

Ethkt4 (version in 5.1 without some figures and tables. Mar. 4, 1998 version)

Chapter 1
Elicitation Techniques
for Cultural Domain Analysis¹

by **Stephen P. Borgatti**, University of Kentucky
Daniel S. Halgin, University of Kentucky

[HEAD: LIST, BULLETS

INTRODUCTION?
ELICITING CULTURAL DOMAINS USING FREE LISTS
ELICITING DOMAIN STRUCTURE USING PILESORTS
ELICITING DOMAIN STRUCTURE USING TRIADS]

We are grateful to H. Russell Bernard, Pertti Pelto, A. Kimball Romney, and Gery Ryan for helping shape our views on cultural domain analysis, which is not to say they necessarily agree with anything we have written. We are also grateful to Mark Fleisher and to John Gatewood for giving us permission to use their data to illustrate concepts. Finally, we thank Jay Schensul and Marki LeCompte for their many helpful comments on earlier drafts.

INTRODUCTION

The techniques described in this chapter are used to understand *cultural domains*

[DEFINITION: MARGIN, **A cultural domain is a set of items or things that are all of the same type or category**] (Lounsbury, 1964; Spradley 1979; Weller & Romney, 1988). A cultural domain is a mental category like “animals” or “illnesses”. It is a set of items that are all alike in some important way. Humans in all cultures classify the world around them into cognitive domains, and the way they do this affects the way they interact with the world. Not all cultures classify things the same way. For example, English-speakers recognize a category called “shrubs” which is different from “trees” and “grasses”. But many other cultures do not recognize the “shrub” category at all: they divide up the plant kingdom differently. Even when cultures have the same domains, the contents may be somewhat different. For example, many cultures have a domain called “illnesses”, but often these cultures include as illnesses things that most Americans would regard as imaginary, such as “evil eye”, or things that Americans regard as symptoms, such as “stomach pains”. Ethnographers often begin their studies by trying to identify and describe the cultural domains that are used by the people they are studying.

The techniques described in this chapter are used to (a) elicit the items in a cultural domain, (b) elicit the attributes and relations that structure the domain, and (c) measure the positions of the items in the domain structure. These techniques, which include freelists, pilesorts, triads, multidimensional scaling, and graph layout algorithms have been incorporated into two commercially available computer programs called Anthropac (Borgatti, 1992) and UCINET (Borgatti, Everett & Freeman, 2002).

Defining Cultural Domains

There are several ways to define a cultural domain. A good starting point is: a set of items all of which a group of people define as belonging to the same type. For example, “animals” is a cultural domain. The members of the domain of animals are all the animals that have been named, such as dogs, cats, horses, lions, tigers, etc. But there is more to the idea than just a set of items of the same type. Implicit in the notion is also the idea that membership in the cultural domain is determined by more than the individual respondent -- the domain exists “out there” either in the language, in the culture or in nature. Hence, the set of colors that a given individual likes to wear is not what we mean by a cultural domain.

One rule of thumb for distinguishing cultural domains from other lists is that cultural domains are about people’s perceptions rather than people’s preferences. Hence, “my favorite foods” is not a cultural domain, but “things that are edible” is. Another way to put it is that cultural domains are about things “out there” in reality, so that, in principle, questions about the members of a domain have a right answer. Consider, for example, the cultural domain of animals. If asked whether a tiger is an animal, the respondent feels that she is discussing a fact about the world outside, not about herself. In contrast, if she is asked whether “vanilla” is one of her favorite ice cream flavors, the respondent feels that she is revealing more about herself than about vanilla ice cream. In this sense, cultural domains are experienced as outside the individual and shared across individuals.

The fact that cultural domains are shared across individuals does not mean that all members of a given population are in complete agreement on which items belong to a given cultural domain. The extent to which a cultural domain is actually shared in any given population is an empirical question --- that is, a question that is open to testing.² Conversely, simple agreement about a set of items does not imply that the set is a cultural domain. If we ask 1,000 randomly sampled informants in our own culture about their ten favorite foods and every one of them happens to give the same list, it is still not a cultural domain because personal preferences are not the kind of thing which in principle could be a cultural domain. In contrast, responses to the question, "What foods are preferred in your community?" could be a cultural domain.

Another aspect of cultural domains is that they have internal structure.

[DEFINITION:MARGIN: THE INTERNAL STRUCTURE OF A CULTURAL DOMAIN REFERS TO THE RELATIONSHIPS THAT EXIST AMONG THE ITEMS OR THINGS IN IT] That is, they are systems of items related by a web of

relationships. For example, in the domain of animals, some animals are understood to eat other animals. The relation here is "eats", and every pair of animals can be evaluated to

Some people use the term "cognitive domain" to refer to domains which are not necessarily shared. For example, a psychologist might make an in-depth study of one person's understanding of nature. Since no other respondents were studied, the psychologist might refer to the person's categories as cognitive domains rather than cultural domains. However, it is important to realize that whether they are shared or not, cognitive domains have all the same properties as cultural domains, including being experienced as outside the individual, as outlined above. In this sense, we can think of cognitive domains as the general category, and cultural domains as a member of that category.

5 Elicitation Techniques

see if the first animal eats the second. Another relation applicable to animals, recognized by biologists at least, is “competes with”.

A relation of particular importance, which seems to be common to all cultural domains, is the relation of similarity. It appears that, for all cultural domains, respondents can readily indicate which pairs of items they consider similar, and which they consider dissimilar. Another relation that seems to apply to most domains is co-occurrence, as in which foods “go with” which others, or which animals live in the same habitats with which others.

Relations among things are a fundamental aspect of how humans think about the world. Lists of “universal” relations have been made by many researchers, including Casagrande and Hale (1967) and Spradley (1979). Spradley’s list includes [LIST:

BULLETS

- cause-and-effect (X causes Y, Y is the result of X),
- inclusion (X is a kind of Y),
- rationale (X is a reason for doing Y),
- means-end (X is a way to accomplish Y),
- sequence (X follows Y)
- function (X is used for Y)
- spatial (X is a part of Y), (X is a place in Y)
- attribution (X is a characteristic of Y)
- location for action (X is a place for doing Y)]

Most of these, however, are not relations among items in the same domain, but rather relate the items from one domain to the items in another domain. For example, <italic> location for action <italic> relates a place, such as "Madrid" with an activity, such as "bullfighting"; places and activities typically belong to different cognitive domains. Similarly, the cause of a given effect is not necessarily a member of the same cognitive domain as the effect. For instance, making love may result in getting AIDS, but most respondents think of these as belonging to different domains. In this chapter we concentrate on only those relations that relate items within a single cognitive domain.

Other largely universal relations are semantic relations among the terms used to label items in a cultural domain. These are relations such as synonymy (same meaning) and antonymy (opposite meaning). For example, in the domain of illnesses, there is often more than one term for a given illness (such as a folk name and a medical term). While the line separating relations among terms from relations among the items themselves may be difficult to draw, in principle our interest here is in the relations among the items rather than among the terms we use to describe them.

An important class of relations among items is the kind that can be reduced to a single attribute. For example, in the domain of illnesses, some illnesses are seen as "more contagious" than other illnesses. This relation is based on a single property of each illness in the domain, which is how contagious it is. This is different from the relation of (perceived) similarity, which is indivisible. We cannot attach a similarity score to an individual item -- it is always attached to a pair. For example, we can say that the similarity between "pneumonia" and "flu" is 8 on a scale of 1 to 10, but it doesn't make

7 Elicitation Techniques

sense to assign a similarity score to just one of the illnesses by itself (as in, ‘the flu has a similarity score of three’). In contrast, it does make sense to assign an individual illness a contagious score: we don’t have to do it in pairs. The difference between attributes of individual items and relations among pairs of items becomes more clear as we go along in this chapter.

In general, an attribute that makes sense for some items in a cultural domain will make sense for all items. In other words, if “sweetness” is a sensible attribute of fruit, then it is meaningful to ask ‘how sweet is ____?’ of all fruit in the domain. If the attribute cannot be applied to all items, this is sometimes because not all the items are at the same level of contrast, which in turn means that there are subdomains. For example, if the domain of “animals” contains the items “squirrel”, “ant”, and “mammal”, informants will be confused if asked whether squirrels are faster than mammals. The real test for items of different levels of contrast, however, is to look at the semantic “is a kind of” relation (Casagrande and Hale, 1967; Spradley, 1979). If any item in a domain is a kind of any other item in the domain (e.g., squirrel is a kind of mammal) then you know that the latter item is actually a cover term (a *gloss*) for a subdomain. [DEFINITION: MARGIN, **Cover terms are summary terms encompassing all the items in a domain or subdomain**]

Even if all the items are of the same level of contrast, however, the inability to apply an attribute to all items is sufficient to suggest that the domain has a hierarchical taxonomic structure and that the attribute belongs to items in one particular class. For example, the attribute “shape of wings” can be applied to some animals, but not to others. This means that the domain of animals contains at least two types — animals with wings

and animals without — and within the set of those with wings, we can ask what shape the wings are (see Figure 1).

INSERT FIGURE 1 HERE

Eliciting Cultural Domains Using Freelists

The *freelist* technique is used to elicit the elements or members of a cultural domain. For domains that have a name or are easily described, the technique is very simple: just ask a set of informants to list all the members of the domain. For example, you might ask them to list all the names of illnesses that they can recall. If you don't know the name of a domain, you may have to elicit that first. For example, you can ask “what is a mango?” and very likely you will get a response like “it's a kind of fruit”. Then you can ask, “what other kinds of fruit are there?” Note that if a set of items does not have a name in a given culture, it is likely that it is not (yet) a domain in that culture. However, you can still obtain a list of related items by asking questions like “what else is there that is like a mango?”.

At first glance, the freelist technique may appear to be the same as any open-ended question, such as “What illnesses have you had?” The difference is that freelisting is used to elicit cultural domains, and open-ended questions are used to elicit information about individual informants (see Table 1). In principle, the freelists from different respondents who belong to the same culture should be comparable and similar because the stimulus question is about something outside themselves and which they have in

9 Elicitation Techniques

common with other members. In contrast, an open-ended question could easily generate only unique answers.

INSERT TABLE 1 HERE

Collecting Freelist Data

Ordinarily, freelists are obtained as part of a semi-structured interview, not an informal conversation. With literate informants, it is easiest to ask the respondents to write down all the items they can think of, one item per line, on a piece of paper. The exact same question is asked of the entire sample of respondents (see below for a discussion of sample size). We then count the number of times each item is mentioned and sort the items in order of decreasing frequency. For example, I asked 14 undergraduates at Boston College to list all the animals they could think of. On average, each person listed 21.6 animal terms. The top twenty terms are given in Table 2.

INSERT TABLE 2 HERE

The number of informants needed to establish a cultural domain depends on the amount of cultural consensus in the population of interest — if every informant gives the exact same answers, you only need one — but a conventional rule-of-thumb is to obtain lists from a minimum of 30 lists. One heuristic for determining whether it is necessary to interview more informants, recommended by Gery Ryan³ (personal communication), is to compute the frequency count after obtaining 20 or so lists from randomly chosen informants, then repeat the count after 30 lists. If the relative frequencies of the top items have not changed, this suggests that no more informants are needed. In contrast, if the relative frequencies have changed, this indicates that the structure has not yet stabilized and you need more informants. This procedure only works if the respondents are being sampled at random from the population of interest. If, for example, the domain is illnesses and the first 20 respondents are all nurses, the method might indicate that no more respondents are needed. Yet if the results are intended to represent more than just nurses, more (non-nurse) respondents will be needed.

³Dr. Ryan is a Senior Behavioral Scientist at RAND and an Adjunct Assistant Professor in the Department of Psychiatry and Biobehavioral Sciences at UCLA

11 Elicitation Techniques

The frequency of items is usually interpreted in terms of their salience to informants. That is, items that are frequently mentioned are assumed to be highly salient to respondents, so that few forget to mention those items. Another aspect of salience, however, is how soon the respondent recalls the item. Items recalled first are assumed to be more salient than items recalled last. The second column from the right in Table 2 gives the average position or rank of each item on each individual's list. With sufficient respondents (more than used in Table 2), it is often the case that a strong negative correlation exists between the frequency of the items and their average rank, at least for the items mentioned by a majority of respondents. This means that the higher the probability that a respondent mentions an item, the more likely it is that they will mention it early. This supports the notion of salience as a latent property that determines both whether an item is mentioned and when. In recognition of this, some researchers like to combine frequency and average rank into a single measure⁴.

Once the freelists have been collected and tabulated, it usually becomes apparent that there are a few items that are mentioned by many respondents, and there are a huge number of items that are mentioned by just one person. For example, I collected freelist data on the domain of "bad words" from 92 undergraduate students at the University of South Carolina. A total of 309 distinct items were obtained, of which 219 (71%) were mentioned by only one person (see Figure 2). As discussed near the end of this section, domains seem to have a core/periphery sort of structure with no absolute boundaries. The

4

One such measure, Smith's S (Smith, 1997), is given in the rightmost column of Table 2. The measure is essentially a frequency count that is weighted inversely by the rank of the item in each list. In practice, Smith's S tends to be very highly correlated with simple frequency.

more respondents you have, the longer the periphery (the right-hand tail in Figure 2) grows, though ever more slowly.

INSERT FIGURE 2

From a practical point of view, of course, it is usually necessary to determine a boundary for the domain one is studying. [LIST: BOX, BULLETS

Ways to Determine a Domain Boundary

--Include all items mentioned more by more than one respondent.

--Look for a natural break or grouping.

--Define a boundary arbitrarily.]

One natural approach is to count as members of the domain all items that are mentioned by more than one respondent. This is logical because cultural domains are shared at least to some extent, and it is hard to argue that an item mentioned by just one person is shared. However, this approach usually does not cut down the number of items enough for further research. Another approach is to look for a natural break or “elbow” in the sorted list of frequencies.⁵ This is most easily done by plotting the frequencies in what is known as a “scree plot” (see Figure 2). When such a break can be found it is very convenient, and may well reflect a real difference between the culturally shared items of the domain and the idiosyncratic items. But if no break is present, it is ultimately necessary to arbitrarily choose the top N items, where N is the largest number you can really handle in the remaining part of the study. In Figure 2, no really clear breaks are

Or salience, as captured by Smith’s S.

13 Elicitation Techniques

present, but there are three “mini-breaks” that one might consider. In the sorted list of words, they occur after the 20th, 26th, and 40th words.

One problem that must be dealt with before computing frequencies is the occurrence of synonyms, variant spellings, subdomain names, and the use of modifiers. For instance, in the “bad words” domain, some of the terms elicited were “whore”, “ho”, and “hore”. It seems likely that “whore” and “hore” are variant spellings of the same word, and therefore pose no real dilemma. In contrast, “ho”, which was used primarily by African-American students, could conceivably have a somewhat different meaning. (There is always this potential when a word is used more often by one ethnic group than by others.) Similarly, in the domain of animals, the terms “aardvark” and “anteater” are synonymous for most people, but for some (including biologists), “anteater” refers to a general class of animals of which the aardvark is just one. Whether they should be treated as synonyms or not will depend on the purposes of the study. It may be necessary, before continuing, to ask respondents whether “aardvark” means the same thing as “anteater”.

Occasionally, respondents will fall into a response set in which they list a class of items separated by modifiers. For example, they may name “grizzly bear”, “Kodiak bear”, “black bear”, and “brown bear”. Obviously, these constitute subclasses of bear that are at a lower level of contrast than other terms in their lists. Occasionally, these kinds of items may lead respondents to generalize the principle to other items, so that they then list such items as “large dog”, “small dog”, and “hairless dog”. In general this is not a problem because these kinds of items will be mentioned by just one person, and so will be dropped from further consideration.

Analyzing Freelist Data

While the main purpose of the freelisting exercise is to obtain the membership list for a domain, the lists can also be used as ends in themselves. That is, several interesting analyses can be done with such lists.

INSERT TABLE 3 HERE

Once we have a master list of all items mentioned, we can arrange the freelist data as a matrix in which the rows are informants and the columns are items (see Table 3). The cells of the matrix can contain ones (if the respondent in a given row mentioned the item in a given column) or zeros (that respondent did not mention that item). Taking column sums of the matrix would give us the item frequencies. Taking column averages would give us the proportion of respondents mentioning each item. Taking row sums would give us the number of items in each person's freelist.

The number of items in an individual's freelist is interesting in itself. Although perhaps confounded by such variables as respondent intelligence, motivation and personality, it seems plausible that the number of items listed reflects a person's familiarity with the domain (Gatewood, 1984). For example, if we ask people to list all sociological theories of deviance they can think of, we should expect to find that professional sociologists have longer lists than most other people. Similarly, dog fanciers are likely to produce longer lists of dog breeds than ordinary people. Yet length of list is obviously not perfectly correlated with domain familiarity, as respondents who are

relatively unfamiliar with a domain can produce impressively long lists of very unusual items -- items with which other respondents would not agree.

To construct a better measure of domain familiarity (or “cultural domain competence”) we could weight the items in an individual freelist by the proportion of respondents who mention the item. Adding up the weights of the items in a respondent’s freelist then gives a convenient measure of what might be called “cultural competence”. Respondents score high on this measure if they mention many high-frequency items and avoid mentioning low-frequency items.

INSERT TABLE 4

INSERT TABLE 5

Another way to analyze freelist data -- now focusing on the items rather than the respondents -- is to examine the co-occurrences among freelisted items. Table 4 gives an excerpt from a respondent-by-item freelist matrix. There are four items labeled A through D. Consider items A and B. Each are mentioned by four respondents. Three respondents mention both of them. That is, A and B co-occur in three of the six freelists. By comparing every pair of items, we can construct the item-by-item matrix given in Table 5. This matrix can then be displayed via multidimensional scaling (MDS), as shown in Figure 3. In a multidimensional scaling map of this kind, two items are close together to the extent that many respondents mentioned both items. Items that are far apart on the map were rarely mentioned by the same respondents.

INSERT FIGURE 3

Typically, such maps will have a core/periphery structure in which the core members of the domain (i.e., the most frequently mentioned) will be at the center, with the rest of the items spreading away from the core and the most idiosyncratic items located on the far periphery. The effect is similar to a fried egg.⁶ This occurs in part because odd items can be odd in so many ways that they tend to be different from each other as well. In contrast, core items are very similar to each other. Another factor is that core items tend to be mentioned so much more often, there is a greater chance of overlapping with other items.

While not an artifact, exactly, of the column sums of the matrix (i.e., some items are mentioned more often than others), the core/periphery structure of co-occurrence matrices is made visible by not controlling for the sums. It is also useful to examine the pattern obtained by controlling for these sums. One way to do this is to simply compute Pearson correlations among the columns. Another way is to count both matches of the ones and the zeros.

There are a number of other ways to analyze freelist data. As Henley (1969) noticed, the order in which items are listed by individual respondents is not arbitrary. Typically, respondents produce runs of similar items separated by visible pauses. Even if we do not record the timing, we can recover a great deal of information about the cognitive structuring of the domain by examining the relative position of items on the list. Two factors seem to affect position on the list. First, as mentioned earlier, the more central items tend to occur first. When we ask North Americans to list all animals, “cat” and “dog” tend to be at the top of each person’s list, and they tend to be mentioned by nearly everyone.

A second pattern is that related items tend to be mentioned near each other (i.e., the difference in their ranks is small). Hence, we can use the differences in ranks for each pair of items as a rough indicator of the cognitive similarity of the items. To do this, we construct a new person-by-item matrix in which the cells contain ranks rather than ones and zeros. For example, if respondent “Jim” listed item “Deer” as the 7th item on his freelist, then we would enter a “7” in the cell corresponding to his row and the deer column. If a second respondent named “Fred” did not mention an item at all, we enter a special code in Fred’s column denoting a missing value (NOT a zero). Then we compute correlations (or distances) among the columns of the matrix. The result is an item-by-item matrix indicating how similarly items are positioned in different people’s lists, when they occur at all. This can then be displayed using multidimensional scaling. It should be noted, however, that if uncovering similarities is the primary interest of the study, it would be wiser to use more direct methods, such as those outlined in the next section.

It should also be noted that while we reserve the term “freelisting” for the relatively formal elicitation task described here, the basic idea of asking informants for examples of a conceptual category is very useful even in informal interviews (Spradley 1979). For example, in doing an ethnography of an academic department, we might find ourselves asking an informant “You mentioned that there are a number of ways that graduate students can get into difficulty. Can you give me some examples?” Rather than eliciting all the members of the domain, the objective might be simply to elicit just one element, which then becomes a vehicle for further exploration.

It is also possible to reverse the question and ask the respondent if a given item belongs to the domain, and if not, why not. The negative examples help to elicit the taken-for-granted characteristics that are shared by all members of the domain and which therefore might otherwise go unmentioned.

Eliciting Domain Structure Using Pileorts

The pileort task is used primarily to elicit from respondents judgments of similarity among items in a cultural domain. It can also be used to elicit the attributes that people use to distinguish among the items. There are many variants of the pileort sort technique. We begin with the *free pileort*.

Collecting Free Pileort Data

The typical free pileort technique begins with a set of 3-by-5 cards on which the name or short description of a domain item is written. For example, for the cultural

domain of illnesses, we might have a set of 80 cards, one for each illness. For convenience, a unique ID number is written on the back of each card. The stack of cards is shuffled randomly and given to a respondent with the following instructions: “Here are a set of cards representing kinds of illnesses. I’d like you to sort them into piles according to how similar they are. You can make as many or as few piles as you like. Go!”

In some cases, it is better to do it in two steps. First you ask the respondent to look at each card to see if they recognize the illness. Ask them to set aside any cards representing illnesses that they are unfamiliar with. Then, with the remaining cards, have them do the sorting exercise.

Sometimes, respondents object to having to put a given item into just one pile. They feel that the item fits equally well into two different piles. This is perfectly acceptable. In such cases, the researcher can simply take a blank card, write the name of the item on the card, and let the respondent put one card in each pile. As discussed in a later section, putting items into more than one pile causes no problems for analyzing the data, and may correspond better to the respondents’ views. The only problem it creates is that it makes it more difficult later on to check whether the data were entered into computer files correctly, since having an item appear in more than one pile is usually a sign that someone has mistyped an ID code.

Instead of writing names of items on cards, it is sometimes possible to sort pictures of the items (see Figure 4), or even the items themselves (e.g., when working with the folk domain of “bugs”). However, in our experience, for literate respondents, the written method is always best. Showing pictures or using the items themselves tends to

bias the respondents toward sorting according to physical attributes such as size, color and shape. For example, sorting pictures of fish yields sorts based on body shape and types of fins (Boster and Johnson, 1989). In contrast, sorting *names* of fish allows non-visible attributes to affect the sorting (such as taste, where the fish is found, what it is used for, how it is caught, what it eats, how it behaves, etc.).

INSERT FIGURE 4

Normally, the pilesort exercise is repeated with at least 30 respondents⁷, although the number depends on the amount of variability in responses. For example, if everyone in a society would give exactly the same answers, you would only need one respondent. But if there is a great deal of variability, you may need hundreds of sorts to get a good picture of the modal answers (i.e., the most common responses), and so that you can cut the data into demographic subgroups so that you can see how different groups sort things differently.

Analyzing Pilesort Data

Pilesort data are tabulated and interpreted as follows. Every time a respondent places a given pair of items in the same pile together, we count that as a vote for the similarity of those two items (see Table 6). In the domain of animals, if all of the respondents place “coyote” and “wolf” in the same pile, we take that as evidence that these are highly similar items. In contrast, if no respondents put “salamander” and “moose” in the same pile, we understand that to mean that salamanders and moose are

The number 30 is merely a convention -- a rule of thumb. More respondents is always more desirable but involves more time and expense.

not very similar. We further assume that if an intermediate number of respondents put a pair of items in the same pile this means that the items are of intermediate similarity.

INSERT TABLE 6 HERE

This interpretation of agreement as monotonically⁸ related to similarity is not trivial and is not widely understood. It reflects the adoption of a set of simple process models for how respondents go about solving the pilesort task.

Process Model

--They use a similarity metric or measure.

--They "bundle" or clump together items with similar

attribut

This means that there is a 1-to-1 correspondence between the rank orders of the data. That is, the pair placed most often in the same pile is the most similar, the pair placed second-most often in the same pile is the second-most similar, etc.

One such model is the metric model. Each respondent is seen as having the equivalent of a similarity metric in her head (e.g., she has a spatial map of the items in semantic space). However, the pilesort task essentially asks her to state, for each pair of items, whether the items are similar or not. Therefore, she must convert a continuous measure of similarity or distance into a yes/no judgment. If the similarity of the two items is very high, she places, with high probability, both items in the same pile. If the similarity is very low, she places the items, with high probability again, in different piles. If the similarity is intermediate, she essentially flips a coin (i.e., the probability of placing in the same pile is near 0.5). This process is repeated across all the respondents, leading the highly similar items to be placed in the same pile most of the time, and the dissimilar items to be placed in different piles most of the time. The items of intermediate similarity are placed together by approximately half the respondents, and placed in separate piles by the other half, resulting in intermediate similarity scores.

An alternative model, not inconsistent with the first one, is that respondents conceptualize domain items as bundles of features or attributes. When asked to place items in piles, they place the ones that have mostly the same attributes in the same piles, and place items with mostly different attributes in separate piles. Items that share some attributes and not others have intermediate probabilities of being placed together, and this results in intermediate proportions of respondents placing them in the same pile.

Both these models are plausible. However, even if either or both is true, there is still a problem with how to interpret intermediate percentages. Just because intermediate similarity implies intermediate consensus does not mean that the converse is true, namely

that intermediate consensus implies intermediate similarity. For example, suppose half the respondents clearly understand that shark and dolphin are very similar (because they are large ocean predators) and place them in the same pile, while the other half are just as clear that shark and dolphin are quite dissimilar (because one is a fish and the other is a mammal). Under these conditions, 50% of respondents would place shark and dolphin in the same pile, but we would NOT want to interpret this as meaning that 100% of respondents believed shark and dolphin to be moderately similar. In other words, the validity of measuring similarity by aggregating pilesorts depends crucially on the assumption of underlying cultural consensus (Romney, Weller and Batchelder, 1986), a topic we take up in more detail a bit further along.

INSERT FIGURE 5 HERE

We can record the proportion of respondents placing each pair of items in the same pile using an item-by-item matrix, as shown in Table 6. This matrix can then be represented spatially via non-metric multidimensional scaling, or analyzed via cluster analysis.⁹ Figure 5 shows a multidimensional scaling of pilesort similarities among 30 crimes collected by students of Mark Fleisher¹⁰. In general, the purpose of such analyses would be to [LIST, BULLETS

An excellent introduction to multidimensional scaling is provided by Kruskal and Wish (1978). For an introduction to cluster analysis, we recommend Everitt (1980).

The data were collected specifically for inclusion in this chapter by: Jennifer Teeple, Dan Bakham, Shannon Sendzimmer, and Amanda Norbits. We are grateful for their help.

-- reveal underlying perceptual dimensions that people use to distinguish among the items, and

--detect clusters of items that share attributes or comprise subdomains.]

Let us discuss the former goal first. One way to uncover the attributes that structure a cultural domain is to ask respondents to name them as they do the pilesort¹¹. One approach is to ask respondents to “think aloud” as they do the sort. This is useful information but should not be the only attack on this problem. Respondents can typically come up with dozens of attributes that distinguish among items, but it is not easy for them to tell you which ones are important. In addition, many of the attributes will be highly correlated with each other if not directly synonymous, particularly as we look across respondents. It is also possible that respondents do not really know why they placed items into the piles that they did: when a researcher asks them to explain, they cannot directly examine their unconscious thought processes and instead go through a process of justifying and reconstructing what they must have done. For example, all native speakers of a language are good at constructing grammatically well-formed sentences, but they need not have conscious knowledge of grammar to do this.

In addition, it is possible that the research objectives may not require that we know how the respondent completes the sorting task but merely that we can accurately

It is best to use a different sample of respondents for this purpose, or wait until they have finished the sort and then ask them to discuss the reasons behind their choices. Otherwise, the discussion will influence their sorts. You can also have them sort the items twice: the first time without interference, the second time discussing the sort as they go. The results of both sorts can be recorded and analyzed, and compared.

predict the results. In general, scientists build descriptions of reality (theories) that are expected to make accurate predictions, but are not expected to be literally true, if only because these descriptions are not unique and are situated within human languages utilizing only concepts understood by humans living at one small point in time. This is similar to the situation in artificial intelligence where if someone could construct a computer that could converse in English so well that it could not be distinguished from a human, we would be forced to grant that the machine understood English, even if the way it did so could not be shown to be the same as the way humans do it. What is common to both scientific theories and artificial intelligence is that we evaluate truth (success) in terms of the behavioral outcomes, not an absolute yardstick.

To discover the underlying perceptual dimensions people use to distinguish among items in a cultural domain, we begin by collecting together the attributes elicited directly from respondents. Then we look at the multidimensional scaling (MDS) map to see if the items are arrayed in any kind of order that is apparent to us.¹² For example, in the crime data shown in Figure 5, it appears that as we move from right to left on the map, the crimes become increasingly serious. This suggests the possibility that respondents use the attribute “seriousness” to distinguish among crimes. Of course, the idea that the leftmost crimes are more serious than the rightmost crimes is based on the researcher’s perceptions of the crimes, not the informants’. Furthermore, there are other attributes that might arrange the crimes in roughly the same order (such as violence). The

It is important to remember that since the axes of MDS pictures are arbitrary; perceptual dimensions can run along any angle, not just horizontal or vertical.

first question to ask is whether respondents have the same view of the domain as the researchers.

To resolve this issue, we then take all the attributes, both those elicited from respondents¹³ and those proposed by researchers, and administer a questionnaire to a (possibly new) sample of respondents asking them to rate each item on each attribute. This way we get the informants' views of where each item stands on each attribute. Then we use a non-linear multiple regression technique called PROFIT (Kruskal and Wish, 1975) to statistically relate the average ratings provided by respondents to the positions of the items on the map. Besides providing a statistical test of independence (to guard against the human ability to see patterns in everything), the PROFIT technique allows us to plot lines on the MDS map representing the attribute so that we can see in what direction the items increase in value on that attribute. Often, several attributes will line up in more or less the same direction. These are attributes that have different names but are highly correlated. The researcher might then explore whether they are all manifestations of a single underlying dimension that respondents may or may not be aware of.

Sometimes MDS maps do not yield much in the way of interpretable dimensions. One way this can happen is when the MDS map consists of a few dense clusters separated by wide open space. This can be caused by the existence of sets of items that happen to be extremely similar on a number of attributes. Most often, however, it signals the presence of subdomains (which are like categorical attributes that dominate res-

Either as part of the pilesort exercise, or by showing the MDS map to informants and asking them to interpret it.

pondents' thinking). For example, a pilesort of a wide range of animals, including birds, land animals, and water animals will result in tight clumps in which all the representatives of each group are seen as so much more similar to each other than to other animals that no internal differentiation can be seen. An example is given in Figure 6. In such cases, it is necessary to run the MDS on each cluster separately. Then, within clusters, it may be that meaningful dimensions will emerge.

INSERT FIGURE 6

We may also be interested in comparing respondents' views of the structure of a domain. One way to think about the pilesort data for a single respondent is as the answers to a list of yes/no questions corresponding to each pair of items. For example, if there are N items in the domain, there are $N(N-1)/2$ pairs of items, and for each pair, the respondent has either put them in the same pile (call that a YES) or a different pile (call that a NO). Each respondent's view can thus be represented as a string of ones ("yesses") and zeros ("no's"). We can then, in principle, compare two respondents' views by correlating these strings.

However, there are problems caused by the fact that some people create more piles than others. This is known as the "lumper/splitter" problem. For example, suppose two respondents have identical views of what goes with what. But one respondent makes many piles to reflect even the finest distinctions (he's a "splitter"), while the other makes just a few piles, reflecting only the broadest distinctions (she's a "lumper"). Correlating their strings would yield very small correlations, even though in reality they have identical views. Another problem is that the strings of two splitters can be fairly highly

correlated even when they disagree a great deal because both say “no” so often (i.e., most pairs of items are NOT placed in the same pile together). Some analytical ways to ameliorate the problem have been devised, but they are beyond the scope of this chapter.

The best way to avoid the lumpers/splitter problem is to force all respondents to make the same number of piles. One way to do this is to start by asking them to sort all the items into exactly two piles, such that all the items in one pile are more similar to each other than to the items in the other pile. Record the results. Then ask the respondents to make three piles, letting them rearrange the contents of the original piles as necessary.¹⁴ The new results are then recorded. The process may be repeated as many times as desired. The data collected can then be analyzed separately at each level of splitting, or combined as follows. For each pair of items sorted by a given respondent, the researcher counts the number of different sorts in which the items were placed together. Optionally, the different sorts can be weighted by the number of piles, so that being placed together when there were only two piles doesn't count as much as being placed together when there were 10 piles. Either way, the result is a string of values (one for each pair of items) for every respondent, which can then be correlated with each other to determine which respondents had similar views.

A more sophisticated approach was proposed by Boster (1994). In order to preserve the freedom of a free pilesort while at the same time controlling the lumpers/splitter problem, he begins with a free pilesort. If the respondent makes N piles,

An alternative here is to ask them to divide each pile in two. This is repeated as often as desired.

29 Elicitation Techniques

the researcher then asks the respondent to split one of the piles, making $N+1$ in total. He repeats this process as long as desired. He then returns to the original sort and asks the respondent to combine two piles so that there are $N-1$ in total. This process is repeated until there are only two piles left.

Both of these methods, which we can describe as *successive pilesorts*, yield very rich data, but they are time-consuming and can potentially require a lot of time to record the data. The the respondent also has a long wait while data are recorded. In Boster's method, because piles are not rearranged at each step, it is possible to record the data in an extremely compact format without making the respondent wait at all. However, it requires extremely well-trained and alert interviewers to do it.

Eliciting Cultural Domain Structure Using Triads

An alternative to pilesorts for measuring similarity is the triad test. [LIST: BOX, BULLETS

Triads are used for:

- very small domains (12 items or less)
- testing hypotheses where it is important that every respondent make an active judgment regarding the similarities among a certain set of items, or
- getting people to define what attributes they use to distinguish among items.]

In a triads test, the items in a domain are presented to the respondent in groups of three. For each triple, the respondent must pick out the one she judges to be the most different. For example, one triple drawn from the domain of animals might be:

Dog

Seal

Shark

Picking any item is equivalent to voting for the similarity of the other two. Hence, choosing “dog” would indicate that “seal” and “shark” were similar, while choosing “shark” would indicate that “seal” and “dog” were similar. If all possible triples are presented, each pair of items will occur $N-2$ times¹⁵, each time “against” a different item. If a pair of items is really similar it will “win” each of those contests and will be voted most similar a total of $N-2$ times. If the pair is extremely dissimilar, it will never win. For example, “oyster” and “elephant” might occur in the following triples:

Again, N is the number of items in the domain.

Oyster	Elephant	Dog
Oyster	Elephant	Shrimp
Oyster	Elephant	Ostrich

In the first one, the respondent might choose “oyster” as the most different. In the second, the respondent might choose “elephant”. In the third, the respondent might choose “oyster” again, and so on. Hence, the triad test in which every possible triple is presented will yield a similarity score for each pair of items that ranges from zero to $N-2$, where N is the number of items. For example, if there 10 items, then each pair will occur against all $10-2 = 8$ remaining items.

The problem with presenting all possible triples is that there are $N(N-1)(N-2)/6$ of them, which is a quantity that grows with the cube of the number of items. For example, if the domain has 30 items in it, the number of triples is 30 times 29 times 28 divided by 6, which is 4,060, which is too many for an informant to respond to, even over a period of days. The solution is to take a manageable sample of triples. For example, out of the 4,060 triples, we might randomly select 200 for the respondent to work with. However, a random sample would allow some pairs of items to appear in several triples, and allow others not to occur at all. The latter would be a real problem because the purpose of the task is to measure the perceived similarity between every pair of items.

The solution is to use a *balanced incomplete block* or BIB design (Burton and Nerlove, 1976). In a BIB design, every pair of items occurs a fixed number of times. The number of times the pair occurs is known as lambda (λ). In a complete design (where all

possible triples occur), λ obviously equals $N-2$, since each pair occurs against every other item in the domain. When λ equals 1, we have the smallest possible BIB design, where each pair of items occurs only once. For a domain with 30 items, a $\lambda=1$ design would have only 435 triads — still a lot, but a considerable savings over 4,060.

In general, however, $\lambda=1$ designs should be avoided, because the similarity of each pair of items will be completely determined by their relation to whichever item happens to turn up as the third item. For example, if “elephant” and “mouse” occur in this triple:

Mouse	Elephant	Rat
--------------	-----------------	------------

it is likely that they will be measured as not similar, since “elephant” is likely to be chosen as most different. But if instead they happen to occur in this triple:

Mouse	Elephant	Oyster
--------------	-----------------	---------------

it is likely that they will be measured as similar. Thus, it is much better to have at least a $\lambda=2$ design, where each pair of items occurs against two different third items. The only exception to this rule of thumb is when you give each respondent in a culturally homogeneous sample a completely different triad test, based on the same set of items but containing different triples. For example, respondent #1 might get “mouse” and “elephant” paired with “oyster”, but respondent #2 might get “mouse” and “elephant” paired with “dog”. In a way, this is like taking a complete design and spreading it out

across multiple respondents. This can work well, but means that you cannot compare respondents' answers with each other to assess similarity of views, since each person was given a different questionnaire.

A nice feature of the triads task is that, unlike the simple pilesort, it yields degrees of similarity for pairs of items for each respondent. In the simple pilesort, each respondent essentially gives a “yes they are similar” or “no they are not” vote. In the triads, the range of values obtained for each pair of items goes from zero to λ . Hence, for a $\lambda=3$ design, each pair of items is assigned an ordinal similarity score of 0, 1, 2, or 3. This means that we can sensibly construct separate multidimensional scaling maps for each respondent.¹⁶

One problem with triad tasks is that respondents often find them tiring and repetitive. They will swear that a certain triad has already occurred, and will suspect that you are trying to see if they are responding consistently, which makes them nervous. Another problem is that respondents tend to become aware of their own thought processes as they proceed, and start feeling uncomfortable about using varying criteria (which is unavoidable) to pick the item most different in each triple. This makes them feel that they are not doing a good job. In general, triads are only useful for very small domains (12 items or less) or for testing hypotheses (where it is important that every respondent make an active judgment regarding the similarities among a certain set of items).

The same was true for the successive pilesort techniques described earlier.

Analyzing Triad Data

Perhaps the most interesting use of triads was by Romney and D'Andrade (1964) who used them to test two theories of cognition about American male kinship roles, such as grandfather, father, son, grandson, uncle, brother, nephew and cousin. One theory, by Wallace and Atkins (1960), held that Americans use two attributes --- generation and lineality [STEVE DEFINE HERE -- **COLLINEAL, ABLINEAL, EGO, RECIPROCAL, DIRECT, AND COLLATERAL**] --- to distinguish among the roles, as shown in Table 7. In the table, "lineal" refers to kin who are either ancestors or descendants of the speaker, who by convention is labeled "ego". The term "collineal" refers to non-lineal kin whose set of ancestors include or are included by ego's set of ancestors. The term "ablineal" refers to all other blood relatives.

INSERT TABLE 7

If the theory is true, in a triads test that included the triple

Grandfather	grandson	father
--------------------	-----------------	---------------

Americans should choose "grandson" as the one most different because grandfather and grandson are the least different with respect to the two attributes in the model (all of the terms are lineal, differing only on generation, where "grandfather" and "father" are adjacent, but "grandson" is a step removed).

In contrast, Romney and D'Andrade propose a model with three attributes --- generational distance, lineality, and reciprocal roles --- as shown in Table 8. In the table, "direct" refers to kin that share the same ancestors as ego, and "collateral" are all others.

INSERT TABLE 8

According to the Romney and D'Andrade model, when faced with the same triad given above, Americans should choose, with equal probability, either "grandson" or "father" as the item most different, and should never choose "grandfather". Given these predictions, it was a simple matter to test the theories by giving the triads to a sample of Americans and seeing which theory best predicted the actual answers on the triads test. Overall, the best theory turned out to be the Romney and D'Andrade model.

Informal Use of Triads

So far, I have only described the formal use of the triads task, which results in the generation of similarities among items. Another way to use triads is as a device to spark discussion of the underlying attributes that people use to distinguish among items in the domain. To do this, we present informants with a small random sample of triples, one at a time. For each triple, the informant is asked to explain in what ways each item is different from the other two. This is an extraordinarily effective way to elicit the attributes that people use to think about the domain. For example, consider this triple:

Cancer**Syphilis****Measles**

This triple can elicit a number of perceived attributes of illnesses including seriousness ("cancer is fatal"), age of the afflicted ("measles is something that kids get"), morality ("you get syphilis from sleeping around too much"), contagiousness ("you can catch

sypilis and measles from other people”), and so on. It is easy to see that it only requires a handful of triples to elicit dozens and dozens of attributes.

Consensus Analysis

In the study of cultural domains, the researcher must be aware of issues of cultural variability among respondents. It is impossible to interpret the results of triads and pilesorts if fundamentally different systems of classification are in use among different respondents. One way to determine whether variability among respondents reflects different systems of classification is the consensus methodology of Romney, Weller and Batchelder (1986). This method, available in both UCINET and Anthropac, provides a way to (a) discover whether a set of questions has multiple correct answer keys (corresponding to different cultures), (b) uncover what the culturally “correct” answers are for each culture, and (c) assess the extent of cultural domain knowledge possessed by a given member of a culture. Thinking back to the example of dolphins and sharks, consensus analysis would indicate whether our sample includes subgroups of respondents with systematically different answers across all similarity judgments (e.g., because some are sorting on the basis of habitat others are sorting based on phylogeny), or just individual variability, where some people simply know more about a given domain than others do.

Once consensus analysis has been used to identify culturally homogeneous groups of respondents, it can then be applied within in each group to determine the amount of knowledge regarding the cultural domain possessed each respondent in that group. In

effect, the theory underlying consensus analysis distinguishes between two kinds of variability in people's responses. One kind is systematic difference that we attribute to cultural differences. The other kind is random or piecemeal difference which we attribute to individual differences in knowledge of the domain. For example, within a culture, people may have similar understandings of types of flowers, but some people simply have more of that cultural knowledge than others. They know more names of flowers, they know more about which are used in what occasions (such as weddings or funerals or romances), and so on. This does not mean they know more in the sense of Truth with a capital "t", but rather that they know more of their own culture.

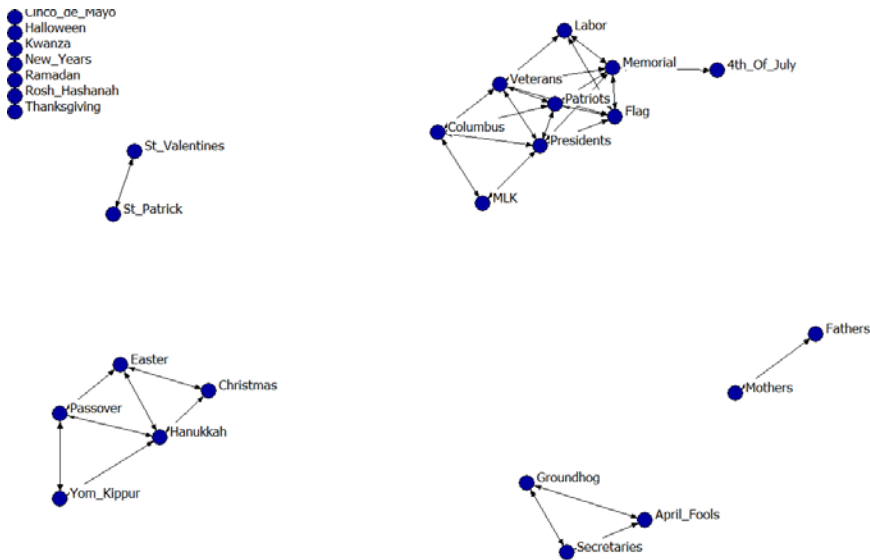
Consider the study of crimes discussed earlier in this chapter in which 30 respondents were asked to sort crimes into piles based on how similar the crimes were to each other. As discussed, we can use the pilesort method to uncover various types of crimes (see figure X for a MDS plot), and then use consensus analysis to determine whether there was agreement among respondents. Consensus analysis methods of pilesort data using UCINET and Anthropac use individual proximity matrices as input. In this example, the input is an item-by-item matrix for each of the 30 respondents in which $x_{ij}=1$ if the respondent placed crime i and crime j in the same pile. The program then correlates the 30 item-by-item matrices with each other and runs a factor analysis. The program output includes factor loadings for all respondents. If one factor is predominant, we can conclude that there is a single culture. In this example, the largest eigenvalue = 14.440 and the 2nd largest eigenvalue = 1.653 indicating strong cultural consensus. If the respondents were not culturally homogeneous we would find multiple factors

indicating systematically different response patterns. We are also able to identify respondent 3 as an individual with the greatest amount of cultural knowledge (see table X for other competence scores) because he or she has the highest loading on the first factor. We also see that respondent 22 has a significantly lower competence score than all the others – it is possible the respondent did not understand the task.

Visualization of Cultural Domains

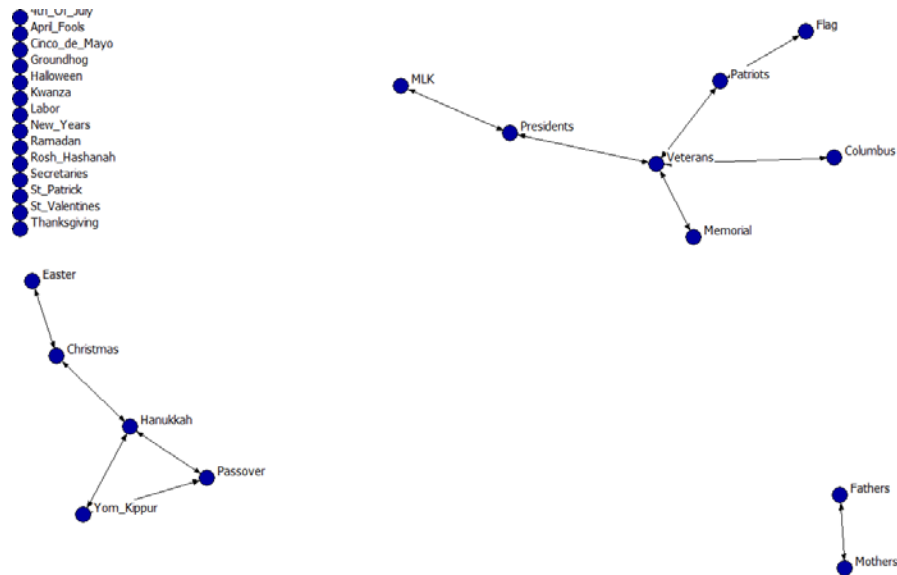
Implicit in these data collection and analysis techniques is the idea of the cultural domain as a network of items related by semantic relations, or families of linked meanings. Thus, we can use visualization and analytic techniques drawn from social network analysis to investigate cultural domains. For example, network visualization tools can be used to elucidate the internal structure of a cultural domain and identify how the position of items within this structure distinguishes the items from each other and gives them their unique meanings.

Network visualization is based on graph layout algorithms (GLAs) which locate nodes in space and connect them with lines indicating a close relationship (cite deJordy et al?) . In the case of cultural domain analysis, the lines are determined by the level of similarity among items determined by the researcher (e.g., how often two items appeared in the same pile, the correlation coefficient of two items, etc.). [something about choosing the level] Boston College student Heidi Stokes used the pilesort method to collect data on the perceived similarity of 24 holidays as part of an undergraduate research methods class. Figure 6 is a GLA representation of the holiday data in which an edge is shown connecting holidays deemed similar by at least 50% of the respondents. The edges allow the viewer to clearly identify various groupings that exist within the domain of holidays among Boston College undergraduates. We can see that there is a large grouping of religious holidays and a large grouping of patriotic holidays. The GLA also reveals structure within these groupings. The existence of a line from the 4th of July to Memorial Day but no other holidays in the grouping of Patriotic holidays indicates that respondents considered the 4th of July more similar to Memorial Day than to other holidays such as Columbus Day and Martin Luther King Day.



Researchers can use GLAs to investigate the perceived similarity of two items at various strengths to better determine the structure of items within groupings and possible attributes that might give items their unique meanings. For example, we can draw a GLA in which a line is shown connecting holidays deemed similar by at least 70% of the respondents (or any other percentage decided by the researcher). Figure X indicates that respondents in the Catholic school deemed Easter more similar to Christmas than to Jewish holidays Hanukkah, Yom Kippur, and Passover. We can also see that Christmas is deemed more similar to Hanukkah than to other Jewish holidays perhaps because both holidays occur at the same time of year.

41 Elicitation Techniques



Conclusion

We have presented basic techniques for eliciting data concerning cultural domains. The freelist technique is primarily used to elicit the basic elements of the domain. The pilesort and triad tasks are used both to elicit similarities among the items, and to elicit attributes that describe the items. Consensus analysis can be used to uncover the culturally correct items in a cultural domain in the face of certain kinds of intra-cultural variability, and enables the researcher to assess the extent of knowledge possessed by an informant about a given cultural domain. In addition, we have touched on the use of multidimensional scaling and graph layout algorithms to graphically illustrate the structure of the domain, and locate each item's position in that structure.

SUGGESTED ADDITIONAL RESOURCES

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

Anthropac is a menu-driven computer program for cultural domain analysis. The program's capabilities include all the techniques discussed in this chapter. More information is available on the web at www.analytictech.com.

Kruskal, J.B. and M. Wish. 1978. *Multidimensional Scaling*. Newbury Park: Sage Publications.

This is perhaps the clearest book available on the mathematics and interpretation of multidimensional scaling.

Romney, A.K., S.C. Weller, and W.H. Batchelder. 1986. "Culture as consensus: A theory of cultural and informant accuracy." *American Anthropologist* 88(2):313-338.

This is a brilliant paper on the theory of consensus analysis. A seminal article in the field.

Scott, J. 1991. *Social Network Analysis: A Handbook*. Newbury Park: Sage Publications.

Scott's book is a popular introduction to the techniques of social network analysis. It discusses everything from data management techniques to advanced analytical methods.

Spradley, J. 1979. *The Ethnographic Interview*. NY: Holt, Rinehart & Winston.

Spradley's book is perhaps the definitive book on interviewing technique in the context of cultural domain analysis. Extremely well-written with lots of examples.

References

- Borgatti, S.P. 1992. *ANTHROPAC 4.0*. Columbia, SC: Analytic Technologies.
- Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.
- Boster, J.S. 1994. "The successive pilesort." *CAM: Cultural Anthropology Methods Journal* 6(June):11-12.
- Boster, J.S. and J.C. Johnson. 1989. "Form or function: A comparison of expert and novice judgments 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.
- Casagrande, J.B. and K.L. Hale. 1967. "Semantic relations in Papago folk-definitions." In *Studies in Southwestern Ethnolinguistics*. D. Hymes and W.E. Bittle, (eds.) Pp. 165-196. The Hague: Mouton.
- Chavez, L.R., J.M. McMullin, R.G. Martinez, S.I. Mishra, F.A. Hubbell. 1995. "Structure and meaning in models of breast and cervical cancer risk factors: A comparison of perceptions among Latinas, Anglo women and physicians. *Medical Anthropological Quarterly* 9:40-74.
- Dejordy et al
- Everitt, B. 1980. *Cluster analysis*. New York : Halsted Press.
- Gatewood, J. 1984. "Familiarity, vocabulary size, and recognition ability in four semantic domains." *American Ethnologist*, 11:507-527.

- Henley, N.M. 1969. "A psychological study of the semantics of animal terms." *Journal of Verbal Learning and Verbal Behavior*. 8:176-184.
- Kruskal, J.B. and M. Wish. 1978. *Multidimensional Scaling*. Newbury Park: Sage Publications.
- Lounsbury, F. 1964. "The structural analysis of kinship semantics." In H.G. Lunt (Ed.) *Proceedings of the ninth international congress of linguists*. The Hague: Mouton.
- Romney, A.K. and R.G. D'Andrade. 1964. "Cognitive aspects of English kin terms." *American Anthropologist* 66(3):146-170.
- Romney, A.K., S.C. Weller, and W.H. Batchelder. 1986. "Culture as consensus: A theory of cultural and informant accuracy." *American Anthropologist* 88(2):313-338.
- Scott, J. 1991. *Social Network Analysis: A Handbook*. Newbury Park: Sage Publications.
- Spradley, J. 1979. *The Ethnographic Interview*. NY: Holt, Rinehart & Winston.
- Weller, S.C., and A.K. Romney. 1988. *Systematic Data Collection*. Newbury Park: Sage Publications.