

1 Surveying Culture

In traditional sample surveys, one person's measurements are presumed unpredictable from measurements on another person. For example, one person's age offers no clues about another person's age. Consequently, knowledge about variations and central tendencies in a population has to be built up piecemeal from many individual measurements. A statistical methodology has developed to guide the process of making inferences about a population from a sample of people (Kish 1965), and this methodology is so fundamental in social science that it frequently is treated as the only viable framework for acquiring and interpreting survey data.

However, work in psychological anthropology (Romney 1994; Romney, Batchelder, and Weller 1987; Romney, Weller, and Batchelder 1986) and in sociology (Rossi and Nock 1982) has clarified that some surveys are conducted in order to ascertain normative features shared by everyone, such as beliefs and sentiments deriving from culture. In this case, information from one person does predict information from others: For example, one person in a traditional society reporting that fathers usually are husbands foreshadows others saying the same thing. When people all provide the same information, it is redundant to ask a question over and over. Only enough people need be surveyed to eliminate the possibility of errors and to allow for those who might diverge from the norm. Romney, Weller, and Batchelder's (1986) mathematical-statistical analysis of ethnographic data gathering demonstrated that as few as a half-dozen expert respondents can provide a very clear picture of some types of shared norms.

In a survey of a population of individual subjects, variability is sought in the answers to every item, and items that would yield meager variability generally are avoided because they seem uninteresting. However, surveys that have been engineered to maximize variability fail to reveal norms.

The measurement practices that have dominated the fields of sociology and social psychology seem designed to avoid finding empirical evidence for norms, beliefs about which there is some large degree of popular consensus. Indeed, when we find measures on which individual subjects agree, we tend to discard them because they will reveal little about differences among individuals. . . . We

prefer items with maximum variance and hence with corresponding minimum agreement among subjects, a strategy that makes good sense in measuring inter-individual variation in the amount of cognitive achievement, but may not make good sense for sociologists who are trying to understand the overall normative patterning of human behavior. Indeed, it is often precisely those measurement instruments that we conventionally reject as useless that are most indicative of norms. Thus few social scientists would ask respondents whether they approve of murder because we would expect almost everybody not to. (Rossi and Rossi 1990, p. 160)

In population surveys, large variances in variables are sought to register the extent and shape of social controversies and to enable causal inferences. However, a survey of culture is intended to build a descriptive database regarding norms, and therefore lack of variability on every item is the ideal, since response variation confounds the delineation of norms. Surveys of cultures do seek variations, but across items rather than across respondents. For example, when describing a culture, the difference in the average evaluation of doctor versus the average evaluation of robber is the interesting variation, rather than differences in evaluations of doctor or robber by respondent A versus respondent B. Stable individual differences actually count as errors when surveying cultural norms because individual deviance obscures underlying uniformity. Consequently, usual notions of reliability no longer hold in culture surveys. In fact, as shown later in the book, an ideal item for assessing a norm would have zero reliability in the traditional sense!

An occasional complaint about surveys of culture is that the respondents providing the data are too few and are not selected randomly from general populations. However, this criticism—posed from the perspective of traditional survey methods—is tangential for ethnographic data gathering. The aim in ethnography is not to describe a population of individuals but, instead, to describe a culture that is being reproduced within some group. Properly chosen respondents are those whose responses are quintessential for their culture, and the more normative the respondents' beliefs and sentiments, the fewer of them are needed to obtain an accurate view of the culture. Whereas there is no notion of respondent goodness in surveying a population, other than representativeness of the sample as a whole, proficiency in the target culture is a key desideratum in choosing respondents for a survey of culture. Indeed, lack of cultural expertise is a legitimate basis for culling respondents or for assigning less weight to a respondent's answers, as we demonstrate later in the book.

The usual procedure in a survey of a population, of sampling individuals in a political unit, is inefficient in surveys of culture because desirable respondents for a survey of culture typically are not evenly distributed throughout a politically defined population. Rather, the best respondents for a culture survey are persons who reproduce the culture—the denizens of settings where the culture is being regenerated. For example, if interested in the middle-class culture that sustains the basic social institutions of American society (e.g.,

commerce, education, medicine, politics), a researcher would seek representative settings in which the activities of those institutions occur and question individuals at those sites about cultural matters. If interested in the culture that sustains black community life, the researcher would go to the homes, churches, and leisure venues where black culture is reproduced and question people at those sites. This emphasis on behavior settings in which culture is reproduced contrasts with the sampling frame in traditional survey research, where individuals themselves are the sampling units, even when geography is a practical consideration in sampling schemes for acquiring respondents.

So surveys of culture differ fundamentally from surveys of populations in at least three respects: (1) questions are asked about matters of agreement rather than about issues that generate diversity of response; (2) respondents are graded on the basis of their expertise; and (3) respondents are acquired by visiting settings where cultural reproduction takes place rather than by random sampling of people in a large geographic area delineated by political boundaries.

Since surveys of populations and surveys of cultures have distinctive goals and methods, one kind of study cannot be adapted easily to the purposes of another. A representative sample from a population rarely is simultaneously a sample that is culturally homogeneous and well inculcated in the culture of interest, so population surveys generally provide little information about the vast number of shared norms and understandings in the dominant culture. On the other hand, culture surveys aim at homogeneity in responses among those who are most knowledgeable about the culture, so they typically provide meager data about variations among diverse individuals within the society at large, and they are poor bases for explaining individual differences.

1.1 CASE STUDIES OF CULTURAL SURVEYS

Subjective culture—the knowledge and motivational structures a society provides for its indigenes—is the focal concern in this book. Surveys can be fielded to assess norms of material culture; for example, Chapin (1932) described how interviewers visiting homes could record household furnishings and equipment. Surveys can also assess some kinds of behavioral norms, such as patterns of pronunciation in American English (Labov, Ash, and Boberg 2006). However, the massive social science emphasis on cognition and emotion during the last quarter century has made the study of subjective culture a richly developed area, worthy of methodological consideration.

Important aspects of subjective culture include the categorization system that establishes culturally acknowledged realities, the implicit rules that interrelate cultural categories in ways that foster sensible decision making, and the sentiment system that orders cultural categories in terms of value, significance, and urgency. In the following case studies, I illustrate how social surveys have been applied in each of these areas.

1.1.1 American Regional English

In 1965, lexicographer Frederic Cassidy fielded a five-year survey of American variations in word usage that acquired such abundant data that the ultimate product, the *Dictionary of American Regional English* (*DARE*), still had not been completed at Cassidy's death in 2000. The first volume of *DARE*, 1,059 pages describing procedures and conventions and definitions for words beginning with the letters A through C, appeared 20 years after the survey began. Volumes 2 and 3, edited by Cassidy and Joan Hall and covering D through O, appeared in 1991 and 1996. Volume 4, edited by Hall and covering P through Sk, appeared in 2002. Volume V, covering the remainder of the alphabet, still was unpublished as of 2009. Another planned volume will contain a bibliography, maps, responses to the questionnaire, and such. (All of the volumes are available from Harvard University Press.)

The general concern of the *DARE* project was folk usage, or language learned at home from relatives or learned in the community from friends and associates (Cassidy 1985, p. xvi). There was no interest in words that could be considered Standard English as learned in schools or acquired from books, or as documented in conventional dictionaries; the project did not deal with technical or scientific words.

The questionnaire administered face to face to respondents in this project was massive. "If every question had elicited one response each time it was asked, there would have been 1,850,694 responses—but multiple answers brought the total closer to 2,500,000" (Cassidy 1985, p. xiv).

Intended for use in personal interviews, the *DARE* Questionnaire (QR) begins with the neutral subject of time in order to allay possible suspicions of some hidden purpose on the part of the investigator. Next come weather and topography, equally neutral and safely concrete. Houses, furniture, and household utensils follow, with dishes, foods, vegetables, and fruits. And so the questions continue to more abstract topics: honesty and dishonesty, beliefs, emotions, relationships among people, manner of action or being, and so on—41 categories in all with a total of 1847 questions. (Cassidy 1985, p. xii)

Some questions described denotative targets and asked respondents to supply the words that they themselves used to refer to such an object. In this case, "the question, if properly phrased and understood, and the answer, if responsibly given, should ideally produce a reversed definition. For example: 'What do you call a container for coal to use in a stove?' Responses: *coal bucket, hod, pail, scuttle*, and so on. Reversed, this becomes a definition: *scuttle*, a container for coal to use in a stove. The method works relatively well for simple material objects like coal scuttles, less well for abstract things or emotional matters" (Cassidy 1985, p. xiii). Other questions named a category in Standard English and asked respondents to supply local synonyms. For example, one question asked what local names were given to the dragonfly (Cassidy 1985, p. lxix). Respondents gave 79 different replies to this question,

among them *snake feeder, snake doctor, mosquito hawk, spindle, and ear-cutter* (Cassidy 1985, pp. xii–xiii). In a subset of questionnaires, respondents identified wildflowers from color photographs.

Completion of the entire questionnaire required 25 to 30 hours (Carlson 2001, p. 2). Since few respondents could commit such a huge block of their time, each questionnaire typically was divided among several respondents from the same community—2.8 respondents on average.

Besides obtaining respondents' answers to the questions, fieldworkers recorded individual information regarding the respondent: name, address, gender, race, age, education, amount of travel, chief occupations, associations, family background on both sides, and a description of the respondent's speech and attitudes toward language. Fieldworkers also prepared a brief description of the respondent's community. For most respondents, a tape recording was made of the respondent speaking freely for 20 minutes or more on a familiar topic and also reading a short story known to expose speech variations.

The fieldworkers mainly were graduate students who had some formal training in English and linguistics and who could make phonetic transcriptions; a few undergraduates and faculty members also worked in the field (Cassidy 1985, p. xiii). In total, the fieldworkers consisted of 51 men and 29 women. To deal with a conceivable biasing factor, both white and black fieldworkers interviewed black respondents; however, results appeared to be unaffected by the race of the interviewer (Cassidy 1985, p. xiv). Some of the fieldworkers have published interesting and valuable reminiscences of their fieldwork experiences in the *DARE Newsletter*, available at the *DARE* Web site (von Schneidemesser 2008).

The *DARE* project sampled communities rather than individuals, with relatively few communities being selected in sparsely populated Western states and many communities being selected in populous Eastern states; for example, New York City had 22 *DARE* communities (Carver 1985, p. xxiii).

Communities were chosen in each state, the number proportional to the population and taking settlement history into account. . . . The aim was to choose relatively stable communities, distributed according to the states' composition, and communities of various types, so that the aggregate would reflect the makeup of each state's population. . . . *DARE* recognized as a "community" any group of people living fairly close to each other and sharing the same commercial facilities, social organizations, and the like. Even within metropolitan areas such communities, or subcommunities, exist with a sense of focus based on ethnic, religious, and other characteristics. Contrariwise, quite small independent rural communities, though close together, may keep themselves apart on similar grounds. (Cassidy 1985, p. xiii)

A total of 1,002 communities were selected in the 50 states of the United States, and one questionnaire was completed in each community, often by several different respondents. "Neither the choice of communities nor that of

informants was randomized; on the contrary, the intention was to maximize the collection of materials by going to the places and people most likely to furnish the largest amount of appropriate data" (Cassidy 1985, p. xiv).

Fieldworkers chose informants, and while they attempted to balance gender, age, education, and race for their geographic area, inculcation into the community was a paramount consideration. "To qualify as a *DARE* informant, a man or woman had to have been born in the community represented or very close by, and could not have traveled or stayed away long enough for his or her speech to be affected. [Choices were made] with a deliberate weighting toward older people. Folk language is traditional, and older people remember many things that young ones have never heard of" (Cassidy 1985, p. xiv). Fieldworkers were advised to search for ideal informants by contacting "local churches, branch libraries, real estate offices, funeral parlors, gas stations, car dealers, and even neighborhood bars" (Carlson 2001, p. 2).

The 2,777 respondents in the study consisted of 1,368 men and 1,409 women. Sixty-six percent were 60 years of age and over (some over 90), 24 percent were middle-aged (40–59), and 10 percent were in the age range 18 to 39. Levels of education were: 3 percent less than fifth grade, 24 percent no more than one year of high school, 41 percent at least two years of high school, 31 percent at least two years of college, and 1 percent unknown. Racially, the respondents were 92.7 percent white, 6.7 percent black, with 0.3 percent each of American Indians and Orientals (Cassidy 1985, pp. xiii–xiv).

The *DARE* survey focused on identifying regional words or meanings, and in this sense it was not a survey of a single culture but of multiple subcultures in the United States, as manifested in America's variant Englishes. Language variations in the United States are not so distinctive as variant Englishes around the world (Bhatt 2001), but American usages do vary in ways that reflect the different social histories and interests of groups in different geographic settings.

Rather than specify the distinct regions of language usage by fiat, the *DARE* researchers took an empirical approach based on a novel form of mapping. "On [a] conventional map, the [*DARE*] communities are concentrated in the most populous areas of the country. . . . At the same time the communities of the western states are widely scattered. The *DARE* map, by contrast, compresses the western states while expanding the more populous eastern states, creating a relatively uniform distribution of communities across the map. This makes it easier to see the clustering of communities where a given response is recorded" (Carver 1985, p. xxiii). In essence, *DARE* maps allow any viewer to partition response data visually into culture regions. "The *DARE* map is essentially a scatter diagram that economically illustrates degrees of clustering—that is, degrees of regionality" (Carver 1985, p. xxvii). The unique maps appear throughout the dictionary volumes, and a collection of them is planned for the final volume.

Cassidy (1985, p. xiii) remarked that the *DARE* project assembled unequalled data on living American speech, with the data containing a great

deal of information on syntax that went unexploited in constructing the dictionaries. The data additionally could be analyzed to study American folk culture. Reading through the *DARE* questionnaire is a humbling experience regarding one's knowledge of American folk culture, which suggests that future analyses could examine the percentage of questions that respondents were able to answer satisfactorily, after eliminating answers with special codes,¹ in order to correlate the extent of folk knowledge with social variables such as age. Data from communities could be analyzed to identify fissures in American folk culture. While the *DARE* maps allow responses to a single question to be partitioned into homogeneous groups, a multivariate analysis of communities' responses to all questions, comparable to the Q-factor analyses discussed in Chapter 7, would partition communities and the content of folk culture simultaneously into general divisions, whether or not related to geography. Such analyses will eventually become possible when the plan to put the *DARE* corpus on the Internet is implemented.

1.1.2 Obligations to Kin

Sociologists Alice and Peter Rossi undertook a survey of family relations among people in the Boston metropolitan area in 1984 and 1985, publishing the results of their study in a 1990 book entitled *Of Human Bonding*. Their study focused on three main questions: How do individuals change over the life course? What determines variations in parent-child solidarity? How are obligations to help others normatively structured? Analyses relating to the last question—the one of interest here—generated findings about norms of help-giving in Boston, and probably elsewhere in America, which allowed the authors to portray how obligations vary from parents and children to siblings, grandparents, aunts, cousins, and on to nonkin such as friends, neighbors, and former spouses.

Data for the study of norms was obtained after an hour-long face-to-face interview that dealt with life-course matters and relations between parents and children. At the end of the interview the respondent took about 15 additional minutes to work through a questionnaire booklet containing 31 brief vignettes, preceded by a practice vignette that the interviewer used to instruct the respondent in the task. While the interviewer checked over answers that had been obtained in the interview, the respondent read through the vignettes and provided ratings about what level of help would be appropriate in each situation.

Rossi and Rossi (1990, p. 45) described items in the vignette questionnaire as follows:

¹Interviewers entered special codes when they had to prompt for an answer and when they had doubts about the response; they also entered special codes when the respondent claimed no such word was used locally and when the respondent gave no response (Cassidy 1985, p. xiv).

A typical vignette might read:

“Your married brother has undergone major surgery and will be disabled for a very long time. This problem is straining his financial resources. How much of an obligation would you feel to offer him some financial help?”

A numeric rating scale follows this question, ranging from “0” (No obligation at all) to “10” (Very strong obligation). Two types of obligation ratings were built into the design: an *instrumental* obligation, in the form of financial help, as illustrated above, or an *expressive* obligation, for which the question was “How much of an obligation would you feel to offer him comfort and emotional support?”

The illustrative vignette above was in the category of crisis events involving a traumatic event in the life of that kinperson creating a situation of need that called for financial help or for comfort (Rossi and Rossi 1990, p. 162). Other vignettes dealt with celebratory occasions calling for recognition or appreciation, and ratings were made on scales relating to obligation to give something appropriate to the occasion or to visit (Rossi and Rossi 1990, pp. 162–164).

The 31 vignettes considered by a respondent were a small sample of the grand total of 1,628 vignettes presented in the survey. Different respondents got different sets of vignettes, each booklet being constructed uniquely as a random sample from the total population of vignettes. Ratings by different respondents later were pooled to analyze all vignettes together.

The 1,628 vignettes were created by conjoining 74 focal characters with various types of crisis or celebratory events. The focal character of each vignette was designated in terms of three dimensions: form of kinship, gender, and marital status.

Using the respondent as the reference point, the kin designated ranged from four grandparents, each described as a grandmother or grandfather, whether in the maternal or paternal line and whether married or widowed, through the parental generation (e.g., mother father, uncle, or aunt), through the respondent’s generation (siblings and cousins) to children and grandchildren. Each kin type was further specified by gender and marital status. Children and grandchildren were implicitly described as adults, the other kin being explicitly defined as such. To provide a contrast with the level of felt obligation toward nonkin, we also included “friends” and “good neighbors” and “former spouses” as designated categories. (Rossi and Rossi 1990, p. 164)

Eight types of crisis situations appeared in the vignettes, including major surgeries, serious personal troubles, losing everything in a household fire, and unemployment. Three types of celebratory events appeared: winning an award, having a birthday, and moving into a new place.

The overall survey involved face-to-face interviews with 1,393 respondents obtained via probability sampling procedures. The overall survey also obtained

telephone interviews with 323 parents of main respondents and 278 adult children of main respondents, but those data were not used in vignette analyses. The vignette questionnaire was completed by 1,176 of the main respondents (84.4 percent).

About 49 percent of the respondents were age 19 to 40; about 9 percent were over 70 years of age. Fifty-eight percent were female. About 62 percent were currently married, and the rest were about evenly divided between never married and previously married. The respondents were overwhelmingly white (94 percent), Anglo (70 percent), and Christian (80 percent, two-thirds of whom were Catholic).

The various kinds of obligations—financial aid, emotional support, gift giving, visiting—conceivably could be complementary types of solidarity helping provided to different sets of relatives, but the Rossis found that the opposite is actually the normative rule (1990, pp. 169–170).

Offering comfort and offering financial aid to a given kin tend overall to go hand in hand, respondents stating the relatives to whom they feel strongly obligated to give comfort and emotional support are also highly likely to be the same kin to whom they feel strongly obligated to offer financial help. . . . [Similarly] kin that induce a strong obligation to visit are also kin that evoke a strong sense of obligation to give gifts, and, correspondingly, kin to whom little or no obligation to visit is acknowledged tend also to be the kin to whom little or no obligation to send gifts is felt.

In essence, each type of kin has a general obligation strength that governs giving emotional support and offering financial aid in crises and that determines gift giving and visiting in happy situations. What differs among the various obligations is the threshold at which they dominate. For example, respondents typically felt more of an obligation to give comfort and emotional support than to provide financial aid, so comfort and emotional support end up being offered in a broader range of crises than is financial aid. For celebratory occasions, making a visit often is less obligatory than giving a gift, at least in the case of more distant relatives (Rossi and Rossi 1990, Table 4.7).

As one would expect, all types of normative obligations are strongest for one's parents or children, and strong obligations to these kin hold across all kinds of situational circumstances. After parents and children, siblings and grandchildren occasion the greatest normative commitment, then daughters-in-law, sons-in-law, and parents-in-law. Grandparents and stepchildren complete the innermost circle of kin. After that, normative obligations to friends often are as great as obligations to kin such as stepparents, nieces and nephews, aunts and uncle, and cousins. Normative obligations are least for ex-spouses, regardless of the type of obligation or the circumstances.

The Rossis classified focal characters in vignettes in terms of their *range*: primary kin (parents and children); first-order kin (connected through another

relative: e.g., siblings, grandparents, and grandchildren); second-order-kin (connected through two other relatives: e.g., aunts and nieces); third-order-kin (connected through three other relatives: cousins); kin related by marriage (in-laws); step kin; friends and neighbors; and ex-kin. They also classified the vignette characters in terms of *depth*, or the other's generation relative to the respondent's: two generations up (grandparents); one generation up (e.g., parents, aunt); same generation (e.g., siblings, cousins, friends, ex-spouse); one generation down (e.g., children, niece); and two generations down (grandchildren).

Great variations in levels of obligations were found among the range categories, with little variation within each category. In the case of the depth categorization, they found higher obligations to descendent than to ascendant generations. Overall, the structure of kin relations, defined in terms of range and depth, largely determined obligations of all kinds, in all types of situations.² The Rossis concluded (1990, pp. 182–183):

There is clearly a robust and consistent underlying structure to normative obligations to kin. Types of situational stimuli or types of obligatory responses show only minor differences, compared to the inherent structure determined by the degree to which respondents are related to the kinperson in the vignette. Obligations radiate out in lessening degrees from the high obligation primary kin, with greater obligations toward descendents in all categories of kin than to ascendants. Secondary affinal kin, acquired through marriage or remarriage, generally evoke greater obligations than distant consanguineal kin. Friendship involves considerable obligation to provide social-emotional comfort, on a par with secondary blood kin, but . . . friends do not stimulate as high an obligation to provide financial aid as do blood and affinal kin.

The Rossis used the range and depth categorizations to compute regression equations predicting the average obligation felt toward each type of vignette character. They summarized these analyses as follows (Rossi and Rossi 1990, pp. 183–185):

The most striking findings are that the handful of coded structural features of kinship accounted for almost all of the variation in the strengths of obligations to the kin in question. The R^2 values for the four kinds of obligation . . . range from .89 to .94, which are considerably higher than ordinarily found in social science research. These findings further indicate that there is a very robust normative structure to American kinship: Obligations to kin vary in a lawful and regular way according to the position of the kinperson in question vis-à-vis ego. . . . Primary kinship relationships are the most obligating, with obligation declining rapidly as the number of links between ego and kinperson increase. Affinal kin are more obligating than step kin, but both are less so than consanguineal kin.

²Malone (2004) found that the kinship structure similarly determines sentiments about relatives when sentiments are measured in the manner described in Chapter 2 of this book.

Finally, kinship relationships that go down the generational ladder are more obligating than those involving the same or ascendant generations. . . . Note also that female kin are slightly more obligating than male kin, a generalization that does not hold for visiting, but does hold for the other obligation types.

In further analyses, the Rossis found that obligations also were determined somewhat by the nature of the person in the connecting link between ego and the focal character. Kin who were related through siblings had lower priority than those related through a parent, child, or spouse; and the obligation level increased when the connecting link between ego and alter was a woman (Rossi and Rossi 1990, p. 207).

Considerable consensus among respondents was necessary to reveal such structured results, but agreement was not perfect concerning levels of obligation to various kin in a range of circumstances. The least variance in ratings occurred for children and parents, the most variance for grandparents, step-parents, and stepchildren, with other kinds of kin and nonkin ranged in between. The Rossis conducted numerous analyses to find causes of variations in obligation ratings.

The ethnicities of respondents had some small impacts on their obligation ratings (Rossi and Rossi 1990, pp. 239–241). Irish and Jews gave somewhat higher obligation ratings than did respondents of British extraction. Relative to respondents of other ethnicities, blacks, Asians, and Portuguese gave relatively high ratings to distant kin.

Respondents' age was an important factor (Rossi and Rossi 1990, pp. 220–225). In general, younger respondents rated all kinds of obligations as stronger than did older respondents. Age 50 was the main breakpoint; after age 50 obligation ratings declined steadily with every decade of aging. Moreover, older respondents rated obligations as closer to zero for every type of kin—primary, secondary, and distant—and even for nonkin. The Rossis could provide no explanation of this phenomenon, although they noted that it was consistent with other data in their survey regarding help-giving between parents and their adult children.

Education explained variation in ratings to some degree, with college-educated respondents rating all kinds of obligations as stronger than did high school graduates, who in turn made ratings farther from zero than did those without high school degrees (Rossi and Rossi 1990, pp. 226–230). The education effect acted independently of the age effect; education accounted for less variation in ratings than did age.

Emotional troubles in a respondent's family of origin (e.g., alcoholism, mental illness, child or sexual abuse) lowered the respondent's obligation ratings for all kinds of kin. On the other hand, many other kinds of early difficulties, such as family quarrels or problems with schools, actually increased respondents' obligation ratings. "The joint effects suggest that adversities that have their roots in events beyond the control of parents—physical illnesses, unemployment, the behavior of children—lead to stronger kin bonding. In

contrast, troubles that have their roots in the behavior of parents (or other adults) lead persons who experience such adversities in their childhood to become adults with lower levels of kin obligations generally” (Rossi and Rossi 1990, p. 235). A respondent’s own divorce also was associated with lower obligation ratings for all kinds of kin (Rossi and Rossi 1990, Table 5.14).

The Rossis summarized their findings as follows (Rossi and Rossi 1990, p. 246):

Virtually all of our respondents feel some degree of obligation to kin as prescribed in our kinship norms. However, some feel strong obligations to kin and in others there are weaker ties to the kindred. Although we cannot explain all of the differences from respondent to respondent, it is also apparent that their childhood families can set down strong or weak patterns, the former by cohesive families presided over by affectionate parents. As adults, the strength of obligations is influenced positively by being better educated, having a strong sense of duty in a variety of roles, and by being an outgoing expressive person.

As will be seen in Chapter 6, many of the Rossis’ findings parallel findings in other studies relating to respondents’ cultural inculcation.

1.1.3 African-American Sentiments

As a part of a larger project devoted to studying the cross-cultural validity of three dimensions for measuring sentiments and to obtaining comparative data on sentiments in more than two dozen cultures (see Chapters 2 and 3), a group of researchers led by Daniel Landis surveyed a broad range of sentiments held by African-American youths living in segregated areas of Chicago during the 1970s. They reported their procedures and a selection of findings in a book chapter (Landis et al. 1976). Additionally, a listing of black sentiment measurements for 611 concepts was sold as a computer printout at the University of Illinois Bookstore during the 1970s (Landis and Saral 1978); see Chapter 3 for more details. The computer printout introduced the African-American study as follows: “The data in this atlas were gathered from lower-class Black male youth in Chicago during 1973–74. Data gathering sites were from high-schools located in center-city and the southern schools. The concepts used as well as the scales were based on the patois in use during the same period (actually 1971–74).”

A goal of the study was to determine if black sentiments reflected a culture different than the dominant white culture recorded in sentiment measures obtained from whites during the 1960s. Landis et al. (1976, pp. 50, 55) noted that some researchers too readily assumed that black and white cultures were the same:

Most American researchers seemed to forget that the United States itself is a very heterogeneous culture. There tended to be an implicit assumption that the

original American English data could be applied willy-nilly to various American subgroups, including black Americans. . . . The statement of the problem here does not mean that we accept (or reject) the hypothesis of black cultural uniqueness. But given the anthropological, historical, and sociological data, such a hypothesis is at least tenable.

Determining if black sentiments reflected a separate culture required Landis and his colleagues to reproduce methodological procedures used in foreign locales, aimed at preventing ethnocentric biases. Only by treating the black sentiments as culturally disparate could it be determined whether or not they were. Accordingly, Landis and his colleagues treated African-American Vernacular English (AAVE) as a language in its own right.

The process of obtaining measurements of black sentiments free of white biases began with a list of 100 nouns common to all of the languages in the larger cross-cultural study (Osgood, May, and Miron 1975, Table 3-1); the list included words such as *house*, *girl*, and *meat*. These words were translated to AAVE by 10 translators,³ “whose ages ranged from late adolescence to early twenties, [and who] had grown up in a lower SES urban ghetto environment. All were training to become teachers in such a setting” (Landis et al. 1976, pp. 57–58).

As Landis et al. (1976, p. 45) remarked: “The words used [in AAVE] are sufficiently at variance from their use in standard American English to provide the potential for serious misunderstandings between some blacks and whites.” For example, translating to AAVE turned a house into a crib, and a girl into a little-moma (Landis and Saral 1978, Table 9-C). Of course, some words (e.g., *meat*) did stay the same.

The first function of the translated nouns was to serve as stimuli eliciting a range of frequently used adjectives in AAVE. The nouns were presented to 100 black high school students, and these respondents were told to give common adjectives used on the street with each noun, in frames like A crib is _, or The _ crib. “The testers were black males, and the instructions that they gave to the Ss stressed that the adjectives should be those likely to be used by the ‘black and beautiful people’” (Landis et al. 1976, p. 58). This resulted in the naming of hundreds of words, the top 200 of which ranged from frequent responses such as *bad*, *good*, *big*, *nice*, *cool*, *eat*, *black*, and *together* to relatively low-frequency responses such as *new*, *scared*, *do-your-thing*, *happy*, *good-looking*, *fun*, and *chicken* (Landis and Saral 1978, Table 2). Sixty adjectives for further study were chosen for their high frequency of usage, their usage with a variety of different nouns, and their complementarity with other adjectives.

³The first three steps of the procedure—translating nouns, eliciting adjectives, and selecting adjective opposites—were performed by New Jerseyites, because that was the original intended locale for the study, but the data were deemed equivalent to data that might be gathered in the final study locale of Chicago.

The purpose of eliciting adjectives was to obtain modifiers that could be turned into anchors for bipolar scales. The process of scale construction was continued by specifying opposites of the adjectives obtained. "The original group of translators was called together and asked to provide the opposites for the 60 adjectives. . . . Again, the emphasis was on the opposites that would be used in everyday street conversation" (Landis et al. 1976, pp. 58–59). Matching adjectives with their opposites and including a set of checkmark positions in between generated a set of 60 bipolar adjective scales (Landis et al. 1976, Table 1). The adjective anchors of these scales ranged from the familiar in standard American English (e.g., *fast–slow*) to pairs that might seem strange (e.g., *straight–stone*).

The 60 bipolar adjective scales were used to rate each of the nouns that had been used previously to elicit adjectives. Rating all nouns on all scales would have been too big a task for any one respondent, so booklets were made up with a random selection of 10 nouns for each booklet. "In constructing the booklets, the 60 scales were randomized on a page in terms of order as well as position (left versus right) of the 'positive' end of the scale. Each booklet then consisted of 20 pages (two pages for each concept, with each page containing 30 scales)" (Landis et al. 1976, p. 59). The number of booklets was doubled by duplicating them with scale order reversed.

Ratings of nouns on the scales were obtained from black youths in Chicago (Landis et al. 1976, p. 59).

Each of the test booklets was then administered to at least 20 black adolescent pupils in the ghetto area of the West Side of Chicago. Although it was originally planned that each *S* would complete an entire booklet (that is, make a total of 600 judgments), this proved impractical given school time constraints and a high absentee rate at test sites. Therefore each *S* completed the rating of at least one concept on all 60 scales. Some, of course, did more.

Factor analyses determined the dimensionality of the scales. The data matrix had 60 columns and a row for each of the rated nouns. Each cell entry was the mean of all ratings that had been made of the noun representing the cell's row, on the scale representing the cell's column.

Three factors, or dimensions, accounted for 52 percent of the variance in the cells (Landis et al. 1976, p. 61). The first dimension—identified as Evaluation—related to ratings on the *good–foul*, *all right–mad*, *hard up–straight*, and *peaceful–ferocious* scales. The second dimension—identified as Potency—related to ratings on the *large–small*, *big–small*, *big–little*, and *wide–frail* scales. The third dimension—identified as Activity—related to ratings on the *active–passive*, *free–tied down*, *fast–slow*, and *loose–tight* scales.

Landis and colleagues conducted a bicultural factor analysis that analyzed mean ratings of the 100 nouns on black scales used by blacks along with mean ratings on white scales used by whites (Landis and Saral 1978, Table 5). The bicultural analysis revealed whether factors emerging in the black data were

similar to factors emerging in white data. Two correlated Evaluation factors emerged for blacks. One of these, with the scales *clean–nasty*, *jam–bad scene*, *good–foul*, and *together–wrong*, was the same as the Evaluation dimension defined by white ratings. The other black evaluation scale was defined by scales such as *hip* versus *dumb*, *cool* versus *silly*, and *hip* versus *lousy*. The Potency factors for blacks and whites aligned, and so did the Activity factors. Landis and his colleagues selected scales to measure each of the three dimensions common to blacks and whites and computed factor scores for each of the nouns that had been rated. The results were presented in a table (Landis et al. 1976, Table 3).

Landis and colleagues compared their measurements of black and white sentiments with regard to selected concepts in the general areas of material possessions, confrontation, personal relations, quality of life, and ecosystem distrust. Their interpretative approach was one of cultural equivalence, legitimating middle-class family norms, and stressing similarities between blacks and whites.

Our feeling from the data presented in the previous pages is that not only do blacks value the same goals, relationships, and ideals that whites do, but in many cases they value them more. Where the differences occur, they seem to be related to perceptions of the amount of effort necessary to achieve those goals and the potency of those aims in changing one's life. In other words, blacks want the same things whites do, but they don't believe that (a) they can achieve them, (b) if they could, their lives would be significantly improved, and (c) they would be engaged in anything less than a constant struggle to maintain those things they do achieve. (Landis et al. 1976, p. 78)

Later Landis and Saral (1978, Table 6) factored mean ratings of the culturally common nouns on scales from 23 cultures simultaneously—a pancultural factor analysis. The intent was to see if the dimensions of black ratings corresponded to rating dimensions that had been found to be present in all of the cultures studied in the overall project (Osgood, May, and Miron 1975, Chapter 4). Black dimensions did indeed emerge that aligned with cross-culturally universal dimensions. Sets of black scales were selected to measure each of the pancultural dimensions, and these scales were then used to rate several hundred concepts that also were rated in all locales within the larger cross-cultural project. This part of the cross-cultural project was described in general terms by Osgood (1974, p. 1): An *Atlas of Affective Meanings* was to be constructed from mean ratings of 620 common concepts rated by respondents in all of the cultures in the study in order to provide a basis for exploring “subjective culture—values, feelings, and meanings—as it is expressed in language.” The 620 concepts sampled many areas, from universal notions prevailing in large societies, such as colors and numbers, to concepts arising in everyday life related to kinship, foods, animals, technologies, and other matters considered routinely by ethnographers.

The Landis and Saral (1978) report presenting black factor scores for 611 concepts was fallow for three decades. However, Sewell and Heise (2009) returned to the data and conducted cross-cultural analyses comparing the black sentiments, white American sentiments, and German sentiments. Sewell and Heise found that while black and white mean factor scores correlated positively on the three dimensions of Evaluation, Potency, and Activity—0.65, 0.38, and 0.03 respectively—these correlations were less than the correlations between white Americans and Germans—0.83, 0.66, and 0.58. That remained the case even when correlations were computed just with concepts that did not translate into different words in AAVE and when those correlations were corrected for unreliability. Summarizing, Sewell and Heise said: “We conclude that during the 1970s, sentiments in some Black groups were distinctive enough to be treated as a parallel subjective culture co-existing with the White subjective culture—as different from White culture as White culture was different from the culture of another nation.”

1.1.4 Observations on the Cases

The *DARE* study attempted to distinguish, delineate, and compare different subcultures of American folk English, whereas the other two studies focused on homogeneous populations in which respondents could be assumed more or less inculcated with uniform norms. One consequence was that the *DARE* study had to employ a method of factoring different groups apart, the ingenious use of population-weighted maps. Another consequence was the substantial amount of time required for analysis. The *DARE* study spanned 20 years from data collection to the first volume of lexicographic results, with the final volume expected 44 years after data collection. Part of the delay in the *DARE* study was due to its qualitative nature: Every definition had to be constructed uniquely. However, another consideration was the study’s cross-cultural aspect: Cross-cultural studies involve problems of synchronization and matching of responses for comparisons, beyond regular data analysis (see Harkness, van de Vijver, and Mohler 2002). The time required for analysis contrasts with the Boston kinship study, which was completed in a half-dozen years from data collection to book publication. The study of black sentiments culminated in a book chapter within a few years of data collection. On the other hand, the larger cross-cultural project of which the black sentiments study was a part did take years to produce results. A first volume of results was published about 10 years after data collection began, but a planned second volume devoted to cross-cultural comparisons was still in process when the project director (Charles Osgood) fell ill, and that volume never appeared.

Some uniformity is evident in all three examples of culture surveys. In all three cases, data were collected by visiting settings where the culture of interest was being reproduced day by day. The *DARE* investigators were explicit about this, designing their project so that it would sample

communities where different varieties of American English might be spoken, and giving instructions to interviewers to visit locales where folk conversations often take place. The study of African-American sentiments used urban black neighborhood high schools for its data-collection venues. Interactions among teenagers are settings where a unique black culture is likely to be found, as reflected in Labov's (1972, p. 257) remark that "the most consistent vernacular is spoken by those between the ages of 9 and 18." Although the study of family relations in Boston used a random probability sample of adults from the Boston Metropolitan Statistical Area, this was implemented by drawing an area probability sample of housing units in the Boston area. Thus the core sampling procedure focused on households where family culture was reproduced.

Often, seeking out settings where the culture of interest is being reproduced is associated with data collection in a single area, as in the Boston study of families, and the assessment of sentiments among Chicago blacks. However, when multiple areas are sought, culture researchers use population-weighted geographic sampling, as in the *DARE* study. [Similarly, Berk and Rossi (1977, p. 9) used numbers of prisoners as a weighting factor in selecting states for studying normative aspects of how elites were involved in prison reform).]

The *DARE* study and the study of black sentiments manifestly selected respondents for their levels of inculcation into cultures of interest. In the *DARE* study, respondents had to be natives of their area, preferably elders, with no diminishment of their cultural inculcation through extensive outside experiences. In the black sentiments study, urban black youths were employed exclusively as respondents and as language experts making study-related decisions. The focus on cultural expertise is less evident in the case of the study of family norms, but even there the respondents, all of whom were selected from Boston households, would have been participating in the reproduction of the family institution as they made daily decisions regarding their own family responsibilities. The sampling design omitted homeless and institutionalized persons, who would have been least likely to provide expert judgments regarding family obligations.

In all three of the case studies presented above, very long questionnaires were partitioned into relatively small subquestionnaires for administration to respondents. The long questionnaires reflected a basic consideration in cultural studies: A considerable range of material has to be covered to provide systematic treatment of a culture, as opposed to entertaining samples of cultural curiosities in the manner of newspaper feature articles. The need to subdivide the long questionnaires arises from the limited time and patience of respondents. The *DARE* study plumbed respondents' upper limits of these resources with interviews averaging 10 hours; that such long interviews were possible speaks to respondents' interest in their folk culture as well as to the adeptness of the *DARE* interviewers. The black sentiments study skirted respondents' minimum involvement, with some respondents providing ratings of only a single concept. The tediousness of repeated ratings on the same

scales, especially for adolescents, probably was the key factor in such limited participation, although limited literacy of some respondents might also have been a factor slowing their use of a paper-and-pencil questionnaire.

1.2 PREVIEW

The rest of this book deals with surveys of culture, but not equally with all types of culture surveys nor with all issues arising in such surveys. The focus is narrowed to some methodological issues in surveying cultural sentiments such as those assessed in the case study regarding American blacks and in the use of vignettes to survey unconscious but normative responses to situations such as those considered in the case study regarding kinship. A primary reason for narrowing the focus this way is that sentiment and vignette studies are the aspects of subjective culture that I myself have researched for decades and that I know best. Beyond that, sentiments (including attitudes) and normative processing of situations prior to decision making are among the most actively researched topics in the social sciences, so the book's focus is a justifiable hub of many research interests.

Surveys of sentiment norms can arise in studies of various kinds of social partitions: for example, race, gender, geography, nationality, ethnicity, education, academic discipline, occupation, and leisure pursuits. In the past, data-collection projects have assessed sentiment norms for several thousand concepts in multiple nations and languages (see Chapter 3), and additional surveys of this kind are in progress. One driving force behind these surveys is affect control theory (Heise 2007; MacKinnon 1994; Smith-Lovin and Heise 1988), which argues in essence that social interaction, institutional roles, emotions, and other social phenomena are a function of people maintaining cultural sentiments. The sentiment surveys, combined with affect control theory's mathematical model and computer simulation software, allow predictions to be generated about specific social processes in various cultures. Applications have been made to courtroom processes (Robinson, Smith-Lovin, and Tsoudis 1994; Tsoudis 2000a, 2000b; Tsoudis and Smith-Lovin 1998, 2001), business organizations (Schneider 2002; Schröder and Scholl 2009; Smith 1995), and international relations (Heise 2006; Heise and Lerner 2006). The surveys also produce culture databases that can be considered with databases for other cultures in comparative analyses (Heise 2001a; Ragin 1987).

The *DARE* study represents a class of studies devoted to recording culturally shared knowledge. I do not focus on this kind of work because it has been examined so thoroughly in psychological anthropology by the ground-breaking originators of the culture-as-consensus approach and their collaborators (Batchelder, Kumbasar, and Boyd 1997; Batchelder and Romney 1988; Romney 1994, 1999; Romney, Batchelder, and Weller 1987; Romney, Weller and Batchelder 1986; Romney and Weller 1984; Romney et al. 2000; Weller 1987; Weller, Romney, and Orr 1987). One of their publications

(Romney, Weller, and Batchelder 1986) is the most highly cited article ever to appear in *American Anthropologist* (Batchelder 2009). Their model has been applied in studies of folk medical beliefs, parental sanctions for rule breaking, judgment of personality traits, semiotic characterizations of alphabetic systems, occupational prestige, causes of death, strategies to control graffiti, national consciousness in Japan, and social network data.

Chapter 2 of the book reviews the development of sentiment measurement procedures. Key theoretical ideas regarding sentiments were in place at the beginning of the twentieth century. However, the development of empirical measurement procedures took another 50 years, with a mammoth cross-cultural study being a key component in the development. The last half of the twentieth century and the turn of the twenty-first century saw the incorporation of computers and the Internet into sentiment-measurement methodology.

Chapter 3 describes repositories of sentiments that can be used in cross-cultural studies. This line of research began in the 1950s and burgeoned from about 1970 onward. Repositories have been assembled in a score of nations, and some of these repositories provide quantitative measurements regarding thousands of sentiment norms. Chapter 3 also describes in detail the sentiment study that provided data used in analyses for Chapters 6 through 8.

Chapter 4 describes a vignette method for assessing norms of unconscious cultural processing. The first part of the chapter focuses on how vignettes have been used to explore the generation of preferences, as in the Boston study of kinship described earlier. Other studies of this kind are also considered in the review. The second part of the chapter describes how vignettes have been used in multiple societies to identify norms in forming impressions from events and from other kinds of observations. The chapter includes an appendix offering guidance for designing impression-formation studies.

Chapter 5 expands Romney, Weller, and Batchelder's (1986) culture-as-consensus model in psychological anthropology to culture surveys in general, dealing with sentiment measurements on quantitative scales instead of just data with dichotomous categories, which has been typical in culture-as-consensus research. A review of contemporary research on errors in surveys is followed by the formulation of a mathematical model that undergirds the subsequent three chapters.

Chapter 6 focuses on respondents' adequacy as informants about norms, and more generally on their levels of cultural inculcation. Performance measures obtained during the rating task are found to identify respondents whose contributions actually undermine the assessment of norms. An enculturation index is used to predict inculcation levels from background information, and it is found that respondents' social characteristics are associated with their levels of enculturation, although the correlations are only modest.

Chapter 7 examines the empirical tenability of assumptions involved in culture-as-consensus methodology. Factor analysis applied to respondents' evaluations of concepts reveal that ratings by different persons form a single

factor, supporting the assumption that respondents are normatively homogeneous. A dominant factor also characterizes ratings of a concept's potency levels and activity levels, but in both of these cases a second appreciable factor also appears, because some respondents transform a concept's evaluation into an assessment of the concept's potency and activity. Additional analyses demonstrate that factor analysis of sentiment data cannot be used to uncover subcultures, because adherents of a subculture have special sentiments for so few concepts relative to the total number of concepts in a culture.

Chapter 8 analyzes test–retest data for a few selected stimuli, plus one-time sentiment measurements for a large set of stimuli, in order to partition rating variances into cultural, individual, and error components. Converting these variances into measures of reliability reveals two key findings. First, ratings of a single concept in a normative study have fairly low reliability when reliability is conceived in a traditional manner, precisely because respondents' sentiments are normatively shaped and therefore similar to one another. Second, measurement reliability is substantially higher when reliability is conceived as the proportion of rating variance that is explained by the meanings of different concepts. Additional analyses show that the reliabilities of culture assessments increase dramatically when aggregating data from multiple respondents.

Chapter 9 contrasts culture surveys with traditional ethnographic studies and with traditional survey research studies. Despite some parallels, it is shown that culture surveys are a distinctive methodological approach that cannot be reduced to either ethnography or traditional survey research. An examination of sentiment data used in this book reveals that ratings of large samples of concepts are largely governed by meanings of the concepts rather than by idiosyncratic views of respondents, thereby emphasizing the extent to which traditional survey research studies focus on concepts that are in cultural play. The chapter also develops some guidelines for optimizing data quality in surveys of cultural sentiments.

1.3 CHAPTER HIGHLIGHTS

- In culture surveys, questions are asked about matters of agreement, so relatively few respondents need be surveyed to establish any particular norm. The more normative the respondents' beliefs and sentiments, the fewer of them are needed to obtain an accurate view of the culture.
- In culture surveys, a considerable range of norms has to be assessed systematically, so large samples of respondents may be required, with each respondent reporting on just a portion of the norms.
- In surveys of culture, lack of variability on every item is the ideal, since response variation confounds the delineation of norms. Surveys of cultures do seek variations in response, but across items rather than across respondents.

- The best respondents for a culture survey are denizens of settings where the culture is being regenerated. Thus, ideally, respondents are acquired by visiting settings where cultural reproduction takes place.
- Surveyable aspects of subjective culture include the categorization system that establishes culturally acknowledged realities, the implicit rules that interrelate cultural categories in ways that foster sensible decision making, and the sentiment system that orders cultural categories in terms of value, significance, and urgency.

