Cultural Domain Analysis*

STEPHEN P. BORGATTI

Sociology Department
University of South Carolina
Columbia, SC 29208 USA

ABSTRACT: In this paper, I outline a framework for cultural domain analysis. The basic goal of such an analysis is the scientific study of cultural domains from an emic perspective. The framework I present is organized around a classification of the kinds of data encountered in cultural domain analysis. In particular, I distinguish between *monadic* data, in which we measure attributes of items, and *dyadic* data, in which we measure attributes of pairs of items. I then discuss key strategies for collecting and analyzing these types of data.

KEY WORDS: domain, CDA, attribute, cognitive, data collection, consensus, monadic, dyadic

In recent years, there has been increased interest in the exploration of cognitive and/or cultural domains.[1] This is perhaps due to two factors. First, there is the rapid growth of applied anthropology, which needs scientific methods for quickly describing circumscribed aspects of cultures, such as hygienic practices or perceptions of (disease-transmitting) insects. Second, there is the availability of appropriate software, such as ANTHROPAC (Borgatti 1985, 1992), together with rapid adoption of microcomputers by social scientists.

The basic ideas of cultural domain analysis (CDA) are not new: they are derived from cognitive anthropology-ethnoscience and marketing research. Many of the analytic methods, such as factor analysis and unidimensional scaling, are quite old, dating back to the beginning of the century. Other methods, such as QAP, consensus analysis, correspondence analysis and network analysis, are relatively new. Most of the methods, like QAP and correspondence analysis, were developed outside of anthropology. Some, however, like consensus analysis (Romney *et al.* 1986), were developed by anthropologists.

In this paper, I describe an abstract framework for conceptualizing CDA, showing how the different data collection schemes and analytical techniques fit together. The framework is data-oriented, drawing heavily from the theories of data proposed by Coombs *et al.* (1954) and Jacoby (1991).

Journal of Quantitative Anthropology **4**, 261–278, 1994.

1. OVERVIEW

A great deal of empirical cultural research can be summarized as follows: Researchers ask informants to evaluate items on attributes. Collecting cultural data, then, involves measuring relations among four sets of entities: researchers, informants, items and attributes. In the language of data analysis, this forms a 4-way, 4-mode matrix.[2] A proper data matrix recording the empirical results, called X, has cells x_{lkij} whose value gives the kth respondent's evaluation of the ith item on the jth attribute according to the lth researcher. For reasons explained at the end of this section, the set of data corresponding to any one entity (such as a respondent or an item) is often called a *profile*.

Of course, anthropology has rarely concerned itself with replicating results, as other sciences have, so each anthropological dataset is usually only measured by a single researcher. This, in effect, reduces the canonical cultural data matrix to a 3-way, 3-mode matrix involving informants, items and attributes. For example, a medical anthropologist might ask respondents to indicate (on a yes/no scale) which medical treatments are appropriate for which illnesses. Similarly, a market researcher might ask consumers to rate brands of air fresheners on a set of attributes like quality, value, convenience, etc.

In many cases, there is only one attribute being measured (such as good/bad, true/false), which then reduces the data matrix to just informants and items. For example, a health research might ask respondents to rate or rank a set of flying insects on how dangerous they are with respect to transmitting diseases (Kendall *et al.*). Here, the single attribute is "dangerousness". This produces a respondent-by-item matrix in which the items are insects. Or a political scientist might ask respondents a series of agree/disagree questions about abortion, the draft, gun control, capital punishment, prayer in the schools, and the like. Here the attribute being measured is the "correctness" of the items. Similarly, a psychologist might ask subjects about their attitudes towards going to parties, reading books, or meeting people. This yields an informant-by-item matrix in which the items are activities, and the sole attribute being measured is the "desirability" of each activity.

In most studies, certain dimensions (actually, modes) of the data are the focus of attention while others are valued only for the light they shed on the primary interests. Often, the modes of interest are thought of as *variables*, while the remaining modes are thought of as *cases* (replications). For example, the market researcher described earlier might be primarily interested in how each brand is perceived on each attribute. This would make the brand-attribute combinations the variables, and the respondents are seen as observations or replications. One typical next step would then be to average across the respondents to produce an item-by-attribute matrix X whose cells x_{ij} give the average rating of the ith item on the jth attribute.

In contrast, the psychologist described earlier generates a subject-by-activity matrix, in which she might not be very interested in how popular each activity is, but is very interested in how many activities each person enjoys (particularly of a given type). The persons then become the variables, and the activities become the cases. A plausible next step might be to average the data across selected subsets of the activities to get a measure of how "extroverted" each person is. Similarly, a teacher might give and score an exam, yielding a student-by-question matrix in which x_{ij} is 1 if student i got question j right, and 0 otherwise. The teacher averages across questions (columns) to get the proportion of questions that each student got right. Thus, the items are used as indicators of some underlying property (i.e., knowledge) of the students, who are the real focus of the analysis.

Of course, in many situations, different dimensions of a data matrix will be regarded as the variables at different times, depending on the analytic task at hand. For example, in the case of the teacher's data, once having calculated each student's mastery of the material, she might turn around and average the data across students to get the proportion of students getting each question right. This provides a measure of the difficulty level of each question. Now the questions are variables, and the students are cases.

When what are normally considered cases (such as people) are treated as variables, it is common to regard the set of scores for each case across all remaining dimensions of the data as a "profile". The term evokes an image of a distinctive pattern or distribution of rising and falling values that characterize a person. In recent years, the term "profile" has been used to describe the values associated with *any* element of a matrix way, be it person, item, or even attribute.

As I have already indicated, one basic analytic step is to average[3] the data across the cases. Often, this is an end in itself. If a pollster wants to know which candidate is going to win the election, she asks eligible voters to rate each candidate on a simple scale (such as "yes, I will vote for him" or "no, I will not vote for him"), and then aggregates across respondents. The average rating of each candidate is the result to be reported. In other cases, this aggregation is just one step along the way, or it is not used at all.

A second basic analytic step is to examine patterns of covariation or similarity and difference among the variables. If the variables are the columns of a 2-way data matrix, we compare each column with every other, and evaluate the extent to which, when one variable has a relatively large value for a given case, the other variable does as well (and vice-versa). For example, if the data are demographic and socio-economic attributes (columns) of individuals (rows), then we could evaluate whether individuals with high (or low) education levels tend to have correspondingly high (or low) incomes. The result is a 2-way 1-mode variable-by-variable matrix P in which P_{ij} gives the correlation or association between variable i and variable j across all respondents. This derived matrix then becomes the basis of the further analysis.

Measures of association between variables can be viewed more generally as summaries of the similarity of corresponding values of two vectors. Hundreds of measures of similarity, covariation, correlation, co-occurrence, equivalence, and association have been proposed in the literature. Some are applicable to the same kinds of data, some are not. Of those applicable to the same kinds of data, each captures a different nuance of the intuitive concepts of similarity, covariation, and the like. I shall refer to all such measures by the generic label "proximities". Note that distances and differences are the flip side of the same coin, and I include them in "proximities" as well.

Sometimes, proximity data are collected directly, rather than computed from an existing data matrix. Thus, we may ask respondents how similar certain pairs of items appear to be. The result is a 3-way, 2-mode respondent-by-item-by-item matrix. If aggregated across respondents, this yields a 2-way, 1-mode item-by-item proximity matrix, exactly like the proximity matrices described above. A characteristic of proximity matrices is that they have fewer modes than ways, unlike the profile data matrices that I described earlier.

This points up a fundamental difference between profile and proximity data: profile data describe entities, while proximity data describe *pairs* of entities. Profile data are, in this sense, *monadic*, while proximities are *dyadic*. However, profile data can be analyzed dyadically, simply by computing proximities among one or more dimensions of the profile matrix. These distinctions are discussed more fully in sections 3 and 4.

2. DEFINING A CULTURAL DOMAIN

A cultural domain analysis typically begins with the selection of a set of items to work with. Sometimes, the set of items is implicit in the research design. For example, one might have a set of fish hooks whose history and perceived character one wants to explore (Borofsky 1987). Or one might have a set of phrases, culled from focus groups, that relate to sexual behavior (Stanton *et al.* 1993). In most cases, the items are elicited directly from informants by giving them a general description of a domain, and asking them to name items that belong in it. This is called *free listing*. The technique consists of asking a small set of informants (say 30) to name all the items matching a given description. For example, one might ask informants to list all the *bad words* they can think of. The free listing approach works best with categories that have names, especially one-word

names, such as the set of animals, fruits or gems. Examples of freelisting are provided by Trotter (1981) and Gatewood (1983).

The lists generated by each informant are then typed into a computer file, and analyzed by a program like the *FreeList* procedure of ANTHROPAC. The main output of such a program will be a sorted frequency distribution of elicited words or phrases. Typically, the result is a small core set of items mentioned by many respondents, followed by a very long list of idiosyncratic items mentioned by just one or two respondents. Ideally, there is a noticeable break in frequency of mention between the core terms and the idiosyncratic ones. In practice, it is usually difficult to find such an "elbow", and other (arbitrary) ways are found to cut the list in two.

No matter how the list is generated, however, it is important to realize that not every list of items is a cultural domain. A cultural domain is a set of items which are, according to informants, of a kind. Whether a given set of items forms a cultural domain is an empirical question. Once a list of candidates has been formed, one should go back to the informants (perhaps taking a new sample) with the list of items, and ask them which items belong in the domain and which ones do not. This can be done using a variety of methods, such as a paper and pencil true/false test or pilesorts. Such data can then be analyzed via consensus analysis, as described in section 6.

3. MONADIC ATTRIBUTES

As described earlier, a monadic attribute is any attribute of an item in a cultural domain. One of the most basic assumptions of CDA is that items have attributes. Some attributes are categorical (e.g., the color of a banana) while others are quantitative (e.g., the size of an animal). The quantitative attributes are seen as continuums along which each item is positioned. If the continuum is pictured as a horizontal line stretching left and right, some items are more left than others while others are more right, and some items are near each other while others are far from each other. If there are many attributes, then each item has a (probably different) position along each continuum.

Although different attributes are of interest for different cultural domains, it is noticeable that two fundamental dimensions seem to underlie perceptions of many domains. One is the good/bad dimension, and the other is some variant of intense/not intense, active/passive, or strong/weak (Osgood *et al.* 1957).

In many studies, discovering the attributes that people use to distinguish items in the domain is the principal objective. There are two basic approaches to solving this problem. One method is to interview informants on the topic, first informally, and then using a free list technique to elicit all attributes that could be relevant to structuring the domain. The second method is to use some of the techniques described in the next section to obtain judged similarities among the items, and then infer the attributes from the observed pattern of similarities among items.

Once a set of key attributes is identified, the next step is to measure them. To do this, we go back to the informants and ask them to rate or rank the items on each attribute.

3.1. *Monadic data*

The most common method of measuring quantitative attributes, such as size or ferocity of animals, is the n-point rating scale, where n is typically 5 but can range from 2 to 100. The larger the value of n, the more discrimination among items is possible, and the more freedom you give the respondent to express subtle differences of feeling or perception. Also, as n increases, so does reliability. However, for n larger than 20, reliability begins to suffer (Guilford 1954). The smaller the value of n, the less cognitive demands are placed on the respondent. When n is 2, the respondent is merely being asked a yes/no question, as in "Is this mosquito dangerous?".

In rating scales with large n, some respondents restrict themselves to answers at the top (or bottom) end of the scale; this is a level bias. In addition, while some respondents use a limited range of possible values tightly clustered around their mean, others will give wildly scattered responses; this is a scatter bias. Rating scales with small n are less subject to these effects, but even 2-point scales can suffer from "yea-saying".

Most n-point ratings are administered as paper and pencil test. However, if n is small (less than 12), the ratings can be administered as a pilesort task, in which n spots are identified on a table and the respondent places objects representing the items on the appropriate spots. For example, the respondent may be given a set of photographs of plants and asked to put them in four piles, arranged linearly, according to how good to eat they are. Similarly, respondents may be given sentences on cards to be placed in one of two piles depending on whether the sentence is grammatically correct or not.

One potential problem with ratings tasks is order-bias. That is, asking certain questions ahead of others can change the responses to the later items. To get around this problem, it is possible to randomize the order of items (and attributes, if needed) independently for each respondent. Then, after the responses have been recorded, they are unscrambled so that each respondent's data are comparable to each other's. When the data are then averaged across respondents, any response biases introduced by the ordering of items on a page are averaged away as well.

If the n-point ratings are administered as paper and pencil questionnaires, programs like ANTHROPAC can be used to generate a series of individually randomized questionnaires in a standard format. Then, when the completed questionnaires are entered into the computer, the program unscrambles the order to ensure comparability across respondents.

Another approach is to use continuous rating scales in which the respondent draws a length of line, turns a dial, or squeezes a dynamometer in proportion to the degree to which a given item possesses an attribute. Often, this approach is used in conjunction with an anchor point provided by the researcher (e.g., "Suppose burglary is a crime this serious: {researcher draws line}. Now compared to burglary, how serious are each of the following crimes? If you think a crime is twice as serious as burglary, draw a line twice as long . . ."). This approach is known as direct magnitude scaling (Lodge 1981).

Yet another approach is the ranking method, in which respondents are asked to rank order a set of items. Technically, this amounts to an n-point rating scale where n is both the number of possible scores and the number of items, and where no two items are allowed to receive the same score. If the ranking is performed as a paper and pencil task (in which the respondent writes a rank number next to each item), there is a sharp limit to the number of items that can be reliably ranked: it quickly becomes a cognitively difficult task. If implemented as a pilesort, as described above, relatively more items can be ranked. In addition, with very large sets of items (50 or more), it is possible to use a sequential rank-sorting task in which the respondent is initially asked to place all items (represented by cards or objects) into two piles such that all those items in the left pile have more of the attribute than all those in the right pile. Then the respondent is asked to repeat the process with each of the two piles. This creates four piles, ordered from left to right by the attribute in question. The process continuous as long as desired, or until there are no more piles to split.

With smaller sets of items, a method of data collection that is cognitively easier than ranking – and potentially richer – is the paired comparison. In this method, the respondent is presented with all possible pairs of items, one at a time, for a judgement regarding which one has more of the attribute.[4] One way to score the results, for each respondent, is simply to count the number of "contests" that each item has "won". For example, in measuring the perceived sweetness of fruits, mangos are likely to win more pairwise contests than, say, bananas. A more elaborate scoring approach, which can handle missing data, is described by Torgerson (1958), and implemented in ANTHROPAC. One important advantage of this algorithm is that it enables researchers to use incomplete experimental designs in which not every pair of items is presented to the informant. This is useful for large domains, where the number of pairs is impractically large. In many cases it is possible to create a "balanced incomplete block design" in which every item occurs the same number of times.

An interesting aspect of paired comparisons relative to rankings is that a respondent's answers are not forced to be transitive. That is, the respondent may say that item A has more of the attribute than item B, and that B has more than C, yet claim that C has more than A. The occurrence of such intransitivities is a sign that either a mistake has been made, or the attribute being measured is in fact multi-dimensional, and different dimensions are evoked by different pairs. Hence, A could beat B on the attributes evoked by that pair, B could beat C on the (potentially different) attributes evoked by that pair, and C could beat A if the right set of attributes is the basis for comparison.

Examples of monadic data collection in the anthropological literature are provided by Roberts *et al.* (1981) and Weller and Dungy (1986).

3.2. *Monadic analysis*

If a single attribute is measured, the resulting dataset is a respondent-by-item matrix X in which any cell x_{ij} is a measure of respondent i's perception of item j's score on the attribute. This matrix can be averaged across respondents to obtain an average rating for each item on the attribute, or averaged across items to obtain an overall score for each respondent. Which way things are averaged depends on which entities (respondents or items) are the focus of the study, and what the items are. For example, if we ask respondents to rate dog breeds on the size attribute, averaging across respondents makes sense: the result is an estimate of the size of each type of dog. In contrast, averaging across items (dogs) gets nothing substantive: if respondents are unbiased, all respondents should get the same score.[5]

Alternatively, similarities (e.g., correlations) can be computed among the rows (respondents) or columns (items) of the data matrix, creating either a respondent-by-respondent proximity matrix or an item-by-item proximity matrix. The respondent-by-respondent proximity matrix is used to reveal groups of respondents with fundamentally similar views of how the items score on the attribute measured. For example, if the items are a non-homogeneous set of college priorities which the faculty have ranked by importance, it may be that the faculty fall into two camps which rate items in systematically different ways. For example, it may be that those faculty who are either in science or applied fields are most concerned with issues of the physical plant, while faculty who are either in humanities or "pure" fields are more concerned with intangibles. Similarly, it may be that some faculty are relatively concerned with teaching issues, while others are principally concerned with research issues. Those faculty with similar interests will be highly correlated with each other, and poorly or oppositely correlated with others.

Detecting the existence of segments of respondents with systematically different views is of critical importance even when the purpose of the data collection was just to get the average rating of the items on the attribute. To see this, consider the exaggerated case where the items are the college priorities discussed above. Assume that the priorities are of two basic types: teaching-related and research-related. If half the faculty rank all the teaching issues high and all the research issues low, and the other half does the opposite, then averaging the data across all respondents gets you nothing: it is like averaging the preferred temperature of tea for all Americans (including iced tea and hot tea drinkers), and concluding that Americans prefer lukewarm tea. Correlating the respondents first avoids this problem by uncovering "subcultures" with systematically different views whose data may be sensibly averaged within groups, but not across groups.

The item-by-item proximity matrix is used to uncover groups of items which are perceived similarly on the measured attribute. When no clustering of respondents is observed, correlations among the items should be nothing but random fluctuations around zero.[6] But if the items fall into types which are reacted to differently by different segments of respondents, the proximities will show clusters corresponding to the types.

With any proximity matrix, if the number of rows/columns is large (more than 20), it becomes very difficult to see patterns of clustering. The remedy is to submit the matrix to multidimensional scaling (MDS) which depicts the items (or respondents) as points on a map in such a way that the distances between the points correspond to the degree of proximity between the items (or respondents) they represent (Kruskal and Wish 1978). This allows the analyst to locate groups at a glance. An example in the anthropological literature is provided by Boster and Johnson (1989).

One problem with MDS has to do with the dimensionality of the representation. The most natural and useful way to depict the items is on a 2-dimensional plane: e.g., a sheet of paper or a computer screen. However, the data may be such that they cannot accurately be represented in two dimensions. That is, no matter how the points are arranged, the distances among all the points cannot be perfectly proportional to the input proximities. This is captured in a goodness of fit measure called *stress*. When stress is high (say, more than 0.2), it means that the data are not being accurately represented by the MDS map, and one must either use additional dimensions or choose another technique altogether.

An alternative to MDS is cluster analysis (Johnson 1967; Hartigan 1972). The goal of most cluster analytic schemes is to partition items into a few mutually exclusive groups such that items within groups are more proximate to each other than to items in other groups. This provides an automated way of detecting homogeneous groups of items. An example of cluster analysis is provided by Boster, Berlin and O'Neill (1986).

If more than one attribute is measured for each item, the resulting dataset is a respondent-by-item-by-attribute matrix. Typically, the next analytical step is to average across respondents to produce an item-by-attribute matrix, such as brands-by-qualities or tools-by-uses. If the original 3-way matrix consisted of ones and zeros ("Yes, this brand has this feature" or "No, it lacks this feature"), then summing across respondents yields an item-by-attribute frequency matrix in which x_{ij} gives the number of times item i was judged to have attribute j. This matrix can then be analyzed by computing proximities among the rows (to obtain an item-by-item matrix) and, separately, the columns (to obtain an attribute-by-attribute matrix), and submitting both to MDS and cluster analyses. Or, the original item-by-attribute matrix can be submitted directly to correspondence analysis, which is described in section 5.

4. DYADIC ATTRIBUTES

As described earlier, a dyadic attribute is an attribute not of an item but of a pair of items. For example, if the items are cities, an attribute of the pair is the distance between them. Of course, any attribute of an item can be converted into an attribute of pairs. For example, "size" is an attribute of cities, but from it we can easily construct the dyadic attribute, "difference in size". However, I generally describe an attribute as dyadic only if it cannot be reduced to a single monadic attribute.

There are several key dyadic attributes commonly used to understand the structure of a cultural domain. These are: "is similar to", "goes with (co-occurrence)", "is a kind of", and "is a part of" (Werner and Schoepfle 1987). While the latter two relations are analyzable using the methods of network analysis (Borgatti *et al.* 1992), in practice only the first two tend to be studied systematically.

4.1. *Dyadic data*

The most direct way to collect dyadic data is to ask the informant to rate each pair of items on an n-point scale. In fact, any of the rating and ranking tasks used to measure monadic attributes can be used to measure dyadic attributes. The pairs to be rated can be presented one by one, or arranged as a grid or matrix. Both kinds can be created by ANTHROPAC, with random orderings.

However, it must be remembered that because pairs of items are being rated, the number of judgements that must be elicited from each respondent is very high. For example, if there are 20 items, the respondent must make a minimum of 190 judgements.

Perhaps the best approach is the pilesort. The pilesort task uses written cards or photographs to represent the items (or the items themselves, if small enough). The informant is asked to sort the items into piles according to how similar they are. Typically, the number of piles used is left up to the respondent, but imposing a fixed number of piles works well in practice. Pilesorts generate a respondent-by-item-by-item matrix X in which x_{kij} = 1 if respondent k placed item i in the same pile as j, and x_{kij} = 0 otherwise.

The pilesort can be seen as an undemanding way to collect data similar to that generated by a 2-point rating scale, in which the respondent is asked to judge, for each pair of items, whether they are similar or not. The pilesort works well with non-numerate and even non-literate populations. The pilesort differs from the 2-point rating, however, in that it imposes transitivity such that if i and j are similar (i.e. are in the same pile), and j and k are similar, then i and k must be similar (because they are physically in the same pile).[7]

If averaged across respondents, pilesort data yields an item-by-item matrix in which the ijth cell gives the proportion of respondents who placed items i and j in the same pile (regardless of how many piles they made and what other items may have been in the same pile). This is interpreted as a measure of the degree of similarity between items. This matrix is then analyzed via MDS or cluster analysis.

An alternative approach is the triads test. Here, the respondent is presented with triplets of items, one triplet at a time, and the respondent must choose the one item most different. By making that choice, the respondent signals that the remaining two items are relatively similar. If every possible triple of m items is presented, each pair of items will occur together $m - 2$ times, each time "against" a different third item. If the respondents choose the third item as the most different every time, this indicates that the pair is exceptionally similar. In contrast, if every time the pair is presented, one of the two is chosen as most different, this indicates that pair are as dissimilar as possible (since they are never left together). Thus the triads test generates a respondent-by-item-by-item matrix X in which x_{kij} gives the degree of similarity between items i and j perceived by respondent k. Averaged across respondents, the result is an item-by-item similarity matrix suitable for analysis via MDS or cluster analysis.

In general, presenting all possible triads is impractical for domains of more than 9 items.[8] Instead, most studies present only a subset of all the possibilities. To ensure that every pair of items occurs the same number of times, researchers use balanced incomplete block (BIB) designs (Burton and Nerlove, 1976). ANTHROPAC is equipped with a number of these designs. The number of times each pair of items occurs in a BIB design is known as λ. A complete design, where all possible triads occur, has $\lambda = m - 2$. The smallest possible design has $\lambda = 1$, where each pair of items occurs exactly once. These designs, while economical, are potentially unreliable, since the similarity of any two items is entirely determined by their relation to just one other, arbitrarily chosen, item.[9]

One way to use $\lambda = 1$ designs safely is to give each respondent an entirely different design, so that in each questionnaire, a given pair of items occurs against a different third item. Aggregating across respondents then yields a good composite assessment of the similarities among items. Of course, for this to work, there must be consensus among respondents, in the special sense discussed in section 6. One drawback of this approach is that individual responses cannot be compared with each other, since each person receives a different questionnaire.

4.2. *Dyadic analysis*

In some cases, collecting similarities among items is an end in itself. On the hunch that people react similarly to similar stimuli, it is often useful to understand what other items a given item is perceived as similar to. For example, a market researcher needs to know which brands a new product is similar to in order to understand which brands the product will be competing most strongly against. Similarly, an applied anthropologist attempting to promote a certain kind of treatment option might collect similarities among the set of alternatives to see which ones the experimental treatment might substitute for.

In other cases, the similarities are used as the basis for uncovering the underlying attributes that people use to distinguish the items. The assumption is that in order to judge whether two items are similar, respondents calculate a kind of correlation coefficient between the profiles of the items across a set of attributes. Based on her own knowledge of the items, combined with the respondents' similarities, the researcher makes guesses about which attributes seem to be important. For example, if it appears that large dogs tend to be seen as similar to each other, the researcher hypothesizes that size was a factor in making the similarity judgements.

These hypotheses can then be tested by collecting ratings of the items on the hypothesized attributes, and regressing each of them on the coordinates of the items in multidimensional space (as determined by MDS). This technique is known as property fitting or PROFIT (see Kruskal and Wish 1976 for details). In this method each attribute in turn is treated as the dependent variable, and the map positions of the items are treated as independent variables. If the regression is able to predict attribute values accurately (i.e. the r-squared statistic is high), this implies that similar items have similar attribute values. This in turn supports the hypothesis that the attribute was a factor in the way respondents judged similarities among the items.

There are two important problems with the PROFIT technique. One is that it uses MDS coordinates as input rather than the raw data. As noted earlier, MDS pictures represent the raw proximity matrices as accurately as possible, but in real data, there is inevitably distortion. Thus, the PROFIT technique uses distorted data to evaluate the hypotheses, which could affect the results in unpredictable ways.[10]

The other problem with PROFIT is that there are certain ways in which respondents can use items attributes to generate similarities which the regression cannot detect. This occurs when respondents judge all pairs of items in relation to how similar each item is to some prototypical reference item. For example, suppose we have collected similarities in meaning among words expressing a positive evaluation. It may be that respondents consider the category as centered around the fundamental term "good". Words that are closely synonymous with "good" will be seen as similar to each other as well. However, words that are seen as only incidentally related to "good", will not be seen as particularly similar to each other since each departs from the meaning of "good" in its own way. Thus, the attribute generating the judged similarities is "close to good". But the MDS map of these data will place "good" in the middle of the picture, with words similar to "good" surrounding it in concentric circles, and words least like "good" in the outer rings. However, the PROFIT regression demands that the values of the dependent attribute increase monotonically with straight-line movement in any direction on the map, such as left to right or bottom left to top right, and so on. The method is unable to capture a pattern of concentric circles in which the values of the attribute increase as one moves towards the center.[11]

An alternative to PROFIT analysis is the QAP method of Hubert and Schultz 1976. QAP is a general technique for comparing two proximity matrices. For example, if one collected similarities on the same set of items from informants drawn from two different villages, one might be interested in seeing how well the two views correspond to each other. The technique consists of first correlating the matrices cell-by-cell. Then, this observed correlation is compared to a distribution of correlations generated by correlating hundreds of random matrices.[12] The proportion of random correlations that are larger than the observed correlation is then interpreted like the p value in a traditional significance test. QAP can be used as an alternative to PROFIT by converting the attribute vector into a proximity matrix (e.g., by calculating the difference in attribute scores between every pair of items).

5. CORRESPONDENCE ANALYSIS

An interesting variant of dyadic data is the kind elicited by sentence-frame techniques (Black and Metzger 1965; Fabrega 1970). For example, for each informant one might collect which kinds of treatments are applicable to which kinds of illnesses. The result is a respondent-by-illness-by-treatment matrix, which can then be aggregated to yield an illness-by-treatment matrix. These data differ from the monadic item-by-attribute data previously discussed in that neither illnesses nor treatments play the role of attributes. Rather, both illnesses and treatments are items, and the data consist of a dyadic relation between two different sets of items: a 2-mode proximity matrix, in a sense.

The appropriate method of analysis here is correspondence analysis (Benzecri 1967; Nishisato 1980; Greenacre 1984). From a user's point of view, correspondence analysis is a method of mapping both sets of items in the same space. Illnesses with similar profiles across treatments will be located near each other, as will treatments with similar profiles across illnesses. Furthermore, illnesses whose treatment profiles feature certain treatments relatively more than others will be placed near those treatments, while treatments whose profiles feature certain illnesses relatively more prominently than others will be placed near those illnesses.

As noted earlier, correspondence analysis is also useful for item-by-attribute data, and even respondent-by-item data. An example of correspondence analysis in the anthropological literature is provided by Schweizer (1991).

6. CONSENSUS ANALYSIS

Consensus analysis (Boster 1986; Romney *et al.* 1986) provides a framework and method of analyzing patterns of agreement among respondents. Romney *et al.* (1986) show that, given appropriate data, it is possible to infer the amount of knowledge each respondent has about a cultural domain from the pattern of agreements among all pairs of respondents. In other words, agreement is a function of knowledge.

The kind of data this inference applies to is multiple choice data, similar to a standardized test, in which the following three conditions hold. First, there is a single culturally correct right answer to each question, which, if known by a respondent, is always given. Another way to put this is that for each question, whatever the cultural truth might be, it is the same for all respondents. Second, respondents answer each question independently of every other question, and independently of each other. Third, all questions are drawn from the same cultural domain: questions about tennis are not mixed with questions about fish.[13]

These conditions imply that the respondent-by-respondent matrix of agreements among respondents across all questions, adjusted for chance getting, can be factor-analyzed to produce just a single factor (essentially, the answer key) which statistically explains the pattern of agreements. In other words, removing the one factor from the raw data matrix would reduce the agreement matrix to random fluctuations around zero. This means that the extent to which two respondents agree is wholly a function of the extent to which each is correlated to the underlying factor. If the conditions hold, this underlying factor corresponds to the answer key. Hence, the two respondents agree to the extent that each is very knowledgeable, or, to put it in less loaded terms, more culturally central.[14]

If a dataset cannot be explained by a single factor, this is strong evidence that the first condition does not hold, and that more than one set of culturally correct answers exists. For example, it may be that the sample contains representatives from more than one subculture, which have systematically different responses to the same set of questions.

The basic ideas of consensus analysis have important implications for the processing of all kinds of data. As noted earlier, if it can be determined that a given sample contains representatives from two distinct subculture, it would be folly to aggregate the data across all respondents to obtain a majority view: averaging attitudinal data from liberals and conservatives on issues such as abortion and gun control would result in middling ratings that were neither for nor against any position. Thus, all data that are aggregated across respondents make the implicit assumption that the respondents are drawn from the same subculture, and that all variations in respondent answers are unsystematic (i.e., due to variance in "cultural centrality") rather than systematic (due to belonging to different subgroups with fundamentally different views).

This assumption is particularly important in the case of pilesort data, where the degree of similarity between items is measured by the proportion of people who put a given pair of items in the same pile. If many people put the items in the same pile, we see the items as similar. If many people put the items in different piles, we see the items as dissimilar. If half the respondents put the items in the same pile, and half put them in different piles, we see the items as having medium similarity. Implicitly, we assume a response model in which people perceive a certain middling degree of similarity between the items, and have to choose between putting them in the same pile or different piles. If the perceived similarity level is close enough to neutral, the respondent more or less flips a coin to decide whether to put the items together or not. Taken across many respondents, this would mean that, in the long run, about half the respondents put the items together, and half keep them separate. This reasoning does not allow the possibility that the reason why half the group put the items together and half kept them separate is that there exists two strong perspectives on the matter, one of which sees the items as clearly similar, and the other of which sees them as clearly different. Thus, the method of scoring pilesort data (as well as $\lambda = 1$ triad tests), implicitly depends on the presence of consensus, which is to say that a single culturally correct answer for each question (i.e. the similarity of each pair of items) exists for all respondents sampled. It is essential, therefore, that before aggregating pilesort data, an effort be made to analyze the pattern of agreements among respondents, to ensure that the pattern is consistent with a single culture, rather than two or more conflicting subgroups.

7. CONCLUSION

This paper lays out the analytic framework that I believe underlies cultural domain analysis. At the heart of the framework is the fundamental distinction between monadic and dyadic data, which are collected and analyzed using substantially similar techniques. The goals of CDA are the traditional goals of cognitive anthropology: the scientific study of culture from an emic perspective. Some of the analytic tools, however, are new. For example, QAP, consensus analysis, correspondence analysis and network analysis are all relatively new developments that are still diffusing through the social science community. Consensus theory is an especially important development, in my view, because it deals with the nature of variability between cases, with implications for when and how to aggregate data, which is the fundamental operation of analysis.

NOTES

* This paper is based on lectures given at the NSF Summer Institute on Research Methods in Cultural Anthropology, 1988–1992. I am grateful to my fellow instructors, H. R. Bernard and Pertti Pelto, and to the participants of the Institute, for many helplful discussions. I would also like to thank Gery Ryan for research assistance.

[1] I use the terms "cognitive" and "cultural" interchangeably in this context, but some may prefer to reserve "cognitive" for the study of individuals and "cultural" for the study of groups.

[2] The number of *ways* in a matrix is the number of subscripts needed to identify a cell. For example, an ordinary table with rows and columns is a 2-way matrix whose cells are identified by x_{ij}. I use "ways" and "dimensions" interchangeably. The number of *modes* in a matrix is the number of distinct sets of objects referenced by the rows, columns, level or other dimensions of the matrix. For example, a matrix of driving distances among cities is a 2-way, 1-mode matrix because both rows and columns refer to the same entities, namely cities.

[3] It should be understood that I use "average" to stand in for any number of aggregations, including summations, counts, weighted averages, principal components,etc.

[4] It may seem odd that paired comparisons, which seem inherently dyadic, are discussed in the section on monadic data. In truth, paired comparisons should probably be seen as dyadic data, particularly in light of the frequently observed intransitivities discussed in the next paragraph. However, I have chosen to discuss the method here because it is almost universally utilized as a means of generating monadic data.

[5] If there is response bias, where some respondents consistently rate all dogs larger than other respondents do, then the average score across items could be considered a measure of that bias. Or one could argue that the average score across items measures each person's fear of dogs, since those afraid should consistently overestimate the size. However, these are special cases.

[6] However, if the measure of proximity is euclidean, then the proximity between any two items will, random error aside, be equal to the difference in their average ratings on the attribute.

[7] An exception occurs if respondents are allowed to place items into more than one pile, which can be achieved by creating multiple cards for the same item.

[8] For 10 items, the number of triples is 120; for 15 items, it is 455.

[9] Based on simulations, Burton and Nerlove come to the conclusion that $\lambda = 1$ designs are risky, but not bad on average. However, my own experiences with such designs using real respondents have never been good.

[10] The effects of this problem can be minimized by using high-dimensional MDS solutions, which have very low stress. In fact, when stress is low, the distortion may constitute a kind of data cleaning or smoothing which lessens the effects of random error and clarifies the underlying pattern. Normally, only 1 to 3 dimensional solutions are used because of the difficulty in visualizing higher dimensional configurations. However, in the context of property fitting, this visualization is not important.

[11] This is why PROFIT cannot be used to confirm respondent competence as the factor underlying similarities among informants with respect to their profiles of responses to multiple choice questions, as would be expected from the consensus theory of Rommey *et al.* (1986).

[12] Actually, rather than create random numbers, the QAP approach entails randomly permuting the rows and columns of the materials, and then correlating these permutated matrices. This preserves any hidden relationships between different matrix cells (such as transitivity) which may be present in the data matrices and which could artificially inflate the correlations.

[13] In the mathematical theory, this condition is expressed technically by requiring that each respondent have a fixed probability of getting each question "right". In other words, each respondent knows the answer to a certain proportion of all possible questions about the domain.

[14] The term "culturally central" was suggested by Douglas Caulkins (personal communication).

REFERENCE

Benzecri, J. P. 1969 Statistical Analysis as a Tool to Make Patterns Emerge from Data. In: W. Watanabe (ed.) Methodologies of Pattern Recognition. Pp. 35–74. New York: Academic Press.

Black, M. and D. Metzger 1965 Ethnographic Description and the Study of Law. American Anthropologist 6: 141–165.

Borgatti, S. P. 1985 ANTHROPAC 3.0. Software Available from the Author.

Borgatti, S. P. 1992 ANTHROPAC 4.0. Columbia: Analytic Technologies.

Borgatti, S. P., M. G. Everett and L. C. Freeman 1992 UCINET IV Version 1.0. Columbia: Analytic Technologies.

Boster, J. S. 1986 Exchange of Varieties and Information between Aguaruna Manioc Cultivators. American Anthropologist 88: 428–436.

Boster, J. S., B. Berlin and J. O'Neill 1986 The Correspondence of Jivaroan to Scientific Ornithology. American Anthropologist 88: 569–583.

Boster, J. S. and J. C. Johnson 1989 Form or Function: A Comparison of Expert and Novice Judgements of Similarity among Fish. American Anthropologist 91: 866–889.

Burton, M. L. and S. B. Nerlove 1976 Balanced Designs for Triads Tests: Two Examples from English. Social Science Research 5: 247–267.

Coombs, C. H., H. Raiffa and R. M. Thrall 1954 Some Views on Mathematical Models and Measurement Theory. Psychological Review 61: 132–144.

Fabrega, H. 1970 On the Specificity of Folk Illnesses. Southwestern Journal of Anthropology 26: 305–314.

Gatewood, J. 1983 Linguistic Competence and Recognition Ability. American Anthropologist 85: 378–386.

Greenacre, M. J. 1984 Theory and Application of Correspondence Analysis. New York: Academic Press.

Guilford, J. P. 1954 Psychometric Methods. New York: McGraw-Hill.
Hartigan, J. A. 1972 Direct Clustering of a Data Matrix. Journal of the American Statistical Association 67: 123–129.
Hubert, L. and J. Schultz 1976 Quadratic Assignment as a General Data Analysis Strategy. British Journal of Mathematical and Statistical Psychology 29: 190–241.
Jacoby, W. G. 1991 Data Theory and Dimensional Analysis. Newbury Park: Sage Publications.
Johnson, S. C. 1967 Hierarchical Clustering Schemes. Psychometrika 32: 241–254.
Kruskal, J. B. and M. Wish 1976 Multidimensional Scaling. Beverly Hills: Sage Publications.
Lodge, M. 1981 Magnitude Scaling. Bevely Hills: Sage Publications.
Nishisato, S. 1980 Analysis of Categorical Data: Dual Scaling and its Applications. Toronto: University of Toronto Press.
Osgood, C. E. 1957 The Measurement of Meaning. Urbana: University of Illinois Press.
Romney, A. K., S. C. Weller and W. H. Batchelder 1986 Culture as Consensus: A Theory of Culture and Informant Accuracy. American Anthropologist 88: 146–170.
Schweizer, T. 1991 The Power Struggle in a Chinese Community, 1950-1980: A Social Network Analysis of the Duality of Actors and Events. Journal of Quantitative Anthropology 3: 19–44.
Stanton, B., R. Aronson, S. Borgatti, S. Feigelman, I. Ricardo, L. Kaljee, J. Galbraith, V. Keane and T. Whitehead 1993 Interdisciplinary Methodologic Challenges: Confirmation of Interpretation from Focus Group Discussions. AIDS Education and Prevention (in press)
Roberts, J. M., T. V. Golder and G. E. Chick 1981 Judgement, Oversight, and Skill: A Cultural Analysis of P-3 Pilot Error. Human Organizzation 39: 5–21.
Torgerson, W. S. 1958 Theory and Methods of Scaling. New York: John Wiley.
Trotter, R 1981 Remedios Caseros: Mexican-American Home Remedies and Community Health Problems. Social Science and Medicine 15b: 107–114.
Weller, S. C. and C. I. Dungy 1986 Personal Preferences and Ethnic Variations among Anglo and Hispanic Breast and Bottle Feeders. Social Science and Medicine 23: 539–548.
Werner, O. and G. M. Schoepfle Systematic Fieldwork: Volume 2. Newbury Park, CA: SAGE Publications.