first intracultural factor reflect hatred and animosity with a volatile nature: "Every time we fight it ends up in a shouting match" and "I just hate working with him!" The second factor is unique to this data set and represents conflict that is conducted calmly, is not verbally intense, and is often about value differences: "We just don't agree on some basic values" and "It's really a small difference based on preferences." The third factor is not well determined (only one item solely loaded above .40 on factor three).

If there were similar factors in the two analyses (intra- and intercultural conflict), we could create scales for the different factors by taking the means of the items that load on each factor and compare them by t-tests across the intra- and intercultural interviews. In this data set, however, the factors were different. This indicates that it is not appropriate to create comparative scales, and we therefore relied on the interpretive analysis. In the next section, we compare the results of the separate factor analyses to multidimensional scaling analyses (MDS) of the intra- and intercultural text files.

Collecting MDS Ratings

The factor analysis of the aspects of conflict identified groups or categories that were useful in describing the conflict scenarios (e.g., hostile, calm, volatile). Using MDS, we tried to identify relationships among the scenarios themselves (as opposed to

relationships between item ratings of scenarios).

This analysis was done on a subset of the scenarios (two sets of 30 scenarios). The subset of scenarios was determined by clarity ratings conducted by two raters blind to the conditions. The raters rated each scenario on a 5-point Likert-type scale (1 = not at all, 5 = very) on the following three questions: "How complete is this scenario?," "How clear is the information in this scenario?," and "How much irrelevant information is included in this scenario?" (The answers to the last question, of course, were reverse coded.) The mean clarity rating for the entire data set of scenarios was 3.67 (sd = 2.56), and the mean of the subset of 30 clear scenarios was 5.89 (sd = .80) for the *intra* scenarios and 4.87 (sd = 1.89) for the *inter* scenarios.

Once the subsets of scenarios were established, we asked a separate group of 50 American managers in the People's Republic of China to rate the similarity among conflict scenario descriptions provided by the American managers in the Sino-American joint ventures. The two subsets of conflict scenarios (i.e., inter- and intracultural) are treated separately at this stage. Each informant considered between 60 and 120 pairs of conflict scenarios and evaluated how similar/dissimilar they were.

We administered a questionnaire to the respondents and had them judge the degree to which each description was similar/dissimilar to a "target" description. Here are the actual instructions to the informants:

Please read each description and compare each one to the description on the bottom of this page (this is called the target description). That is, compare the behaviors in each description to those in the target and tell us how similar each description is to the target. Think about the content of the situation, the amount of emotion, and how the situation was handled or resolved. To make your judgment, use the scale shown below. A '1' means that the example is not at all similar to the target and a '9' means that they are very similar.

1 2 3 4 5 6 7 8 9
Not at all Very
Similar Similar

Each description was used as the target description (the scenario that every other scenario was compared to) at least four times. Similarity ratings were made on a nine-point Likert-type scale (see above). The rule of thumb is that each description should be used as a target description $n/10$ times. Therefore, in this example, each target was used at least 30/10, or 3 times. We found enough participants to use each target 4 times. This allowed eight comparisons of all possible combinations of the descriptions (4 times as a target and 4 times as an item when another scenario is the target).

Each manager completed the questionnaire for at least two randomly assigned target descriptions (i.e., 60 pairs). At the end of the survey, respondents were asked to list criteria they used to make their ratings (e.g., degree of anger, degree of resolution, amount of work versus nonwork focus).

Create a Similarity Matrix of Means for MDS

The following procedure was done independently for each of the two sets of conflict scenarios. For each set, we created a matrix of similarity ratings containing the mean rating of similarity between each pair of scenarios for the multidimensional scaling analysis. Classical nonmetric, ordinal multidimensional scaling analysis was used to determine the number of dimensions necessary to describe the similarities and differences among the descriptions. The MDS program in SAS (ALSCAL; Young and Lewyckyj 1979) maps the scenarios on the dimensions uncovered.

MDS maps a set of relations (like a matrix of similarities) so that the distance between points on the map reflect the similarities between items (scenarios) in the matrix. Scenarios correlated at .80 for example (quite similar), ought to be half the distance from each other as scenarios that are correlated at .40 (not so similar). With many similarities to account for, it may be impossible to map them onto just two

dimensions. Kruskal's (1964) measure of stress, or badness-of-fit, provides an index of how well any set of relations can be mapped onto n dimensions.

For the intracultural analysis the stress for a two-dimensional plot of relations among the objects (scenarios) was .29, the stress for a three-dimensional plot was .21, and the stress for a four-dimensional solution was .16. For the intercultural analysis the stress for a two-dimensional plot of relations among the objects (scenarios) was .30, the stress for a three-dimensional plot was .21, and the stress for a four-dimensional solution was .16. Thus, the stress tests indicate a four-dimensional solution for both the intercultural conflict scenarios and for the intracultural scenarios.

Label Categories of MDS Results

To determine potential labels for the categories, we examined the plots of the scenarios along the MDS dimensions and looked specifically at the scenarios that fell at the extremes of each dimension. There are six plots for four dimensions (1 and 2, 1 and 3, etc.) and with two studies (one for intercultural and one for intracultural conflict) there are 12 plots. We show 4 of the plots here to demonstrate the method. Figure 1 shows dimensions 1 and 2 for the intracultural analysis; Figure 2 shows dimensions 3 and 4 for the intracultural analysis; Figure 3 shows dimensions 1 and 2 for the intercultural analysis; and Figure 4 shows dimensions 3 and 4 for the intercultural analysis.

For example, in Figure 1, scenarios 26, 25 and 7 fall at one extreme of the horizontal axis, while, 9, 21, 8 and 27 fall at the other extreme. To label this axis, we sought to identify the criterion that distinguished scenarios 26, 25 and 7 from scenarios 9, 21, 8 and 27.

We relied on the criterion that the respondents provided in response to the open-ended question at the end of the similarity rating survey ("Please list the criteria you used to determine how the scenarios were similar or different from one another") to interpret the dimensions. We scanned these criteria, searching for the one

Figure 1: INTERcultural Conflict: Dimensions 1 and 2*

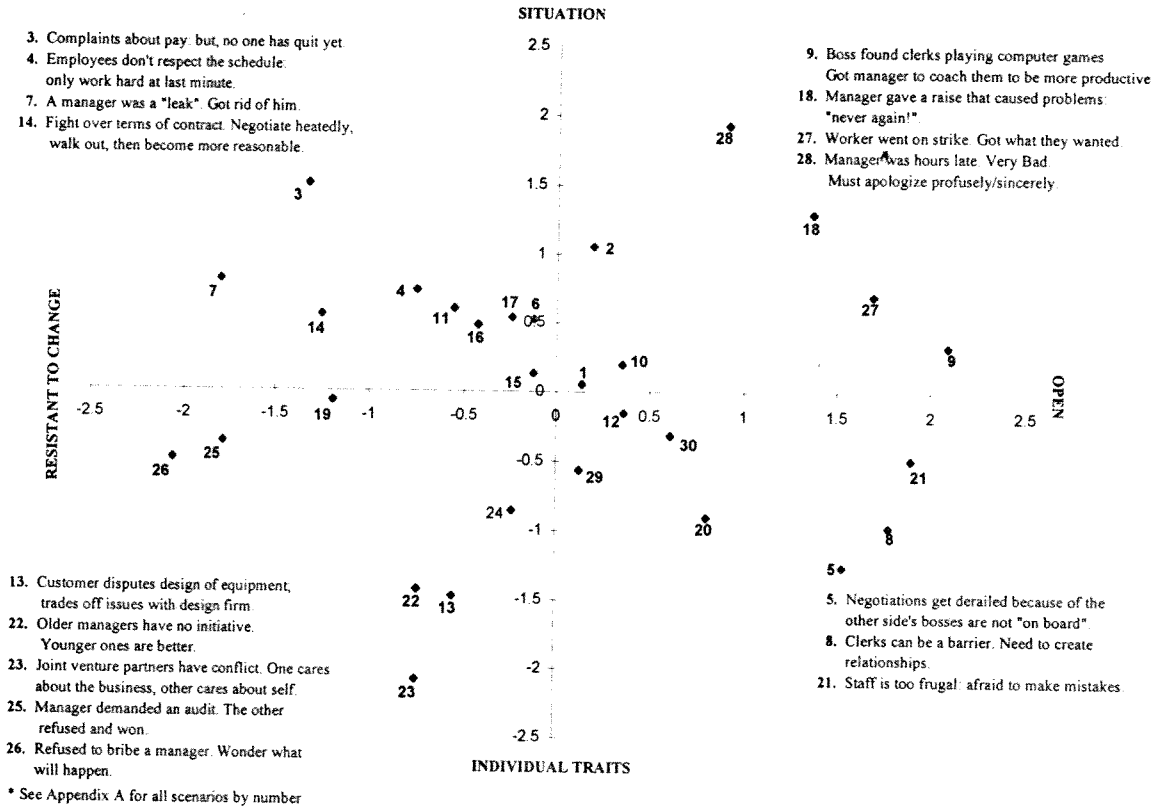


Figure 2: INTERcultural Conflict: Dimensions 3 and 4*

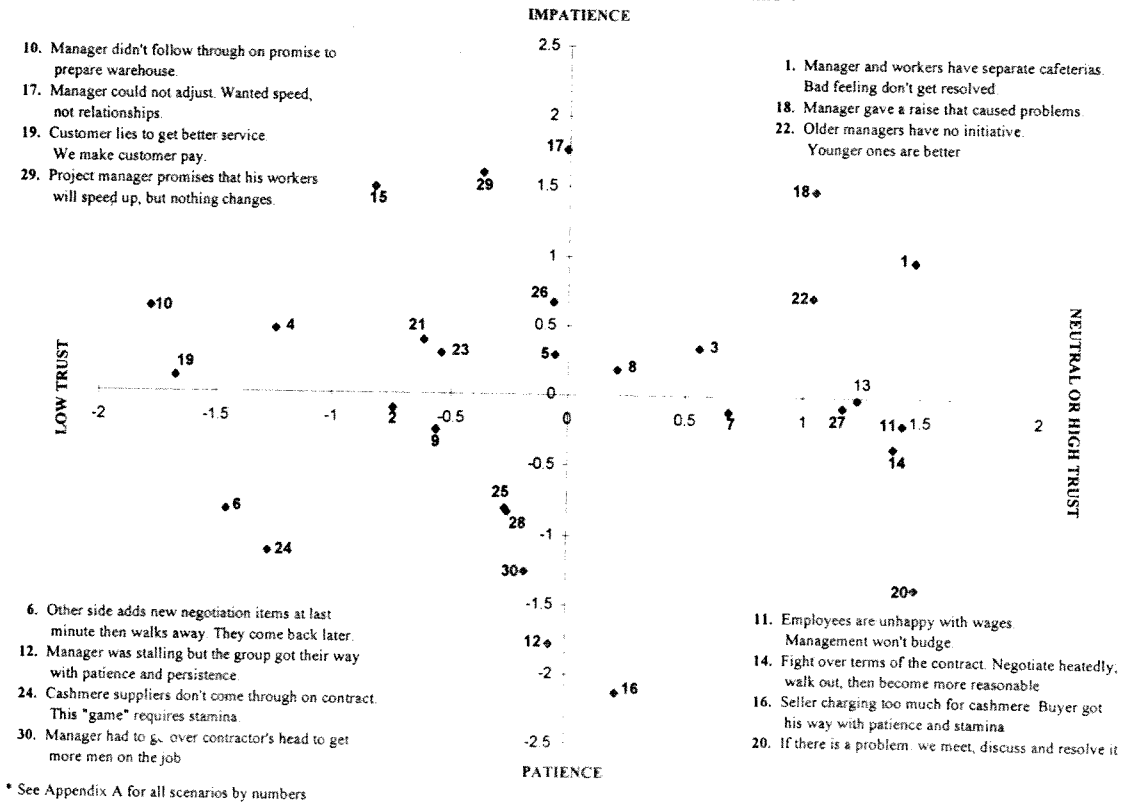


Figure 3: INTRAcultural Conflict : Dimensions 1 and 2 *

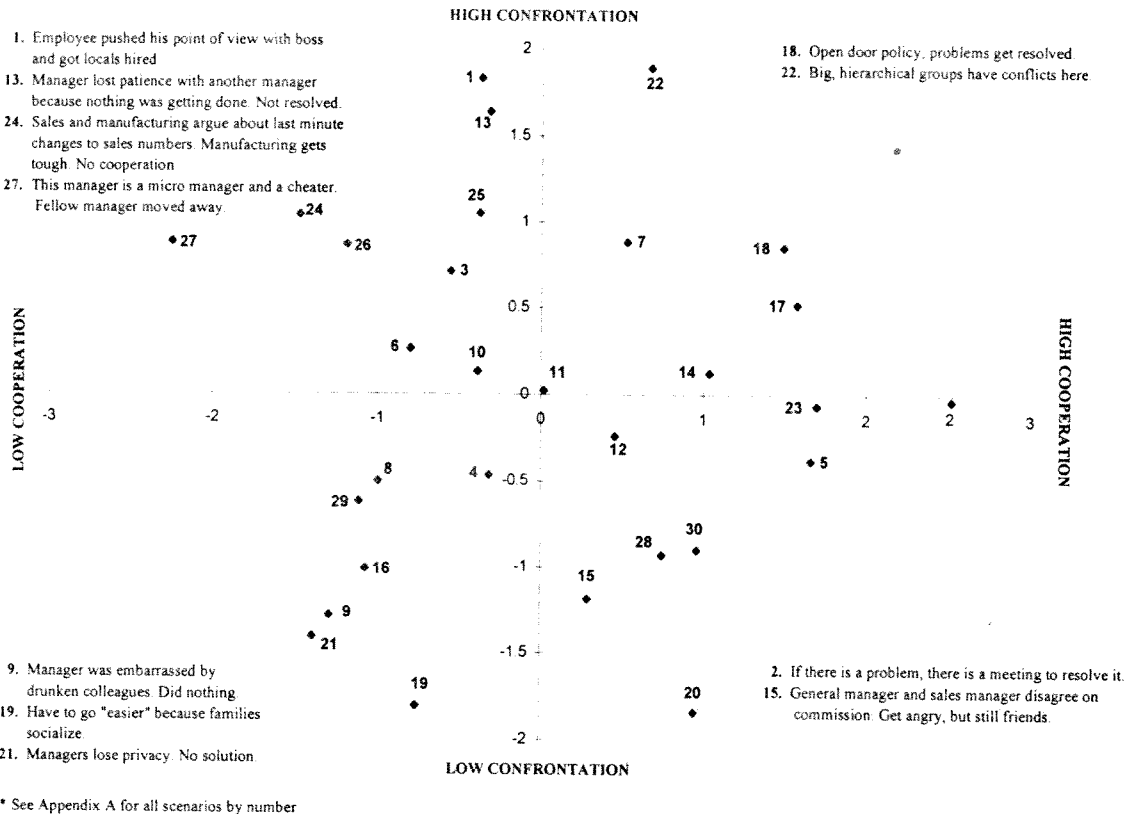
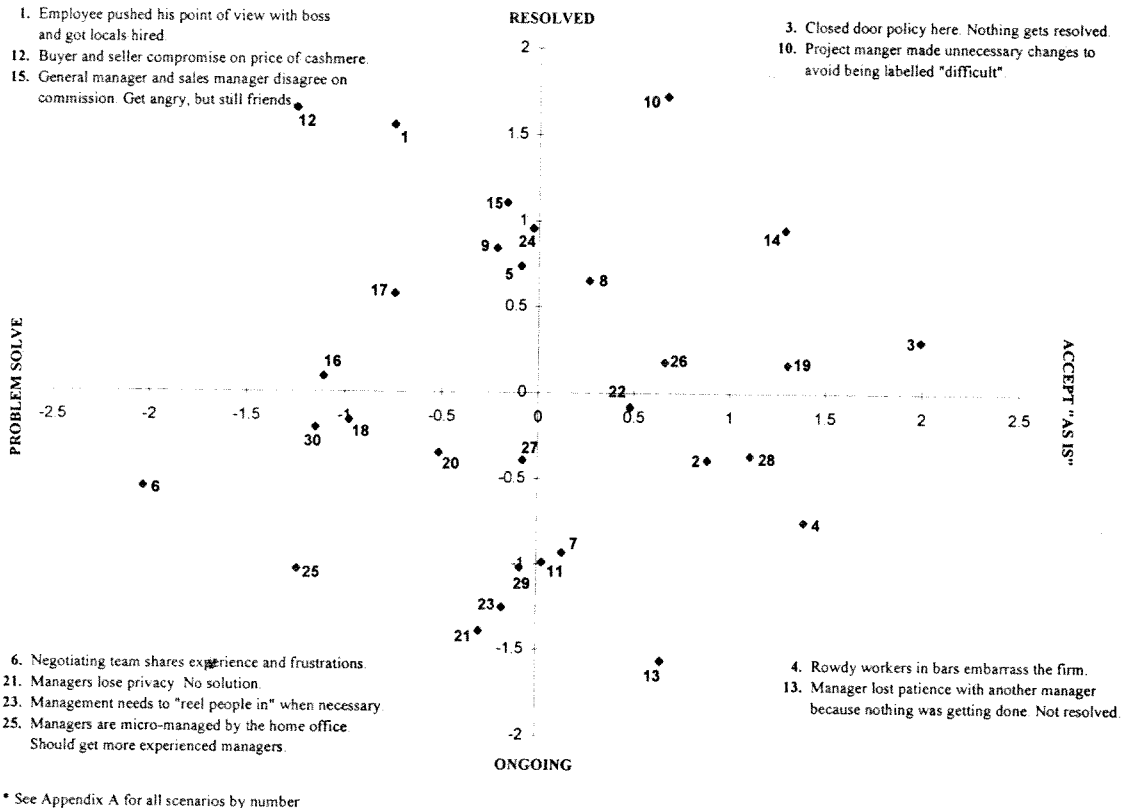


Figure 4: INTRAcultural Conflict: Dimensions 3 and 4*



that best captured the distinction between the two extremes of scenarios falling on the dimension. This procedure was repeated for each of the four intercultural and four intracultural dimensions.

In the intercultural MDS, the four dimensions were: (1) Open Versus Resistant to Change, (2) Situational Causes Versus Individual Traits, (3) High Versus Low Resolution Potential Based on Trust, and (4) High Versus Low Resolution Potential Based on Patience. Figures 1 and 2 illustrate these four dimensions and the scenarios in the space with the axes representing the dimensions.

The intracultural MDS also had four dimensions, but they were different from the ones we saw in the intercultural MDS plots: (1) High Versus Low Cooperation, (2) High Versus Low Confrontation, (3) Problem-Solve Versus Accept as Is, and (4) Resolved Versus Ongoing. Figures 3 and 4 show these four dimensions and where the scenarios fall on the graph. Since the intra- and intercultural analyses are based on different data sets we cannot directly, quantitatively compare them; however, we provide an interpretive comparison below.

Compare Categories from Factor Analysis and MDS

In MDS and factor analysis, the unit of analysis is the conflict scenario. The factors from the factor analysis and the dimensions from the multidimensional scaling analysis can be compared to examine consistency across the various methods of category development and to examine differences in Americans' descriptions of conflict with other American employees as compared to conflict with Chinese employees.

In this example, with data from Jehn, Doucet, and Weldon (1995), initial interpretation revealed that some of the categories were fairly consistent across the methods, and some were not. This was determined by examining the factors and dimensions for the intracultural conflict scenario file and the factors and dimensions for the intercultural scenario conflict file.

For the intercultural data, the first factor was

represented by intense feelings of hate, hostility and insult, while the third factor was more behaviorally oriented and focused on volatile outbursts of rowdy, abusive behavior. The second factor represented a feeling of aggravation, a lower level of feeling indicating that something is a nuisance, someone has perhaps been unprofessional, which causes grumbling and arguing but not the intense feelings of hate and hostility or the rowdy, abusive, volatile behavior of the other factors. In general, the factor analysis shows distinctions of feeling and behavior and high and low intensity.

MDS results illustrate significantly different bases for comparison. Intercultural dimensions focus on individual characteristics of the participants to the conflict (open vs. resistant to change, patient vs. impatient), relational qualities (cooperation), and situational factors (resolved vs. ongoing). These dimensions seem to provide a more general or gestalt view of the conflict and its participants. "Resistance to change" can be manifested and described in behavioral terms as well as affective or state ones. Underlying similarities on dimensions like "cooperation" may only be accessible via gestalt, scenario-level comparisons.

The MDS dimensions seem to rely more heavily on attributions of intent and meaning as opposed to simple observation of behaviors and displayed emotions. Note, for example, that the categories produced by factor analysis pertain largely to negative emotions and behaviors, whereas these negative emotional components seem remarkably absent from the MDS dimensions.

This may reflect different methods of category development. MDS (using pairwise comparison of scenarios) seems to capture more general facets of the conflict, whereas the factor analysis (which aggregates fine-grained individual items or descriptors) seems to capture fine-grained, specific emotional and behavioral components. In the MDS the coders were unconstrained and could distinguish along whatever dimensions they saw fit. On the other hand, the coders in the factor analysis were con-

strained to categories developed by them as described in Part 1 of this article (Jehn and Doucet 1996).

An alternative hypothesis is that since the raters for the MDS and factor analyses are different, this might be a rater difference. In addition, the MDS raters examined 30 of the clearest scenarios for each condition while the factor analysis raters examined all 76 scenarios for each condition.

The statistical and methodological differences between MDS and factor analysis, however, should be noted. Factor analysis tends to address how things covary with one another, or how survey items "hang together." The results of MDS demonstrate the perceived similarities and differences across scenarios by identifying different distinguishing dimensions.

Conclusion

We believe that conducting the above analyses on ethnographic data will produce many interesting categories of whatever construct is being studied. Categories will range from specific behavioral and emotional descriptions of the actors to more general categories, stages, or components of the social interaction of interest. In addition, using the various methods to determine these categories assures that the categories you choose for future analysis of your data are robust.

Using these analyses to compare inter- and intracultural conflicts produced interesting differences across conflict types. We found aspects of conflict that are consistent using the different coding and analysis strategies. We also found the methods are not completely redundant; that is, they vary on the level of interpretation (e.g., concrete behaviors versus general conflict atmosphere) and thus provide a more comprehensive picture of the phenomenon under study.

References

- Jehn, E. and L. Doucet 1996. Developing Categories from Interview Data: Text Analysis and Multidimensional Scaling. *CAM* 8(2):15-16.

Jehn, K., L. Doucet, and E. Weldon 1995. Conflict in Sino-American Joint Ventures: The Differences between Intracultural and Intercultural Conflicts. Paper presented at the Academy of Management Meetings, Vancouver, British Columbia.

Kruskal, J. B. 1964. Multidimensional Scaling by Optimizing Goodness of Fit to a Non-metric Hypothesis. *Psychometrika* 29:1-27.

Young, F. W. and R. Lewyckyj 1979. ALSCAL-4: *User's Guide*. Chapel Hill, NC: Data Analysis and Theory Associates.

Sampling Guidelines for Cultural Data

W. Penn Handwerker

Jeanne Hatcherson

Julie Herbert

University of Connecticut

handwerk@uconnvm.uconn.edu

Introduction

The usual statistical tests contain the assumptions that data observations occur independently and that prediction errors associated with any one case remain uncorrelated with errors made for other cases. Violation of these assumptions invalidates statistical tests, so statistically trained researchers routinely use diagnostic techniques to evaluate whether their data conform to the assumptions of observation and error independence, as well as techniques for avoiding the analytical confusion such violations create.

Consider this question: "How many children did you bear before you turned 20?" This is the sort of question typically asked in survey research. The answers reflect each respondent's unique life trajectory and therefore yield data consistent with the assumptions of observation and error independence. Now consider the following questions: "What do these children mean to you?" "What did you go through while you were pregnant?" "How did people treat you after your baby was born?" These are the kinds of questions typically asked in an in-depth interview—the kind of interviews that generate what are

often called ethnographic, or cultural data.

Complications of Cultural Data

When we take "culture" to refer to the meanings that people use to understand and interact with themselves and the world around them, then cultural data blatantly and inescapably violate the assumption of observation and error independence. The answer any one person gives to questions like "How many children did you bear before you turned 20?" has no necessary relationship to the answer any other person gives to the same question. In contrast, no one can respond to a question like "What do these children mean to you?" or "What did you go through while you were pregnant?" without talking about the social interaction necessary to formulate the answers.

Cultural data thus necessarily reflect the social (interactive) processes by which we construct our knowledge of each other and of how social processes work. Cultural data points exist only insofar as they reflect the *lack of independence* between sources.

By virtue of their construction only through social interaction, cultural data also make the usual sampling criterion (unbiased, "random" case selection) not only impos-

sible to attain, but also undesirable. Indeed, we need knowledgeable informants, whom we select very carefully, at least for cultural domains with clear rules or "culturally correct" answers (Johnson 1990).

Few cultural domains possess clear rules, however. Indeed, even to claim the existence of a set of "rules" or "culturally correct" answers reifies and so distorts the interaction dynamics that generate the meanings we wish to study. We have no clear rules for informant selection when investigating cultural domains like these. The rule we suspect most field workers follow in practice is this: talk to anyone you can about anything and everything.

Skeptics properly question the generalizability of cultural data collected in what appears to be a haphazard manner. The generalizability of noncultural data requires unbiased/random case selection. How can it be that cultural data do not?

The answer, of course, lies in the social processes through which people create cultural data. Cultural differences will reflect differences in the social and historical processes in which people have participated over their lives. Cultural data thus require us to look for cultural differences ("boundaries") that reflect these differences in experience. Absence of these difference constitutes broad agreement about the structure of meaning in a particular domain and, by implication, participation in common social processes experienced in equivalent ways.

We report here an explicit test of the null hypothesis that we discern the same structure of meaning whether or not we select informants randomly. To get right to the point, this null hypothesis is confirmed. In our test, at least, findings show that it makes no difference how we select informants.

Research Procedures

Our data consist of responses to a 16-item triads questionnaire. Triad tests comprise sets of three items that are chosen to represent a cultural domain. Informants look at each set and choose "the one that least fits the other two." So, for example, in a triad test involving the domain of