Borgatti, S. 1998. Elicitation techniques for cultural domain analysis. In J. Schensul & M. LeCompte (Eds.), The Ethnographer's Toolkit, Vol. 3. Walnut Creek, CA: Altimira Press.

Chapter 1
Elicitation Techniques for Cultural Domain Analysis[1]

by Stephen P. Borgatti

The techniques described in this chapter are used to understand cultural domains (Lounsbury, 1964; Spradley 1979; Weller & Romney, 1988). There are several ways to define a cultural domain. A good starting point is: a set of items that are all of the same type. For example, "animals" is a domain. The members of the domain of animals are all the animals people are aware of, such as dogs, cats, horses, lions, tigers, etc. Implicit in the notion, however, is also the idea that membership in the domain is not solely determined by the individual respondent, but that it exists "out there" either in the language, in the culture or in reality. Hence, the set of colors that a given respondent 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 perceptions rather than preferences. Hence, "my favorite foods" is not a cultural domain, but "things that are edible" is. However, a better rule of thumb is that cultural domains are about things which in principle have a right answer which is universally true. Consider, for example, the cultural domain of animals. If asked

[1]I am grateful to H. Russell Bernard, Pertti Pelto, A. Kimball Romney, and Gery Ryan for helping shape my views on cultural domain analysis, which is not to say they necessarily agree with anything I have written. I am also grateful to Mark Fleisher and to John Gatewood for giving me permission to use their data to illustrate concepts.

whether a tiger is an animal, for example, the respondent feels that she is discussing a fact about the world, not about herself. In contrast, the truth about whether ice cream belongs in the category of her favorite foods is local: it is not necessarily true for others. Hence, cultural domains are in principle shared.

This does not mean that for a domain to be called "cultural", each member of a group must agree that each item is really in the domain. The extent to which a cultural domain is shared is an empirical question. Conversely, agreement about a set of items does not imply it is a cultural domain. If we ask informants in our own culture about their 10 favorite foods and every one of them gives the same list, it is still not a cultural domain because asking for personal choices is not the kind of thing which in principle could be a cultural domain.

Another aspect of cultural domains is that the items are linked by semantic relations. In other words, the domain has internal structure. For example, for any pair of items in a domain, we can ask respondents how similar they are, or whether they are more similar than another pair, and this would not be an unanswerable question. Similarity is a binary relation. The elements of a cultural domain also have attributes, such as size, color, and goodness. These can be seen as relations (e.g., is bigger than, is the same color as, is better than), but unlike similarity these relations can be reduced to a single property of each item. In contrast, similarity is indivisible: it is not a property of the item but of the pair of items.

There is an expectation that 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 exist subdomains. For example, if the domain of "animals" contains the items "squirrel", "ant", and "mammal", it will not make sense to informants to be asked whether squirrels are faster than mammals. The real test for items of different levels of contrast, however, is to look at the "is a kind of" relation. 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 gloss for a subdomain.

Even if all the items are of the same level of contrast, however, the inability to apply an attribute to all items still suggests that the domain has a hierarchical taxonomic structure and that the attribute belongs to items in one particular class. For example, the attribute "color of antlers" can be applied to some animals, but not to others. This means that the domain of animals contains at least two types — animals with antlers and animals without — and within the set of those with antlers, we can ask what color the antlers are.

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 position of the items in the domain structure. All of the techniques have been incorporated into a commercially available computer program called Anthropac (Borgatti, 1992).

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 "a kind of fruit". Then you say "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 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 the informant. In principle, the freelists from different respondents (belonging to the same culture) should be comparable and similar because the stimulus question is about something outside themselves and which they have in common with other members. In contrast, an open-ended question could easily generate only unique answers.

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. This is repeated with about 30 informants (but the number depends on the amount of cultural consensus — if every informant gives the exact same answers, you only need one). Then we count the number of times each item is mentioned and sort 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 1.

Table 1. Top 20 Animals Mentioned, Ordered by Frequency

Rank	Item Name	Frequency	Resp. %	Avg. Rank	Smith's S
1	CAT	13	93	4.85	0.758
2	DOG	13	93	3.62	0.814
3	ELEPHANT	10	71	8.20	0.471
4	ZEBRA	9	64	11.11	0.341
5	SQUIRREL	8	57	12.88	0.266
6	TIGER	8	57	5.50	0.440
7	COW	7	50	10.86	0.291
8	FISH	7	50	13.29	0.224
9	BEAR	7	50	7.00	0.366
10	WHALE	7	50	13.86	0.215
11	DEER	7	50	11.29	0.259
12	MONKEY	7	50	10.00	0.293

13 GIRAFFE	6	43	12.00	0.228
14 GORILLA	6	43	14.67	0.146
15 MOUSE	6	43	8.83	0.299
16 SNAKE	6	43	13.33	0.180
17 LION	5	36	11.00	0.199
18 ANTELOPE	5	36	11.00	0.197
19 LEOPARD	5	36	12.40	0.166
20 TURTLE	5	36	7.80	0.238

One way to determine whether it is necessary to interview more informants, recommended by Gery Ryan[2], is to compute the frequency count after 20 or so informants, then again after 30. If the relative frequencies of the top items have not changed, it 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.

Once you have finished collecting the freelists and have computed the sorted frequencies, the first thing you will notice is that there are a few items that are mentioned by many respondents, and a huge number of items that are each mentioned by just one person. As discussed near the end of this section, domains seem to have a core/periphery sort of structure with no absolute boundary. The more respondents you have, the longer the tail grows, though ever more slowly. 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 just one person.

From a practical point of view, of course, it is usually necessary to determine a boundary. One natural approach is to count as member of the domain all items that are mentioned by more than one respondent. This is logical because items mentioned by just one person don't fit the notion of

---

[2]Personal communication

a domain being shared. However, this 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. This is most easily done by plotting the frequencies in what is known as a "scree plot" (see Figure 1). 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.

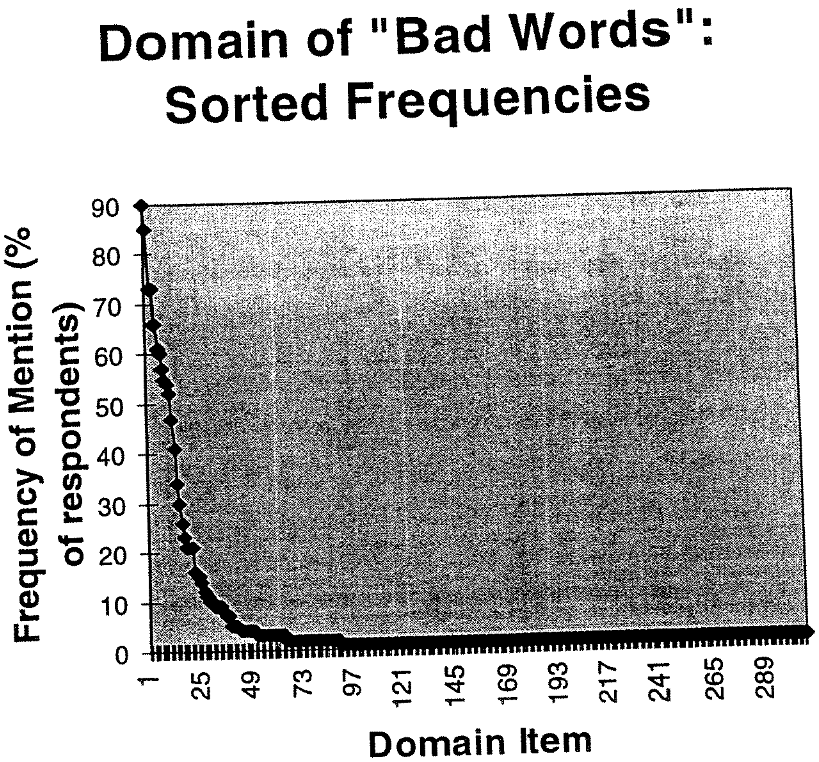

Figure 1. Sorted frequency of items in a freelist.

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 is 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 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 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 may be at a lower level of contrast than other terms in their lists. Occasionally, these kinds of items may lead to a generalization in the respondent's mind, so that they list "large dog", "small dog", and "hairless dog". In general these are not a problem because they will be mentioned by just one person.

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, there are several analyses that can be done which may be quite interesting.

Once we have a master list of all terms mentioned, we can arrange the freelist data as a matrix in which the rows are informants and the columns are items (see Table 2). 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.

Table 2. First 10 Rows and 8 Columns of Respondent-by-Item Freelist Matrix

	Cat	Dog	Elephant	Zebra	Squirrel	Tiger	Cow	Fish
1	1	1	1	1	1	1	1	0
2	1	1	1	1	1	0	0	0
3	1	1	0	1	0	0	0	0
4	1	1	1	0	1	0	1	1
5	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1
7	1	1	0	1	1	0	1	1
8	1	1	1	1	1	1	1	1
9	1	1	1	1	0	1	1	0
10	1	1	1	1	1	1	1	1

The number of items in a freelist is interesting in itself. Although confounded by such variables as respondent competence 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 the names of sociological theories of deviance they can think of, we should find that professional sociologists have longer lists. Similarly, dog fanciers are likely to produce longer lists of dog breeds than ordinary people. In this sense we can use freelist length to construct a crude measure of "cultural competence", where competence is used here in the special sense of access to received wisdom (conventional thinking). However, rather than use the raw number of items mentioned (since many could be wholly idiosyncratic), we weight the items mentioned by a given respondent by the proportion of all respondents who mentioned it. Hence, if respondents A and B both list 10 items, but respondent A mostly lists the items that everyone else gave while respondent B mostly lists unusual items, then respondent A will be assigned a higher "cultural competence" or "cultural centrality".

Table 3. Computing co-occurrence.

	A	B	C	D
1	1	0	1	1
2	0	1	0	0
3	1	1	1	1
4	1	1	0	0
5	0	0	1	0
6	1	1	0	0

a) Respondent-by-item freelist matrix X

	A	B	C	D
A	4	3	2	2
B	3	4	1	1
C	2	1	3	2
D	2	1	2	2

b) Item-by-item matrix of co-occurrences, P = X'X

Moving our focus now from the away from the respondents and onto the items, an important avenue of analysis is to examine the co-occurrences among freelisted items. Using matrix algebra, we can compute the transpose[3] of the respondent-by-item freelist matrix and then pre-multiply it with the matrix itself (in matrix notation we are computing P = X'X). This results in an item-by-item matrix (P) in which the cells record the number of respondents that listed both the row item and the column item in their lists (see Table 3). Taking the product of the transpose with the

[3]Where the columns become the rows and the rows become the columns, resulting in an item-by-respondent matrix.

matrix is equivalent to comparing each pair of columns in the original respondent-by-item freelist matrix by counting up the number of times both columns (items) have a "1" for the same row (respondent).

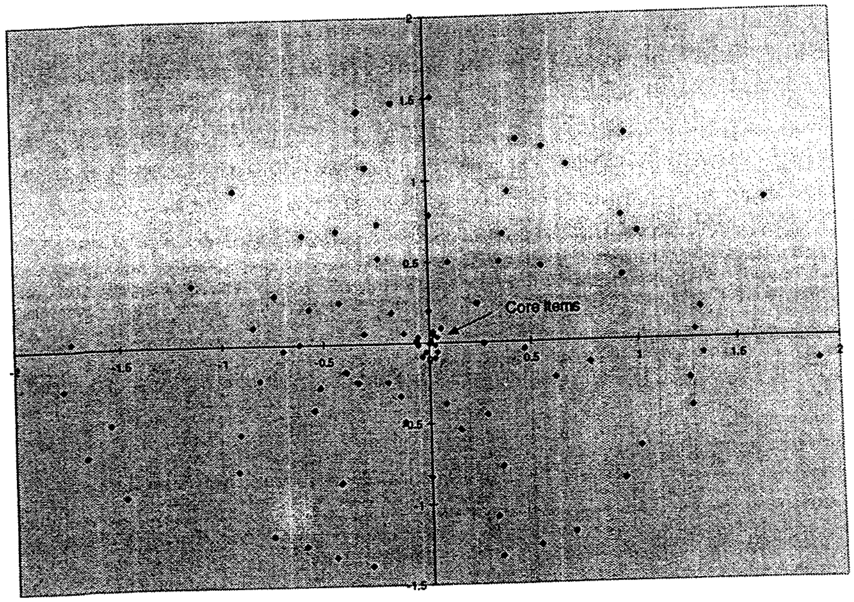

Figure 2. MDS of 90 "bad words" based on their co-occurrence in freelists. Core items of the domain are found in the middle of the space. Items mentioned by only a few respondents are on the periphery.

The resulting matrix can be displayed via multidimensional scaling, as shown in Figure 2. Typically, such maps will have a core/periphery structure in which the core members of the domain (i.e., the most freq-uently 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.[4]

A number of other analyses of freelist data may be made as well. As Henley (1969) noticed, the order in which items are listed by individual respondents is not arbitrary. Instead, we find that respondents will produce runs of similar items, one quickly following the other, followed by a visible pause, and then a new run will begin of different items. Even if we do not record the pauses, 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 seems to affect position on the list. First, 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 everyone. Hence, average position on the list (average rank), is correlated with frequency of mention, across all respondents.

Second, 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 an 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 respondent did not mention an item at all, we enter 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 people's lists, when they occur at all. This can then be

[4]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.

displayed using multidimensional scaling. It should be noted, however, that if the primary interest of the study is to uncover similarities among the members of a domain, it is probably wise 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 screw up. 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 the 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 characteristics that are shared by all members of the domain and which therefore might otherwise go unmentioned.

Pilesorts

The pilesort task is used to primarily to elicit from respondents judgements 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 pilesort sort technique. We begin with the *free pilesort.*

The typical free pilesort 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 animals, we might have a set of 80 cards, one for each animal. 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 animals. 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 animal. Ask them to set aside any cards containing items 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, I simply take a blank card, write the name of the item on the card, and let them 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 input correctly, since usually having an item appear in more than one pile is 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, or even the items themselves (e.g., "bugs"). However, it is my belief that, 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 hidden 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.).

Normally, the pilesort exercise is repeated with at least 20 respondents, although the number depends on the amount of variability in responses. For example, if everyone in a society gives exactly the same answers, you only needed 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.

The data are then 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 4). 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 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 pair are of intermediate similarity.

Table 4. Percent of Respondents Placing Each Pair of Items in the Same Pile.

	Frog	Sala-mander	Beaver	Raccoon	Rabbit	Mouse	Coyote	Deer	Moose
Frog	100	96	6	2	2	0	0	2	2
Salamander	96	100	4	0	0	2	0	0	0
Beaver	6	4	100	62	65	56	17	25	13
Raccoon	2	0	62	100	71	58	23	29	15
Rabbit	2	0	65	71	100	75	17	27	15
Mouse	0	2	56	58	75	100	17	15	10
Coyote	0	0	17	23	17	17	100	21	15
Deer	2	0	25	29	27	15	21	100	77
Moose	2	0	13	15	15	10	15	77	100

*Data collected by Sandy Anderson under the direction of John Gatewood.

This interpretation of agreement as monotonically 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. One such model is as follows. Each respondent has 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 judgement. 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 people think of items as bundles of features or attributes. When asked to placed 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 quite plausible. However, even if either or both is true, there is still a problem with the interpretation of intermediate percentages of respondents placing a pair of items in the same pile. Just because intermediate similarity implies intermediate consensus does not mean that intermediate consensus implies intermediate similarity. For example, suppose half the respondents clearly understand that shark and dolphin are very similar (especially in contrast to land animals) and place them in the same pile, while the other half are just as clear on the idea that shark and dolphin are quite dissimilar (because one is a fish and the other is a mammal). The result would be 50% of respondents placing shark and dolphin in the same pile, but we would NOT want to interpret this as meaning that 100% of respondents saw shark and dolphin as moderately similar. In other words, the measurement of similarity via aggregating pilesorts depends crucially on the assumption of underlying cultural consensus, in the special sense defined by Romney, Weller and Batchelder

(1986). There cannot be different systems of classification among the respondents or else we cannot interpret the results.

To some extent, this same problem afflicts the interpretation of freelist data as well. Items that are mentioned by a moderate or small proportion of respondents are assumed to be peripheral to the domain. Yet, this interpretation only holds if the definition of the domain is not contested by different groups of respondents. This could happen if we unwittingly mix respondents from very different cultures.

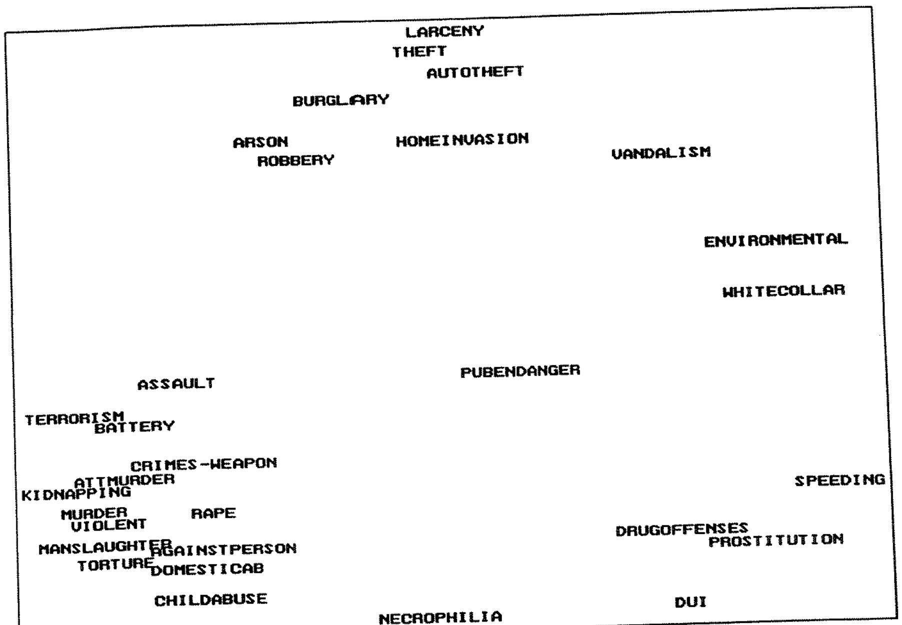

Figure 3. Pilesort of 30 crimes, represented by multidimensional scaling.

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 4. This matrix can then be represented spatially via non-metric multidimensional scaling, or analyzed via cluster analysis. Figure 3 shows a multidimensional scaling of pilesort similarities among 30 crimes collected by students of

Mark Fleisher[5]. In general, the purpose of such analyses would be to (a) reveal underlying perceptual dimensions that people use to distinguish among the items, and (b) 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 pilesort[6]. 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 semantically related, 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. Clearly, informants are good at telling you whether a sentence in their native language is grammatically well-formed, but this does not necessarily imply that they can tell you the syntactical rules that govern the language.

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 predict the results. In general, scientists build descriptions of reality (theories) that are expected to make accurate pre-

---

[5]The data were collected specifically for inclusion in this chapter by: Jennifer Teeple, Dan Bakham, Shannon Sendzimmer, and Amanda Norbits. I am grateful for their help.

[6]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.

dictions, but are not expected to literally be 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 can construct a computer that can converse in English so well that it cannot be distinguished from a human we will be forced to grant that the machine understands English, even if the way it does it cannot be shown to be the same as the way humans do it.

To discover underlying dimensions we begin by collecting together the attributes elicited directly from respondents. Then we look at the MDS picture to see if the items are arrayed in any kind of order that is apparent to us.[7] For example, in the crime data shown in Figure 3, 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 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 respondents[8] 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

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

[8] Either as part of the pilesort exercise, or by showing the MDS map to informants and asking them what to make of it.

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 respondents' thinking). For example, a pilesort of a wide range of animals, including insects, birds, mammals, reptiles, fish, etc., will result in tight clumps in which all the insects are seen as so much more similar to each other than to other animals that no internal differentiation can be seen. 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.

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 and zeros. We can, in principle, compare two respondents' views by correlating these strings. However, there are problems caused by the fact that some people have more piles than others[9]. For an example of one problem,

---

[9]This is known as the "lumper/splitter" problem.

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 two splitters can have fairly high correlations 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). There are some analytical ways to ameliorate the problem, but these are beyond the scope of this chapter.

The best way to avoid the lumper/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.[10] 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 lumper/splitter problem, he begins with a free pilesort. If the respondent makes N piles, the researcher then asks the respondent to split

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

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 are time-consuming and can potentially require a lot of time to record the data (while the respondent looks on). 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.

Triads

An alternative to pilesorts for measuring similarity is the triad test. Here, 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:

Mouse	**Elephant**	**Rat**

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 times[11], each time "against" a different item. If a pair of items is really

[11]Again, N is the number of items in the domain.

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:

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*.

The problem with presenting all possible triples is that there are $N(N-1)(N-2)/6$ of them, which is a number that grows with cube of the number of items. If the domain has 30 items in it, the number of triples 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 sample from the population of all possible triples. However, a random sample would allow some pairs to appear in several triples, and allow others not to occur it 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* design (BIB design). 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). When λ equals N-2, we have the complete design (where all possible triples occur). 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 λ=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 homogeneous sample a completely different triad test, based on the same domain 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.[12]

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 check their consistency, 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 that they are using varying criteria to pick the item most different in each triple. They feel that they are not doing a good job. This is in fact not true but it is difficult to convince respondents. 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 judgement regarding the similarities among a certain set of items).

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 to present informants with a small random sample of triples. 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, for this triple:

Seal	**Shark**	**Dog**

The respondent might indicate that a seal is like a combination of shark and dog. It has the shape of a shark (aerodynamic, flippers/fins instead of legs), and lives in the water like a shark, but is a mammal, like a dog, and barks like a dog. The shark differs from the seal and dog in that it is a fearsome predator, whereas seals and dogs are cute. The dog differs from the other

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

two in being a land animal. It can quickly be seen that it only requires a handful of triples to elicit dozens and dozens of attributes.

Conclusion

I have presented three 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. In addition, I have touched on the use of multidimensional scaling to graphically illustrate the structure of the domain, and locate each item's position in that structure.

Implicit in these data collection and analysis techniques is a distinctive notion of cultural domain as a system or network of items related by families of links (semantic relations). Thus, a cognitive domain has internal structure, and it is the position of items within this structure that distinguishes the items from each other and gives them their unique meanings. Viewing domains in this manner emphasizes their fundamental similarity to other systems, such as economies, societies, ecologies, machines, and brains. Consequently, I would suggest that to obtain additional tools for studying cognitive domains we should look to those disciplines that have explicitly conceptualized their objects of study as systems or networks. In particular, I would recommend the techniques of social network analysis, which are reviewed by Scott (1991).

References

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

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., 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.