Chapter 3

Measurement Techniques

In Chapter 1 we provided conceptual definitions of belief, attitude, intention, and behavior. Attitude was defined as a person's location on a dimension of affect or evaluation. Belief was defined as his location on a probability dimension that links an object and an attribute. Intention was also defined as a dimension of probability, but the link here involves the person and some action with respect to the object. Finally, behavior was defined as a person's observable response when studied in its own right.

This chapter will discuss some of the techniques that have been used to assess beliefs, attitudes, and intentions. As noted in Chapter 1, an almost unlimited variety of measurement procedures have been employed in the attitude area. We will first try to show that most of these different procedures can readily be identified as measures of beliefs, attitudes, or intentions. There will then follow a discussion of the major attitude-scaling methods, namely, Guttman, Thurstone, Likert, and semantic differential scales. We will try to show that all of these standard scaling methods yield a single score that represents the person's location on an evaluative dimension, i.e., that they are all measures of attitude as here defined. Further, we will try to show that these attitude scores are always obtained by a consideration of beliefs or intentions and their associated evaluations. We will then consider some alternative attitude-measurement procedures, such as disguised and physiological measures. Finally, we will take up multidimensional scaling techniques and other measures of cognitive structure.

SINGLE-RESPONSE MEASURES

In the final analysis, all measurement involves observation of one or more responses made by a subject, whether they are verbal (e.g., a questionnaire response) or overt behavioral responses. Much research in the attitude area has

relied on single-response measures to infer beliefs, attitudes, and intentions. Although more complex indices of beliefs, attitudes, and intentions can be obtained, these indices are always in some way derived from single-response measures. It may therefore be instructive to examine the nature of single-response measures employed in the attitude area.

Most single-response measures are verbal in nature; the subject is asked to make a judgment about himself or about some other person, object, or event. Any response of this kind involves three different aspects: the concept, the judgment, and the format. That is, using a certain response format, the subject makes some judgment about a given concept.

Response formats. Many different response formats have been used. Most frequently, some form of graphic scale is employed and the subject responds by placing a check mark on the scale, which may be a continuous line or a line divided into several categories. Different response categories may also be presented in a multiple-choice format. Other response formats are qualitative judgments and quantitative estimates. For example, the subject may be asked to indicate whether a person is intelligent or not intelligent (qualitative judgment) or he may be asked to estimate the person's I.Q. (quantitative estimate).[1]

Irrespective of the response format used, every judgment places some concept into one of several categories. The categories defined by the response format may be discrete, nominal categories (qualitative judgment), they may be discrete categories ordered along some dimension, or they may represent points along a continuous dimension.

Much more than by the response format, however, diversity of judgment is introduced by variations in the concept being judged and the nature or content of the judgment that is required.

The concept. Judgments can be made with respect to any concept whatsoever. The concept may be a physical object, an institution, a person, a trait, an attribute, a behavior, etc. Such concepts can be described in detail or in a more general fashion. For example, one concept might be "the church" (general) whereas another concept might be "the Church of the Latter-Day Saints." Different concepts are also involved when "Joe Namath as a quarterback" and "Joe Namath as a playboy" are rated. Similarly, "highly informative" and "moderately informative" constitute different concepts.

The concept may not consist of a single object, person, or attribute but may instead be a statement linking an object to an attribute. Examples: "Joe Namath is a quarterback." "I intend to donate money to the church." "Reading the *New York Times* is highly informative." Again, the latter concept is not the same as "Reading the *New York Times* is *moderately* informative."

---

1. Subjects are sometimes asked to describe a concept using their own words. Typically, more than one response is elicited. When the investigator singles out one of these responses, he must somehow quantify or categorize it.

Nature of the judgment. Just as there is no limit to the number of concepts that can be rated, there is no limit to the number of labels that can be associated with the judgmental categories or dimensions. The content of a judgment is defined by these labels. Sometimes only the dimension's endpoints are labeled, but at other times each category on the dimension is given a label. For example, endpoints of a scale might be labeled *intelligent–not intelligent,* or the dimension might be divided into four categories with the labels *extremely intelligent, quite intelligent, slightly intelligent,* and *not at all intelligent.* Other examples of endpoints used in attitude research are *justifiable–not justifiable, approve–disapprove, I will attend church–I will not attend church, strongly agree–strongly disagree,* and *good–bad.*

Depending on the labels associated with its endpoints, a scale may be unipolar or bipolar. For example, the scale ranging from *hot* to *cold* is a bipolar or bidirectional scale whereas scales such as *not at all hot–hot* and *not at all cold–cold* are unipolar or unidirectional. Note that judgments on a unipolar scale are not the same as judgments on a bipolar scale, and different results may be obtained. For example, consider the bipolar scale

clean	+3	+2	+1	0	−1	−2	−3	dirty
	extremely	quite	slightly	neither or both	slightly	quite	extremely	

and the following two unipolar scales.

0	+1	+2	+3
not at all clean	slightly clean	quite clean	extremely clean

0	−1	−2	−3
not at all dirty	slightly dirty	quite dirty	extremely dirty

There is considerable evidence that knowledge of a person's responses to the two unipolar scales allows specification of his response to the bipolar scale (e.g., K. J. Kaplan, 1972). However, a given bipolar response may be associated with very different unipolar responses. For example, the following three pairs of unipolar responses would all be associated with a bipolar response of −1: (1) quite clean and extremely dirty; (2) slightly clean and quite dirty; (3) not at all clean and slightly dirty. If only the first unipolar belief were used, some subjects would be classified as judging the concept to be slightly clean whereas on the bipolar measure they would be classified as judging the concept as slightly dirty. However, note that unipolar and bipolar scales often tend to be highly intercorrelated, and this problem may be of greater theoretical than practical significance.

By the variation of concepts, formats, and judgmental content, an infinite number of single response measures can be generated. It is amazing how many different measures have indeed appeared in the literature. What is more problematic, most of these measures have at one time or another been viewed as measures of

"attitude." In the following sections we will try to show that most of the different measures can be classified as measures of attitudes, beliefs, or intentions.

Measures of attitude. Our definition of attitude requires a measurement procedure whereby a person assigns some concept to a position on a bipolar evaluative dimension. We noted earlier that many single-response measures ask the person to locate a concept along some bipolar dimension. The crucial question concerns the extent to which this dimension is evaluative in nature. Data concerning the evaluative nature of single bipolar scales are available (e.g., Osgood, Suci, & Tannenbaum, 1957). These data show that it cannot be assumed that such items will prove to measure affect under all circumstances. There is clear evidence that such scales may take on different meanings with respect to different concepts. For example, a scale ranging from *warm* to *cold* may have evaluative implications in reference to some concepts but not to others. If the concept rated is a person, a high correlation may be found between responses to the *warm–cold* scale and responses to a scale such as *good–bad*. That is, cold people are seen as being bad, and warm people as being good. On the other hand, in the judgment of an inanimate concept such as fire, the *warm–cold* scale may be used in a purely descriptive manner without any evaluative implications. Similar examples are scales such as *hard–soft, clean–dirty,* etc. A given scale may also be used differently by different subjects, even when the same concept is rated. Thus, in judging a person, one subject may use the *clean–dirty* scale in a descriptive manner, whereas another may use it evaluatively.

However, there are some scales that appear to be related to the evaluative dimension under most circumstances. Among these measures are a subject's self-rating of liking for the object (e.g., *like–dislike*), of his favorability with respect to the object (e.g., *favorable–unfavorable, approve–disapprove*), and his evaluation of the object on a single *good–bad* scale. Even these scales, however, may not always be perfect indicants of a person's attitude. Thus, in the absence of appropriate empirical evidence, single-response measures should not be taken as indicants of attitude, irrespective of the investigator's intuition regarding their affective nature. By the same token, unipolar scales, such as *not good–good* and other measures that involve only one pole of the evaluative dimension (e.g., rating of the concept "I like Joe Namath" on an *agree–disagree* scale), should not be viewed as measures of attitude. As noted earlier, unipolar scales may not yield the same results as bipolar scales, and thus they also require empirical evidence before they can be used as measures of attitude. Although many different scales may sometimes correlate with appropriate measures of attitude, they cannot be assumed to measure attitude under all conditions. The concept "attitude" should be employed only where there is clear evidence that the obtained measure places the concept on a bipolar affective dimension.

In sum, single-response measures often place a concept along some bipolar dimension. Whenever this bipolar dimension can be shown to be affective in nature, the judgment can be viewed as indicative of attitude. In all other instances,

the judgment merely assigns the concept to a given content category or to a position on a unipolar or bipolar content dimension. According to our conceptual definitions, however, when the response scale is a measure of subjective probability—i.e., when it represents a probability dimension—the strength of a belief or of an intention is being assessed. Which of these two variables is being measured depends on the concept involved in the judgment.

Measures of belief. We have seen that a belief associates an object and an attribute. To measure beliefs about an object, it is therefore first necessary to identify the attribute that is linked to the object. One can do so by asking a person to place the object into one of several content categories. The label associated with the category selected defines the appropriate attribute. For example, a person might be asked to rate Humphrey Bogart on a seven-place scale ranging from *wise* to *foolish*. It is possible to assign numbers to these positions ranging from 1 (foolish) to 7 (wise) or from −3 to +3. Unfortunately, the interpretation of these numbers is not always clear.

Assume that the categories were defined such that −3 (or 1) meant it was *extremely probable* that Humphrey Bogart was *foolish,* −2 (or 2) meant *quite probable* that he was *foolish,* +1 (or 5) meant *slightly probable* that he was *wise,* +2 (or 6) meant *quite probable* that he was *wise,* etc. Here the endpoints of the scale define two attributes (*wise* and *foolish*), and the response indicates the *strength of the subject's belief* that Humphrey Bogart is associated with one of these two attributes.[2]

Alternatively, suppose that the categories were defined such that −3 (or 1) meant "extremely foolish," −2 (or 2) meant "quite foolish," −1 (or 3) meant "slightly foolish," etc., and that the respondent judged Humphrey Bogart to be "quite foolish" (i.e., rated him −2 on the seven-place *wise–foolish* scale). This judgment identifies the attribute the respondent associates with Humphrey Bogart but does not provide any indication of subjective probability or belief strength.[3]

---

2. Obviously, this is a *relative* rather than an *absolute* measure of belief strength. A judgment on this type of bipolar scale does not provide an unambiguous measure of either the strength of the subject's belief that Humphrey Bogart is wise or the strength of the belief that he is foolish. Rather, the bipolar judgment that Humphrey Bogart is *quite probably foolish* is best viewed as a function of the strength of the subject's two unipolar beliefs. For example, this bipolar response might reflect his beliefs that it is *extremely probable* that Bogart was *foolish* and *slightly probable* that he was *wise.*

3. Note again that a judgment on a bipolar content dimension such as *wise–foolish* does not permit an unambiguous identification of the attribute category. The rating on this scale may be viewed as a function of the ratings that would be made on the two unipolar scales it comprises, namely, *wise–not wise* and *foolish–not foolish.* For example, the bipolar judgment *quite foolish* may be a compromise response based on the unipolar judgments *extremely foolish* and *slightly wise.* Thus, in order to unambiguously identify the attribute associated with a given object, we would recommend use of a unipolar response measure.

Once the object-attribute link has been established, it is possible to measure *belief strength*. For example, the concept "Humphrey Bogart is quite foolish" could now be rated on a probabilistic scale such as *probable–improbable, true–false, yes–no, agree–disagree,* or *likely–unlikely*.

The discussion above should make it clear that, depending on the instructions provided to a subject (i.e., the labels attached to the various response categories), a response on a scale defined by the same endpoints (e.g., *wise–foolish, not clean–clean*) may provide a measure of *belief content* (i.e., only identify the *attribute* associated with the concept), or it may also provide a measure of *belief strength* (i.e., of the subjective probability that the object has, or is associated with, a given attribute).

Many single-response measures found in the literature ask a subject to locate an object along some attribute dimension or to place it into one of several discrete attribute categories. These judgments constitute measures of *belief content*. In a more general sense, they may also be viewed as measures of belief strength since they assign a value of 0 or 1 to the probability that the object has the attribute in question. A *quantitative* measure of belief strength, however, is obtained only when the subjective probability associated with a given object-attribute link is assessed.

In conclusion, any judgment linking an object to an attribute category or to a position on an attribute dimension constitutes a measure of belief content. To obtain a measure of belief strength, the subjective probability associated with the object-attribute link has to be assessed. In some instances a relationship may be found between belief content and belief strength. For example, the more intelligence a person is perceived to possess (e.g., if he is seen as *quite intelligent* rather than *slightly intelligent*) the higher should be the subjective probability that he is intelligent.[4]

Our discussion of single-response belief measures implies that the placement of a concept on any dimension constitutes a measure of belief content or belief strength. Thus, even when the concept is placed on a bipolar dimension, such as *good–bad* or *like–dislike*, these judgments, strictly speaking, must be treated as measures of belief. Our discussion of single-response attitude measures, however, indicated that these particular bipolar measures of belief, under most circumstances, are highly related to the evaluation of the concept. In a similar manner, some other measures of belief strength may also be related to evaluation. For example, when an object-attribute link such as "the church is good" is rated on a probability dimension, belief strength may be related to attitude toward the church. Strictly speaking, however, this measure must be interpreted as a measure of belief strength that happens to be related to evaluation.

Measures of intention. From our point of view, behavioral intentions constitute a special case of beliefs, where the object is the person and the attribute is a behav-

4. This relation is likely to be monotonic but not necessarily linear.

ior. As with beliefs, a distinction must be made between an intention's content and its strength. The content of an intention is determined by the behavior that is to be performed. Behaviors are sometimes quantified along some dimension, such as donating \$10, \$20, or \$30 to the church or attending church never, sometimes, often, always. At other times, the subject is asked to indicate which of several behaviors he would perform in a given situation. The behavioral category selected defines the intention's content. The strength of the intention is assessed by asking the subject to rate a concept such as "I intend to donate \$20 to the church" on a probability dimension.

Most frequently, subjects are presented with a single behavior and are asked to indicate whether or not they would perform it. When a single qualitative judgment is made (*yes* or *no, would* or *would not*), the response again defines the intention's content, and its strength is assessed by a measure of probability. Thus, if a person indicated that he would donate money to the church, the strength of his intention could be measured by a probabilistic rating of the concept "I would donate money to the church" on scales such as *probable–improbable* or *agree–disagree.*

Instead of obtaining qualitative judgments, many investigators have assessed intentions on seven-place or nine-place scales defined by the endpoints *would–would not, willing–unwilling, intend–not intend,* or *will try–will not try.* Although, strictly speaking, these scales may merely assess content of the intention, it appears that judgments on them are probabilistic in nature and can be used to measure strength of the intention to perform the behavior in question.

STANDARD ATTITUDE SCALING

We have seen that it may be possible to obtain a direct measure of attitude by asking the person to rate an object on a single bipolar scale, such as *good–bad* or *I like the object–I dislike the object.* In many instances, however, attitudes are assessed by computing an index over responses to a set of belief items or a set of intentional items. In Chapters 1 and 2 we discussed the relations between attitude on the one hand and beliefs and intentions on the other. Specifically, we argued that a person's attitude toward an object is related to his beliefs that the object possesses certain attributes and his evaluations of those attributes. Similarly, attitude is related to the set of a person's behavioral intentions with respect to an object, each intention weighted by its evaluative implications. For beliefs this relation was expressed symbolically as follows:

$$A = \Sigma b_i e_i.$$

When b is replaced by I (for intentions), the same formulation holds for the relation between attitude and intentions.

This formulation has interesting implications for attitude measurement. In order to measure a person's attitude toward an object, one can obtain measures of the strength of his beliefs (b) that the object has certain attributes and measures

of his evaluation (e) of each attribute. The belief strength associated with a given object-attribute link is then multiplied by the person's evaluation of the attribute involved, and the resulting products are summed. This sum serves as an estimate of attitude toward the object under consideration. The same procedure would be followed for items of intention.

For example, if one is attempting to measure attitudes toward psychology, the first step involves identification of a set of attributes relevant for the subject population under investigation.[5] As a result of a pretest, the following statements might be constructed.

1. Psychology is not an exact science.
2. Psychology contributes to the solution of social problems.
3. Psychology is slightly interesting.
4. Psychology leads to control over man's mind.
5. Psychology is popular among students.
6. Psychology deals mainly with rats.
7. Psychology helps you to understand yourself.

Each of these statements associates the attitude object (psychology) with some other concept or attribute. In terms of our probabilistic definition of belief, an object and an attribute are perceived to be either associated to some degree or not to be associated at all; but there can be no negative association since probabilities cannot take on negative values.[6] Although some theorists (e.g., Heider, 1958; Osgood and Tannenbaum, 1955; Feather, 1971) have viewed object-concept links as either associative or dissociative, in the present conception all beliefs are of an associative nature. Thus, in the belief statements above, the attributes associated with psychology are as follows:

1. Not an exact science
2. Contributing to the solution of social problems
3. Slightly interesting
4. Leading to control over man's mind
5. Popular among students
6. Dealing mainly with rats
7. Helping you to understand yourself

Respondents are first asked to provide evaluations of these attributes (e). For example, each attribute could be rated on a seven-point *good–bad* scale or on a set of such evaluative scales. In keeping with the notion of a bipolar evaluative dimension, these ratings are scored from -3 (*bad*) to $+3$ (*good*). In order to obtain

5. Procedures for obtaining relevant attributes will be discussed in Chapter 6.

6. The same conclusion is derived from a stimulus-response definition of belief (Chapter 2).

measures of belief strength (b), each statement is rated on one or more probability scales. For example, a four-point scale ranging from *improbable* to *probable* could be used and could be scored from 0 to 3, allowing both beliefs and evaluation to carry equal weight in the prediction equation.[7] The $b \times e$ products can be computed for each item, and the attitude score is obtained by summing these products. (For a numerical example, see Table 2.1.)

The logic of this approach to attitude measurement underlies most standard attitude-scaling procedures. However, these standard procedures measure only the subject's belief strength (b) or the strength of his intentions. Attribute evaluations are not measured but are instead assumed to be the same for all subjects.

Beliefs as a Basis for Attitude Measurement

As in the procedure described above, standard attitude-scaling methods take an indirect approach by attempting to infer the person's location on the evaluative dimension on the basis of other responses. For example, Thurstone (1931) argued that "opinions" are verbal expressions of attitude and that they may therefore be used to measure attitude. Most attitude-scaling procedures arrive at an attitude score on the basis of a person's responses to a set of such opinion items. Specifically, these items are statements of belief or intention, and the person's response indicates his location along a probability dimension; i.e., it is a measure of the strength of his belief or intention. For example, the following items might be used to measure attitude toward Russia: (1) Russia is a totalitarian state, and (2) I would buy Russian products. The respondent usually is asked to indicate agreement or disagreement with each item. One major purpose of any scaling procedure is to select items that permit accurate inferences about the respondent's attitude.

To get a better understanding of how it is possible to use responses to items (i.e., the person's beliefs or intentions) to infer attitudes, it may be helpful to consider the balance model involving the focal person, p, another person, o, and an object, x, and the relations between these entities. Our present interest is in p's attitude toward o, and we will show how this attitude can be inferred from his beliefs about o.[8] Heider's (1958) theory assumed a tendency for triadic configura-

7. From a strict probability view, beliefs have to be measured on a scale that ranges from 0 to 1. However, our conceptual framework uses the notion of probability in a more general sense. For example, we do not assume that the beliefs about an object are mutually exclusive and exhaustive, and thus the probabilities are not expected to sum to 1. Although a probability scale ranging from 0 to 1 can be used, in the present example a four-point scale was used so that its range would be the same as that of the evaluative scale. Later in the chapter we shall see that use of a seven-point probability scale and bipolar scoring may be preferable in many practical applications.

8. A similar analysis would show how p's attitude toward o can be inferred from his intentions with respect to o. Here, only the p-o dyad needs to be considered; the intention indicates a positive (U) or negative ($\overline{\text{U}}$) unit relationship between p and o. In a balanced dyad, $p\text{U}o$ implies a favorable attitude ($p\text{L}o$), and $p\overline{\text{U}}o$ implies an unfavorable attitude ($p\overline{\text{L}}o$).

tions to be balanced. Assuming that a balanced state exists, knowledge of any two relations provides information about the third. For example, in a balanced configuration, $p\mathrm{L}x$ and $o\mathrm{U}x$ imply $p\mathrm{L}o$. Thus, p's attitude toward o can be inferred from knowledge concerning his belief about o (i.e., the perceived relation between o and x) and his evaluation of x. Although it is possible to assess both of these relations, the standard attitude-scaling methods assume a certain evaluation of x and then measure the o-x link. To make this assumption, it is necessary to eliminate items for which it is impossible to specify the evaluation of x. For example, consider the belief statement or item "The President is superstitious." If it is known that being superstitious is negatively evaluated by all respondents, it is possible to use this item in order to assess p's attitude toward the President. Specifically, p would be asked to indicate his agreement or disagreement with the item. If he agreed, a negative attitude toward the President would be inferred ($p\bar{\mathrm{L}}x$ and $o\mathrm{U}x$ imply $p\bar{\mathrm{L}}o$), but a positive attitude would be inferred if he disagreed with the item ($p\bar{\mathrm{L}}x$ and $o\bar{\mathrm{U}}x$ imply $p\mathrm{L}o$). However, if different respondents had different evaluations of being superstitious, the item would have to be eliminated by the standard scaling procedures. This follows from the fact that when the liking relation between p and x is unknown, no inference about the p-o relation can be made, even when knowledge of the o-x relation is available. Thus one purpose of a scaling procedure is to identify items that will have the same attitudinal meaning for all respondents.

The assumption of a balanced configuration also implies that knowledge concerning p's attitude toward o and x specifies p's belief about o, that is, his agreement or disagreement with the item. For instance, in the example above it is assumed that if p has a positive attitude toward the President and a negative evaluation of being superstitious, he will disagree with the statement "The President is superstitious." Indeed, the standard attitude-scaling procedures are based on the assumption that since the item has been selected such that p's evaluation of x is given, his agreement or disagreement with the item must be a function of his attitude toward o, and thus p's agreement or disagreement with the item can be used as an indication of p's attitude. An attempt is made to eliminate items that fail to meet this expectation. For example, respondents will tend to agree with the statement "Kennedy is a Democrat" regardless of their attitudes toward Kennedy or their evaluation of being a Democrat. In this case responses are clearly determined by factors other than the person's attitude, and the item will be eliminated by standard scaling procedures.[9]

Since elimination of items on this basis presupposes knowledge about the person's attitude toward the object (and the assumption that his evaluation of x is fixed), the question arises as to how it is possible to have knowledge of the person's attitude if the item itself is being used to measure the attitude. One obvious solution to this problem is to use samples of respondents whose attitudes are as-

9. As we shall see below, items eliminated on this basis may still serve as valid indicants of attitude (Fishbein, 1967d).

sumed to be known. For example, in the construction of a scale for the measurement of attitudes toward "labor unions," company executives (with assumed negative attitudes) and union officials (with assumed positive attitudes) could be used. Given a statement that links o (labor unions) to some positively evaluated x (job security), union officials should agree with the statement (pLo and pLx imply oUx), company executives should disagree ($p\overline{L}o$ and pLx imply $o\overline{U}x$). This technique of "known group comparisons," however, is usually employed to "validate" the total attitude scale rather than to select items. We will see shortly that most scaling methods contain a procedure for obtaining a preliminary estimate of the person's attitude toward the object, and they use this estimate in their item analyses.

One additional point needs to be made before we turn to a consideration of attitude scaling. In the discussion above, attitude was inferred from a single belief. Clearly, such a procedure will usually not provide a satisfactory measure of attitude. According to classical test theory, a person's response to a given item is composed of a "true" score reflecting the underlying dimension (in this case, his attitude) and some measurement error. It is usually assumed that as the number of items increases, measurement errors cancel each other out, and the sum or average across all items is thus a more accurate reflection of the "true" attitude. Indeed, all standard attitude scales use multiple-item formats.

Bogardus's Measure of Social Distance

In an attempt to measure social distance or prejudice toward members of various national, religious, and racial groups, Bogardus (1925) developed an instrument composed of seven intentional items. For each stimulus group (e.g., Armenians, Germans, Jews) his respondents were asked to indicate whether or not they would willingly admit members of this group as follows:

1. To close kinship by marriage
2. To my club as personal chums
3. To my street as neighbors
4. To employment in my occupation in my country
5. To citizenship in my country
6. As visitors only to my country
7. Would exclude from my country

Bogardus found that responses to these items tended to follow a certain pattern. A person who would exclude members of a given stimulus group from his country would also exclude them from all other settings. Similarly, if he agreed with Item 6, he would disagree with Items 1 through 5. Conversely, a person who would admit members of a given group to close kinship by marriage also tended to agree with Items 2 through 5 and to disagree with Items 6 and 7. In other words, the order of these seven items "seems to constitute (further experimenta-

tion is needed) a gradation in social . . . distance" (Bogardus, 1925, p. 303). Thus, although no item analysis was conducted, Bogardus assumed that the seven items implied increasing degrees of social distance.

A social-distance score is obtained by simply counting the number of settings from which members of a given group would be excluded.[10] This score can range from 0 to 6. That is, a score of 6 is obtained when a respondent agrees only with Item 7 and thus excludes the given group from the other 6 settings; a score of 0 is obtained when he agrees with Items 1 through 5, disagrees with Items 6 and 7, and thus does not exclude members of the group from any setting. The higher the score, the greater the degree of social distance, i.e., the more negative the attitude. This score may be viewed as a function of the respondent's intentions (i.e., his agreements or disagreements) and of the evaluations of the behaviors in question. From our point of view, an attitude score could have been obtained by assigning values of 1 (agreement) or 0 (disagreement) to the probability of an intention, and values of +1 (Items 1 through 5), 0 (Item 6), or −1 (Item 7) to the evaluation associated with each intention. Consistent with an expectancy-value model, the products of these values are summed to obtain a measure of attitude. The higher the score, the *less* social distance, or the *more* favorable the attitude. The Bogardus social distance measure can be obtained by subtracting 5 from this attitude score.

Note that assigning an evaluation of +1 or −1 to each item ignores the assumption that the different items express different *degrees* of social distance. This would not constitute a problem if all items followed the general cumulative pattern described above. If all responses followed the pattern, a score of 3 could be obtained only when a subject agreed with Items 4 and 5 and disagreed with all others. In comparison, consider a person who agrees with Items 2 and 3 and disagrees with all others. He, too, would have a score of 3. However, one may argue that this person exhibits less social distance since he would be willing to admit members of the group in question to settings that are assumed to imply less social distance. If the assumed differential evaluations associated with each item are to be neglected in computing the attitude score, it is necessary to demonstrate that responses to the set of items employed do indeed fall into a cumulative pattern. This is a major objective of the Guttman scaling procedure.

Guttman's Scalogram Analysis

In order to clarify the properties or characteristics of a cumulative scale, consider an example from the area of ability testing. A test designed to measure some ability (e.g., mathematical reasoning) is composed of a number of items varying in difficulty level; that is, the items differ in terms of the number of respondents who solve the problem posed in the item. Such a set of items forms a perfect cumula-

10. In practice, this can be done by reversing the response to Item 7 (i.e., treat agreements as disagreements and vice versa), ignoring the response to Item 6, and counting the number of disagree responses.

tive scale (i.e., a perfect Guttman scale) under the following conditions: A person solves all items up to a certain difficulty level and no items beyond that level. Thus, the more items a person passes, the higher his level of ability, and the most difficult item passed corresponds to his ability level. It follows that the number of items passed can be taken as an index of his ability. Response patterns of a perfect cumulative ability scale are shown in Table 3.1 (cf. Guttman, 1944).

Table 3.1 Response Patterns in a Perfect Cumulative Scale

Response patterns	Item difficulty						Ability score
	Low	1	2	3	4	5 High	
A		0	0	0	0	0	0
B		1	0	0	0	0	1
C		1	1	0	0	0	2
D		1	1	1	0	0	3
E		1	1	1	1	0	4
F		1	1	1	1	1	5

Note: A zero indicates failure, a one success with respect to a given item.

As Table 3.1 shows, such a cumulative scale has two interesting properties. First, knowledge of a person's ability score allows one to reproduce his performance on each item of the scale. A person with a score of 3 must have response pattern D, indicating that he passed the three easiest items and failed the two most difficult items. Second, of two respondents, A and B, if A has a higher score than B, then A has passed all items that B has passed as well as at least one additional (more difficult) item.

These properties imply that items on a perfect cumulative scale are ordered along a single dimension—in our example, along an ability dimension. Note that items are merely *ordered* in terms of their difficulty levels, and no assumptions are made about their exact locations on the dimension. Thus only an ordinal and not an interval scale is implied; there is no assumption that the distances between items are of equal magnitude. In the same vein, a person's ability score is also ordinal in nature.

As pointed out above, the properties of a cumulative scale provide information about the relationship between ability scores and performance on items on the scale. This relationship is expressed graphically in Fig. 3.1. We can see that the relationship between the person's location on the ability dimension and the probability that he will pass a given item can be described as a step function. This relationship is usually referred to as the item *operating characteristic* or *traceline*

Fig. 3.1 Tracelines of Guttman items with varying difficulty levels.

(B. F. Green, 1954). Figure 3.1 shows that both persons and items can be located on the ability dimension. A basic property of a perfect Guttman scale is that when the person's location on the dimension is lower than that of a given item, the probability that he will pass the item is 0; conversely, when his location on the dimension is as high as or higher than that of the item, the probability that he will pass the item is 1.

Note that the relation between total score and the responses to any given item is not linear; rather, it is step-shaped. High correlations between total score and single-item responses should not be expected since a correlation coefficient is an index of the degree of *linear* relationship.

Thus far we have discussed cumulative scales designed to assess an ability dimension. The same considerations apply to items designed to measure an evaluative dimension, i.e., attitude. The investigator begins by obtaining a number of belief or intentional items that are related to the attitude object under consideration. Bogardus's measure of social distance might comprise such a set of items. As pointed out earlier, a given set of items may not form a perfect cumulative scale; the Guttman scaling procedure is designed to assess the degree to which a given set of items form a cumulative scale.

As a first step, responses are obtained from a sample of subjects, and the items are ordered in terms of their frequencies of endorsement. This step is analogous to ordering items on an ability test in terms of their difficulty levels. Although it appears reasonable to assume that more difficult items (i.e., items with low probabilities of passing) require greater ability, it is not clear what is implied by probabilities of endorsement of belief or intentional items. The usual assumption is that such probabilities reflect the degree of favorableness (or unfavorableness) that is implied by the different items. Therefore the scale is assumed to measure an evaluative dimension, i.e., attitude toward the object.[11]

11. However, probabilities of endorsement conceivably could reflect an evaluative dimension with respect to an object other than the one considered by the investigator,

Once the items have been ordered, the obtained response patterns of each subject can be examined. Usually the subjects are also ordered in terms of their total scores. Thus, if the set of items formed a perfect Guttman scale, the pattern of responses would be triangular in form and would correspond to the pattern depicted in Table 3.1.[12] However, perfect Guttman scales are seldom obtained, and the working assumption is that deviations from the expected response patterns are due to measurement error. Clearly, however, such deviations may also indicate that the items under consideration are not unidimensional, i.e., that more than one dimension is necessary to account for the observed responses. The question then revolves around the amount of deviation that can be tolerated under the assumption of unidimensionality.

To this end, the "errors" in the response patterns are counted. That is, for each respondent whose pattern deviates from the expected, the minimum number of changes necessary to produce a scale pattern is considered the number of errors. The total number of errors in the sample of respondents is taken as an indication of deviation from unidimensionality. More precisely, Guttman suggested the coefficient of reproducibility as an index of the degree to which an empirical scale approaches a perfect Guttman scale. This coefficient, R, is presented in Eq. 3.1,

$$R = 1 - \frac{\text{Number of errors}}{\text{Total number of responses}}, \tag{3.1}$$

where the total number of responses is equal to the number of subjects multiplied by the number of items to which they respond. Guttman (1944) suggested that when $R \geqslant .85$, the scale may be considered acceptable.

This criterion, however, does not take into account the degree to which responses can be reproduced simply on the basis of relative frequencies of item endorsement. As A. L. Edwards (1957) pointed out, "the reproducibility of any single statement can never be less than the frequency present in the modal category" (p. 191). For example, if 70 percent of the respondents disagreed with a given item, one could predict disagreement for every respondent and be right 70 percent of the time. It is possible to compute a coefficient of minimal marginal reproducibility, MMR, as in Eq. 3.2:

$$MMR = \frac{\text{Number of responses in modal categories}}{\text{Total number of responses}}. \tag{3.2}$$

---

or they may reflect some other dimension altogether. For example, instead of reflecting degree of favorableness toward war, probabilities of endorsement may reflect favorableness toward the military-industrial complex. Alternatively, they may reflect social constraints, social desirability, or degrees of implied aggressiveness. Thus, one must take care in interpreting a Guttman scale as a measure of attitude toward some object.

12. The scoring for negative items will usually be reversed prior to the analysis.

The coefficient of reproducibility must be compared with this index of minimal marginal reproducibility; only when R greatly exceeds MMR (and $R \geqslant .85$) is there an advantage in using the Guttman scale.

When the coefficient of reproducibility is relatively low, one should conclude that the items in question do not form a unidimensional scale.[13] Frequently investigators attempt to raise the coefficient of reproducibility by eliminating items that contribute a large number of errors. Since this practice amounts to a redefinition of the attitude domain under investigation, it is not clear that the remaining set of items will measure the attitude originally intended by the investigator.

Like Bogardus's social-distance measure, the Guttman scale can be viewed as arriving at an attitude score through a consideration of beliefs or intentions and their associated evaluations. Unlike the scaling procedures to be discussed below, Guttman scaling is primarily concerned with testing the assumption that a set of items forms a unidimensional cumulative scale, rather than with selecting appropriate items from a larger pool. To some degree, selection is involved when items are eliminated to increase the coefficient of reproducibility. This procedure eliminates items that cannot be ordered along the evaluative dimension under consideration. That is, they are eliminated because they may have ambiguous attitudinal meaning or because they may be responded to on the basis of one or more irrelevant underlying dimensions.

The remaining standard attitude-scaling procedures also try to locate respondents on an evaluative or affective dimension, but unlike the Guttman scale, they do not result in a cumulative set of items. In Thurstone as in Guttman scaling, an attempt is made to order items along the evaluative dimension. Thurstone scaling, however, goes beyond ordinal measurement and attempts to locate the position of items on an equal-interval scale.

Thurstone's Equal-Appearing Interval Scale

The first step in Thurstone scaling involves the collection of a large pool of belief or intentional items related to some attitude object. As mentioned earlier, Thurstone (1931) assumed that responses to such items (i.e., the person's beliefs or intentions) are expressions of the person's attitude. More specifically, he made the assumption that different items may express different degrees of favorableness or unfavorableness toward the attitude object. A major purpose of Thurstone scaling is to specify the location of each item on the evaluative dimension by assigning a scale value to the item. On the basis of his work in psychophysical scaling, Thurstone first applied the method of paired comparisons to achieve this aim. Since this procedure becomes unwieldy as the number of items increases,[14] he developed the method of equal-appearing intervals as an approximation to the more sophisti-

13. It is usually also assumed that the universe from which the items are sampled is not unidimensional.

14. The number of judgments required for n items is equal to $n(n-1)/2$.

cated paired-comparisons technique. Although these two methods tended to yield similar scale values, he noticed some lack of correspondence for items with extreme scale values. The method of successive intervals was developed to adjust for this effect. By far the most popular and most widely used Thurstone scaling procedure is the method of equal-appearing intervals, and this method will be discussed in the present chapter. Interested readers are referred to A. L. Edwards (1957) for detailed discussion of the other techniques.

The pool of items collected by the investigator is given to a sample of judges representative of the population of subjects whose attitudes are to be assessed. Instead of being asked to agree or disagree with these items, the judges are required to indicate the amount of favorableness or unfavorableness toward the attitude object implied by agreement with a given item. More specifically, the judges sort each item into one of eleven categories that they are to consider equal intervals along the evaluative dimension, ranging from "unfavorable" through "neutral" to "favorable" toward the attitude object. For example, if a judge felt that the belief "The President is superstitious" indicated an extremely favorable attitude toward the President, he should place this item in the eleventh category; if he felt that the belief indicated an extremely unfavorable attitude, he should place it in the first category; and if he felt that it indicated neither a favorable nor an unfavorable attitude, he should place it in the neutral category. The remaining categories express equally spaced degrees of intermediate favorableness or unfavorableness.

The basis for item placement by judges can again be viewed in terms of a balanced triangle. The belief or intentional statement determines a given link between o and x. The statement "The President is superstitious" exemplifies an $o\mathrm{U}x$ link, and the statement "The President opposes medicare" exemplifies an $o\overline{\mathrm{U}}x$ link. The judge's perception of most people's (p's) evaluation of x determines the p-x link, thereby allowing the judge to make a direct inference about what p's attitude toward o would be if p agreed with the $o\mathrm{U}x$ link. For example, if a judge perceived that most people evaluate medicare favorably, he would assume that agreement with the statement "The President opposes medicare" implied a negative attitude toward the President ($p\mathrm{L}x$ and $o\overline{\mathrm{U}}x$ imply $p\overline{\mathrm{L}}x$). The more positive or negative the perceived evaluation of x, the more favorable or unfavorable is the implied attitude.

We can see that an item will express an unambiguous degree of favorableness only when judges agree as to how most people evaluate x. Thurstone proposed the *criterion of ambiguity* to eliminate items on which judges disagree. By this criterion, an item is eliminated if judges place the item into widely discrepant categories. More specifically, the interquartile range (i.e., the range that includes the middle 50 percent of categories into which the item was placed) or the standard deviation of the item placements is computed. These indices measure the degree of dispersion (i.e., disagreement among judges) in the placement of a given item. A large interquartile range or standard deviation leads to rejection of the item. The median or the mean is computed for each remaining item and this score is

taken as the scale value of the item. A set of approximately 20 items representing more or less equally spaced scale values ranging along the entire continuum are then identified.

Most investigators employ this set of items as their attitude scale and give it to a new sample of subjects whose attitude scores are to be measured. In contrast to the instructions given to the judges, the instruction to these subjects is to check all items with which they agree. Each subject's attitude score is obtained by computing the median or mean scale value of all items endorsed. Thus a subject who endorsed three items with scale values of 2.6, 3.0, and 3.4, respectively, would have an attitude score of 3.0.

According to the Thurstone scaling procedure, however, identification of a set of equally spaced and nonambiguous items is not sufficient as a basis for constructing an adequate attitude scale. Thurstone argued that a second criterion, the *criterion of irrelevance,* must also be met by each item on the scale. As mentioned earlier, scaling procedures should eliminate not only items that are ambiguous but also items that elicit responses which are determined by factors other than the attitude that is being assessed, i.e., by irrelevant factors. The criterion of irrelevance is designed to achieve the latter purpose in Thurstone scaling. The basic assumption underlying the criterion of irrelevance is that an item with a given scale value is most likely to be endorsed by respondents whose attitudes are located at the same position on the attitude dimension. The greater the discrepancy between the person's location on the evaluative dimension (i.e., his attitude score) and the item's location on that dimension (i.e., the item's scale value), the lower the probability that the person will agree with the item. This relationship between attitude score and probability of endorsement for an item with a given scale value is shown in Fig. 3.2.

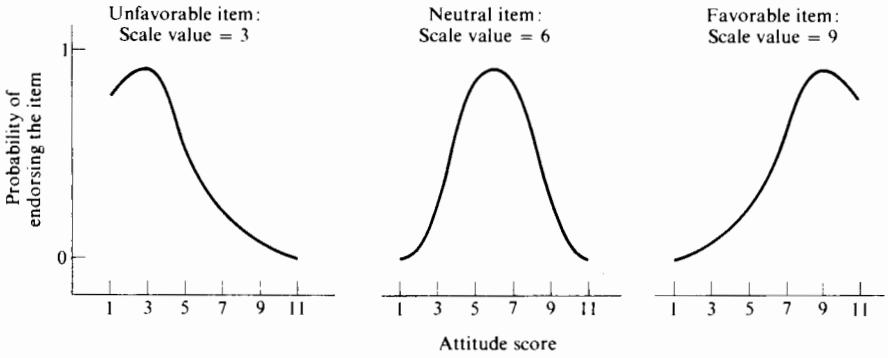

Fig. 3.2 Tracelines of Thurstone items with different scale values.

Note that the relation between the person's location on the evaluative dimension and the probability that he will endorse a given item, i.e., the item operating characteristic or traceline, takes the form of an inverted U. That is, the relation is

nonmonotonic with a single maximum. This kind of traceline indicates first that the items constituting a Thurstone scale are not cumulative. Thus, knowledge of a person's attitude score does not provide information about his responses to all items on the scale. For example, a person with an attitude score of 8 may endorse some items that have higher scale values, and he will usually not have endorsed all items with lower scale values. In contrast to a Guttman scale, the more the item's scale value falls below 8, the less likely he is to endorse it.

Further, even more so than for a Guttman scale, a high correlation between attitude scores and endorsement of items cannot be expected for a Thurstone scale since the correlation coefficient measures linear rather than curvilinear, nonmonotonic relations.

Inspection of Fig. 3.2 suggests that the criterion of irrelevance may be tested by plotting the subjects' responses to a given item against their attitudes. This, however, would require knowledge of the attitudes. A preliminary estimate can be obtained on the basis of the initially selected set of items, as described earlier. Thurstone and Chave (1929) developed an alternative procedure which does not require knowledge of attitudes but instead is based on the item's scale value and its probability of endorsement. Basically, the assumption is that items with similar scale values will exhibit similar patterns of endorsement.[15] Items failing to meet the criterion of irrelevance as assessed by either procedure are eliminated from the scale. The remaining items have passed both the criterion of ambiguity and the criterion of irrelevance, and they constitute the Thurstone scale. This set of items can now be used to assess attitudes in the manner described above. That is, an investigator administers the scale to a sample of subjects and computes their attitude scores by obtaining the median or mean scale value of all items endorsed.

One major criticism of the equal-appearing interval procedure has focused on Thurstone's assumption that the judges' own attitudes do not influence their judgments and thus have no influence on item scale values. Empirical evidence suggests that this assumption appears to be justified under most conditions. Later in this chapter, the issue will be considered in greater detail in the context of the "own categories" procedure. Partly in response to this criticism and partly to avoid the time and effort spent in obtaining item scale values, Likert (1932) suggested an alternative scaling procedure.

Likert's Method of Summated Ratings

Collecting a large pool of items is also the first step in constructing a Likert scale. As with Guttman and Thurstone scales, these items may be statements of either beliefs or intentions. For each item, the investigator first decides whether it indicates a favorable or unfavorable attitude toward the object in question. If the item is ambiguous or appears to indicate a neutral attitude, it is immediately

15. Interested readers are referred to Thurstone and Chave's (1929) discussion of the criterion of irrelevance.

eliminated. The investigator thus fulfills the function of a judge in Thurstone scaling, except that his task is simplified because items are merely placed into three categories: favorable, unfavorable, reject. The remaining items are administered directly to a sample of subjects representative of the target population. Typically, subjects are asked to respond to each item in terms of a five-point scale defined by the labels *agree strongly, agree, undecided, disagree,* and *disagree strongly.*

A preliminary estimate of each respondent's attitude is obtained as follows: First, responses to each item are scored from 1 to 5. Strong agreements with favorable items are given a score of 5, and strong disagreements with these items are given a score of 1. Scoring is reversed for unfavorable items, such that disagreement with an unfavorable item results in a high score. The person's preliminary attitude score is obtained by summing across all his item scores. For a set of 100 items, these attitude scores could range from 100 to 500; the higher the score, the more favorable the attitude.

Since constructing a Likert scale requires the elimination of items that do not reflect the attitude under consideration, an item analysis is performed. To be retained, an item must meet Likert's *criterion of internal consistency.* According to this criterion, the more favorable a person's attitude, the more likely he should be to endorse favorable items and the less likely he should be to endorse unfavorable items. This relationship between attitude score and probability of item endorsement is illustrated in Fig. 3.3.

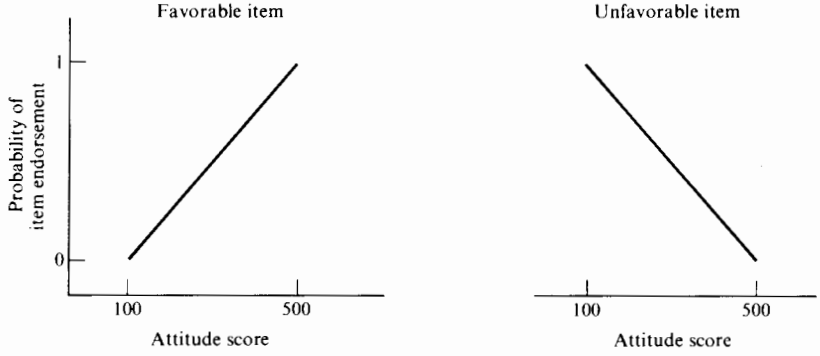

Fig. 3.3 Tracelines of favorable and unfavorable Likert items.

Note that in Likert scaling the item operating characteristic or traceline is assumed to be linear. Again, this traceline indicates that items on a Likert scale will not be cumulative. Here, however, a correlation between attitude score and item endorsement is not only expected but actually constitutes the criterion for inclusion of an item in the scale.

It follows that a given item meets the criterion of internal consistency if the item score correlates significantly with the attitude score. In practice, the pre-

liminary attitude score is used and is correlated with the item score.[16] Prior to the introduction of computers, a simpler procedure was employed whereby the preliminary attitude score was used to select the top and bottom 25 percent of the sample, i.e., subjects assumed to have the most and least favorable attitudes, respectively. Mean item scores are computed within each group. An item meets the criterion of internal consistency if it discriminates between the two extreme groups, i.e., if the favorable group has a significantly higher mean score on the item than the unfavorable group.[17]

The twenty or so items with the highest correlations (or the most discriminating items) constitute the Likert scale. This scale can now be given to a new sample of subjects whose attitude scores are computed in the manner described earlier with reference to the preliminary attitude score. The Likert scaling procedure, then, ensures that ambiguous items as well as items that elicit responses based on factors other than the attitude under consideration are eliminated.

All scaling procedures discussed thus far place individuals along a single dimension of affect by considering their responses to a set of items assumed to reflect this underlying dimension. The derived attitude score represents the person's location on the evaluative dimension. Throughout our discussion, we have emphasized that items reflecting the respondent's location on other dimensions are eliminated by the scaling procedures. The fourth standard scaling procedure to be discussed, the semantic differential technique, permits the identification of items reflecting not only the evaluative dimension but also other dimensions.

Osgood's Semantic Differential Technique

Osgood's measurement technique was developed not for purposes of assessing attitudes but rather as an instrument for the measurement of meaning. The reader will recall that in Osgood's behavior theory the implicit anticipatory response to a stimulus object is viewed as the object's meaning. Since this implicit response cannot be directly observed, overt responses to the object have to be considered. In his search for overt responses that are "maximally dependent upon and sensitive to meaningful states, and minimally dependent upon other variables" (Osgood, Suci, and Tannenbaum, 1957, p. 11), Osgood (1952) settled on verbal responses to the object or concept. He argued that since the basic function of ordinary language was assumed to be the communication of meaning, ordinary language could be used to differentiate between concepts and measure their meaning. The semantic differential technique is based on this premise.

Osgood, Suci, and Tannenbaum argued that as a first step it was necessary to devise a "sample of alternative verbal responses which can be standardized across

16. A negative correlation indicates that the investigator's initial decision concerning the item's favorableness was in error, and the item score is reversed.

17. Alternatively, groups assumed to have positive or negative attitudes can be compared in terms of their responses to each item. This is the "known group comparison" technique mentioned earlier.

subjects . . . and [which would] be representative of the major ways in which meanings vary" (p. 19). Ascertaining an object's meaning was viewed as similar to playing the game of "Twenty Questions" with the respondent. Thus, to identify the meaning of a given object, the respondents might be asked questions such as: "Is it hard or soft?" "Is it pleasant or unpleasant?" "Is it fast or slow?" etc. "Just as in 'Twenty Questions' the selection of successive alternatives gradually eliminates uncertainty as to the object being thought about, so selection among successive pairs of common verbal opposites should gradually isolate the 'meaning' of the stimulus" (Osgood, Suci, and Tannenbaum, 1957, pp. 19–20). To increase the measuring instrument's sensitivity, a seven-point scale is inserted between the bipolar adjective pairs so that the subject can indicate both the direction and intensity of each judgment.

A large number of such bipolar adjective scales were constructed in an attempt to obtain a representative sample of the possible dimensions along which concepts can be judged. The semantic differential technique involves providing the respondent with one or more concepts to differentiate and a set of bipolar adjectives against which to do so. The respondent's task is to rate each concept on each scale. In this manner, a profile of ratings is obtained for each concept; it is assumed that two concepts are similar in meaning to the extent to which their profiles are similar. Hypothetical profiles for three concepts are illustrated in Fig. 3.4. The degree to which two profiles are similar, i.e., the degree to which two

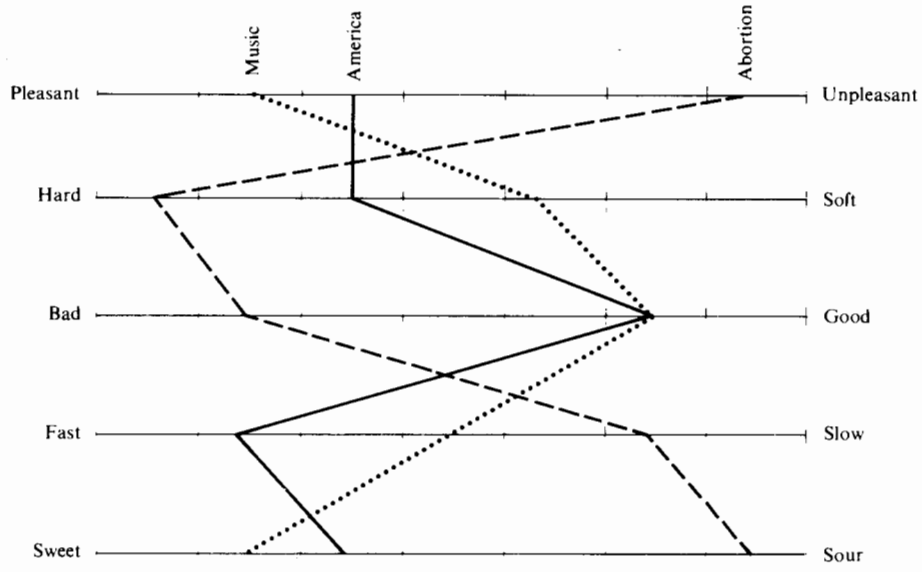

Fig. 3.4 Hypothetical profiles for three concepts rated on five bipolar adjective scales.

concepts have similar meanings, is measured by the generalized distance formula presented in Eq. 3.3,

$$D = \sum_{i=1}^{n} d_i^2, \tag{3.3}$$

where D is the distance between two concepts (i.e., their dissimilarity) and d_i is the difference in ratings of the two concepts on the ith bipolar adjective scale. The squared differences are summed across the n scales on which the two concepts are rated.

Although the generalized distance formula provides an index of the degree to which two concepts are similar in meaning, it does not identify the meaning of any given concept. For such an identification to be made, the concept's position on each of the bipolar scales would have to be described. Not only is this highly impractical when the concept is rated on a large number of bipolar scales, but it should be possible to describe a concept's meaning in terms of a smaller number of basic dimensions since different scales may be reflecting the same aspect of meaning. Indeed, the major aim in developing the semantic differential was to identify the major dimensions of meaning. To do this, Osgood, Suci, and Tannenbaum applied the statistical procedure known as factor analysis.

The general idea behind factor analysis can be illustrated by considering responses to two bipolar adjective scales. When these responses are highly correlated, the scales may be said to be alternative measures of the same variable or dimension. To generalize this principle to a large set of scales: If all scales are highly intercorrelated, they are all considered to assess the same underlying dimension or factor. In the context of the semantic differential, they are all considered to measure the same basic aspect of meaning. The purpose of factor analysis is to identify the number of dimensions that are being assessed by a large set of scales. More precisely, it specifies the degree to which each scale is a measure of each of the underlying dimensions. The basic data matrix on which a factor analysis is usually performed consists of the set of correlations between all possible pairs of scales.[18] The resulting factor matrix can be conceptualized as representing the correlations between the scales and the underlying factors or dimensions. Bipolar adjective scales which are themselves highly intercorrelated should load on (i.e., correlate with) the same factor or factors. Factor analysis is one convenient method for summarizing the interrelationship among a large number of scales since, ideally, the number of factors should be considerably less than the number of scales (N. Wiggins, 1973).

In a large number of studies involving different scales and different concepts, Osgood and his associates have repeatedly found three basic factors or dimensions underlying semantic differential ratings. Based on inspection of the scales that had high loadings on (i.e., high correlations with) these factors, the three major di-

18. Note that factor analysis is a general procedure that can be applied to any intercorrelation matrix.

mensions were interpreted as *evaluation, potency,* and *activity.* For purposes of illustration, Table 3.2 shows some of the results of a factor analysis presented by Osgood, Suci, and Tannenbaum (1957, p. 39). For this analysis, 100 subjects rated 20 concepts on 50 bipolar adjective scales. Correlations were computed between all possible pairs of scales. Since each subject responded to each scale 20 times (once for each concept), each correlation was based on 2000 pairs of observations. The concepts were selected from different domains and included, for example, "lady," "symphony," "lake," and "America." Table 3.2 presents the obtained loadings of nine selected scales on the three major factors. As we can see, the evaluative factor is characterized by scales such as *good–bad, clean–dirty,* and *beautiful–ugly;* the potency factor is characterized by scales such as *large–small, strong–weak,* and *thick–thin;* and some of the scales loading on the activity factor are *fast–slow, active–passive,* and *hot–cold.*

Table 3.2 Selected Factor Loadings of Scales on the Three Major Dimensions (from Osgood, Suci, and Tannenbaum, 1957, p. 37)

	Factor		
Scale	Evaluation	Potency	Activity
good–bad	.88	.05	.09
clean–dirty	.82	−.05	.03
beautiful–ugly	.86	.09	.01
large–small	.06	.62	.34
strong–weak	.19	.62	.20
thick–thin	−.06	.44	−.06
fast–slow	.01	.00	.70
active–passive	.14	.04	.59
hot–cold	−.04	−.06	.46

Osgood argued that a person's attitude toward an object is equivalent to the object's evaluative meaning for the person. Consistent with the view expressed in this book, Osgood's definition of attitude was in terms of a bipolar evaluative dimension. He thus concluded that one could measure a person's attitude toward any object by having him rate that object on a set of scales known to have high loadings on the evaluative factor. Once a set of evaluative scales has been identified, it can be used to measure attitudes toward a large number of concepts. This apparent generality of the semantic differential has resulted in its being used in a variety of contexts, and indeed, it is probably today's most widely used attitude-measuring instrument. Typically, an investigator will select five to ten adjective scales that were found to have high loadings on the evaluative factor by Osgood, Suci, and Tannenbaum or by some other investigator. Sometimes activity and potency scales are also employed. Subjects are then asked to rate one or more

concepts (i.e., attitude objects) on each scale. A measure of attitude toward each object is obtained in the following manner: Responses to each evaluative scale are scored from +3 to −3, with positive values assigned to the positive side of the scale. The scores are summed across all evaluative scales, and this value or the average score is taken as an index of attitude. Similarly, activity scores and potency scores can be calculated.

It is seldom realized, unfortunately, that this practice may not meet some of the requirements of the semantic differential technique. Investigators have found that the same bipolar adjective scale may load on different factors when it is used to measure different concepts (cf. Osgood, Suci, and Tannenbaum, 1957). For example, in studies of person perception, the *hard–soft* scale has often been found to correlate with the evaluative dimension. The same scale may not correlate highly with the evaluative dimension when the concept "war" is rated; instead, it will tend to load on the potency factor. And although the *hard–soft* scale tends to correlate with the evaluative dimension for person concepts, *soft* will be the positive end of the scale for feminine concepts (e.g., "mother") but the negative end for masculine concepts (e.g., "father"). Further, different factor structures tend to emerge when the analysis is based on the ratings of a single concept than when it is based on ratings of that concept and other concepts on the same set of scales. In the present example, the *hard–soft* scale correlates with the evaluative dimension when the ratings of either "father" or "mother" are analyzed. However, in an analysis based on the ratings of both concepts, the *hard–soft* scale would not correlate with the evaluative dimension, since that correlation takes on different signs for the concepts "mother" and "father." When both concepts are considered, therefore, the correlation of the *hard–soft* scale with the evaluative dimension will be close to 0. Thus, in this example, the potency dimension (as represented by the *hard–soft* scale) merges with the evaluative dimension when a single concept is considered, but the two dimensions emerge as separate factors when both concepts are included in the analysis.

This tendency for scales to load on different factors when different concepts are rated has been called "concept-scale interaction" (Osgood, Suci, and Tannenbaum, 1957). It follows from these considerations that scales found to load on a given factor in a multiconcept analysis cannot be assumed to load on the same factor for any given concept, even when that concept is part of the original set. These phenomena, therefore, preclude the notion of a generalized attitude-measuring instrument. Indeed, Osgood and his associates were unable to find sets of scales that would consistently load on a given factor across a wide variety of concepts. Thus, if the semantic differential is to be used as a measure of attitude, it (like all other standardized instruments) must be submitted to an item analysis. Since most attitude measurement is concerned with assessing attitudes toward a single concept, it would be desirable to identify a set of adjective scales that best reflect the evaluative dimension for that concept.

Use of factor analysis in the development of the semantic differential assumes a linear (or monotonic) traceline; that is to say, responses to evaluative scales

correlate with the subject's attitude score, i.e., with the evaluative dimension. Thus the item operating characteristic of items on the semantic differential is the same as that for items on a Likert scale (see Fig. 3.3). This suggests that Likert's criterion of internal consistency could be used to analyze items on a semantic differential. To do so, one would consider only scales that appear to be evaluative in nature. If the investigator could not decide which side of the bipolar scale was positive and which negative, the scale would be immediately eliminated. Separate scale-total score correlations would be computed, and only scales with high correlations would be retained.

More in line with Osgood's approach, subjects would rate the concept in question on a heterogeneous set of scales, and their ratings would be submitted to a factor analysis. Scales with high loadings on the evaluative factor would be retained. This set of items constitutes a semantic differential scale for the measurement of attitude toward the concept in question.

Bipolar adjective scales, such as those constituting the semantic differential, have previously been described as measures of either belief content or belief strength. It is important to note that Osgood's original instructions for using the semantic differential defined the bipolar scales as measures of belief *strength*. More specifically, Osgood asked subjects to indicate the degree to which "the concept is . . . *related* to one or the other end of the scale" (Osgood, Suci, and Tannenbaum, 1957, p. 83; italics added). That is, Osgood asked subjects to indicate the strength of the relationship between the concept and one of the two bipolar adjectives. Given these instructions, the subject indicates by his response the likelihood that the concept is associated with one or the other attribute in the pair. That is, his response can be viewed as a probability judgment, and it constitutes a measure of belief *strength*.[19] Also recall, however, that the *good–bad* scale was viewed as a possible direct measure of attitude. We can now see that when Osgood's original instructions are given to subjects, a *good–bad* scale, like any other bipolar adjective scale, constitutes a measure of belief strength. Consider for example ratings of "war" on a *safe–dangerous* and on the *good–bad* scale. Each scale links the object "war" to an adjective pair. The subject's response indicates the degree to which he perceives the object to be associated with one or the other attribute in the pair, and his response can be viewed as a probability judgment. Unlike most other bipolar adjective scales, however, the *good–bad* scale is found to have a consistently high loading on the evaluative factor, irrespective of the concept under consideration. This particular belief, therefore, almost always tends to be

19. Note again that when bipolar scales are used, *relative* rather than *absolute* belief strength is being measured. That is, one is assessing the probability that the concept is attribute A (e.g., fair) relative to attribute B (unfair) rather than the probability that the concept is, or is related to, attribute A independent of any other attribute. As we have seen, however, knowledge of the two absolute judgments should permit prediction of the relative judgment.

highly correlated with attitude, and thus, as suggested earlier in this chapter, it may be used as a direct measure of attitude.[20]

A Comparative Analysis of Standard Scaling Procedures

In this chapter we have discussed four standard attitude-scaling procedures, all of which arrive at single attitude scores based on responses to statements of beliefs or intentions. Of course, these scaling methods can also be based on behavioral observations, where performance or nonperformance of the behavior is treated as equivalent to agreement or disagreement with a belief or intentional statement. This notion plays an important role in Chapter 8, which deals with the relation between attitude and behavior.

Irrespective of scaling technique or item content, the resultant attitude score represents an individual's location on a bipolar evaluative dimension vis-à-vis a given object. Beyond this fundamental similarity, several important differences between the standard scaling procedures must be recognized. Table 3.3 indicates

Table 3.3 A Comparison of the Four Standard Attitude Scales

	Guttman	Thurstone	Likert	Semantic Differential
Properties of Items				
Quantification	ordinal	"interval"	qualitative	—
Neutral items retained	yes	yes	no	no
Tracelines	monotonic (step-shaped)	nonmonotonic (inverted U)	monotonic (linear)	monotonic (linear)
Cumulative scale	yes	no	no	no
Item selection	response-inferred	response-inferred and judgmental	response-inferred	response-inferred
Properties of Attitude Scores				
Values of b	$0, \pm 1$	$0, 1$	-2 to $+2$	0 to 3
Values of e	± 1	-5 to $+5$	$-1, +1$	$-1, +1$
Values of $b \times e$	$0, 1$	-5 to $+5$	-2 to $+2$	-3 to $+3$
Disbeliefs	yes	no	yes	no
Computational formulas	$\Sigma\, b_i e_i$	$\Sigma\, b_i e_i / \Sigma\, b_i$	$\Sigma\, b_i e_i$	$\Sigma\, b_i e_i$ or $\Sigma\, b_i e_i / n$
Quantification	ordinal	"interval"	ordinal	"interval"

20. As noted previously, the same may be true for certain other single bipolar scales, such as *I like–dislike* or *favorable–unfavorable*.

some of the differences between scales in terms of item characteristics and properties of attitude scores.

Properties of items. All scaling procedures locate individuals along an evaluative dimension (by assigning attitude scores), but Guttman and Thurstone scales also locate each item on the same evaluative dimension. By using judges, the Thurstone procedure assigns a scale value to each item, which represents the item's location on the evaluative dimension. It is assumed that these judgments are made on an equal-interval scale, and thus item scale values as well as attitude scores are assumed to have equal-interval properties. In contrast, Guttman scaling results in an ordering of items along the evaluative dimension without the assumption of equal intervals. Consequently, Guttman scores represent an ordinal scale.

Items on a Likert scale do not reflect different *degrees* of favorableness; instead, they are classified as either positive or negative. This categorical assignment of items also results in attitude scores with ordinal properties. Finally, since semantic differential items are bipolar, they can be assigned to neither a positive nor a negative category and are merely assumed to correlate with the evaluative dimension. The seven places on the bipolar scales are assumed to represent equal intervals, and thus semantic differential attitude scores, like those obtained from a Thurstone scale, are assumed to have equal-interval properties.

One interesting implication of these differences in item quantification concerns the utility of neutral statements in the assessment of attitude. In Thurstone scaling, since neutral items are viewed as valid indicants of attitude, they are part of the attitude scale. Neutral items may be included in a Guttman scale, but there is no way of distinguishing them from other items because Guttman scaling examines the degree to which a set of items can be ordered along a single dimension. Thus, it is possible for the items under consideration to be all favorable, to be all unfavorable, or to range from favorable through neutral to unfavorable.

In contrast, Likert scaling explicitly excludes items that cannot be classified as positive or negative, and the implicit assumption is therefore that neutral items are not valid indicants of attitude. Finally, the issue of item neutrality applies to the semantic differential in that it eliminates bipolar adjective scales (i.e., items) that are not clearly evaluative in nature. At the same time, however, a neutral response on a bipolar evaluative scale is permissible.

All scaling procedures are designed to identify items that are unambiguous indicants of a person's attitude. The different criteria employed result in the selection of items with different operating characteristics or tracelines. Thus Guttman, Likert, and semantic differential items have monotonic tracelines, whereas items on the Thurstone scale have nonmonotonic tracelines. A monotonic traceline indicates that as the attitude score increases, the probability of item endorsement increases or remains constant but never decreases. The Likert and semantic differential tracelines are assumed to approach linearity, but in the Guttman scale it has a step-shaped form. In Thurstone scaling, the probability of item endorsement is assumed to increase as the person's attitude score approaches the item's scale

value (from either end of the bipolar dimension), resulting in an inverted U-shaped traceline.

A related difference between the scales is that items on the Guttman scale are cumulative whereas items on the other scales are not. Thus each attitude score obtained in Guttman scaling is associated with one—and only one—pattern of responses to the items on the scale. In contrast, many different response patterns may result in the same attitude score when any of the other techniques is used. Thus item responses can be predicted from the attitude score only with the Guttman scale.

Before we turn to a more detailed analysis of attitude scores, one other comment concerning item selection procedures is in order. Since only the Thurstone scale makes use of an independent sample of judges to obtain item scale values, it has often been called a "judgmental scale." In contrast, the other scales have been termed "response-inferred" since responses to the items (i.e., agreements or disagreements) are basic to their item-selection procedures. Recall, however, that in Thurstone scaling it is also necessary to obtain responses from subjects in order to test for the criterion of irrelevance. Thus, when properly constructed, the Thurstone scale is at least in part response-inferred.

Properties of attitude scores. Most remaining differences between the four attitude scaling methods are related to the computation of attitude scores. Generally speaking, in all procedures subjects agree or disagree with items that indicate some degree of favorableness or unfavorableness toward the attitude object. A score is obtained for each item by multiplying degree of agreement and item favorableness. It can be shown that this item score is equal to the product of belief about the object (b) and the evaluation of the object's attributes (e), discussed earlier.[21]

In all four standard scaling procedures, attitude scores are computed by taking the sum or average over all item scores or $b \times e$ products. Thus, as Table 3.3 shows, attitude scores are always obtained in accordance with an expectancy-value formulation. The difference between scaling procedures concerns the relative weights placed on beliefs and evaluations of associated attributes. More specifically, Table 3.3 shows that the semantic differential assigns a value of $+1$ to the positive adjective and a value of -1 to the negative adjective on a given scale, and the subject's response can be viewed as indicating the strength of his

21. More specifically, a person agrees or disagrees (b') with a favorable or unfavorable statement (e') which associates or dissociates (u) the attitude object and some attribute (in the case of a belief statement) or the person's behavior (in the case of an intentional statement). The following two equations show that b and e can be derived from b', e', and u.

$$b = b'u, \quad \text{and} \quad e = e'/u.$$

Despite the differences between b and b' on the one hand and e and e' on the other, the next equation shows that the $b \times e$ product is equivalent to the $b' \times e'$ product.

$$be = b'u\,(e'/u) = b'e'.$$

belief that the concept is associated with one or the other adjective; belief strength can vary from 0 to 3. In Likert scaling, each item is assumed to indicate either a favorable (+1) or unfavorable (−1) attitude, and responses are given on a five-point scale ranging from strong agreement (+2) to strong disagreement (−2). Thus belief strength varies from −2 to +2, and evaluations of associated attributes are either +1 or −1.

Both the semantic differential and Likert scales place greater weight on b than on e in computing attitude scores, but the opposite is true in Thurstone scaling. Here a subject either agrees (+1) or disagrees (0) with each item, and the item's favorableness can range from −5 to +5. Finally, a respondent either agrees or disagrees with items on a Guttman scale. For some items agreement is scored +1 and disagreement is scored 0; for other items scoring may be reversed in order to produce a cumulative response pattern. Although items may reflect different degrees of favorableness, they are all assigned a score of +1 in computing the attitude score. Thus, as in Likert scaling, the assumption is that disagreement (−1) with a negative item (−1) is equivalent to agreement (+1) with a positive item (+1). In terms of the $b \times e$ formulation, then, b may take on values of 0, +1, or −1, and e may be either +1 or −1.

These considerations suggest that the standard attitude-scaling procedures differ in the way they treat nonagreements and disagreements with belief or intentional items in the computation of attitude scores. In Likert scaling, disagreement with an item is assumed to be indicative of the person's attitude. Thus disagreement with a negative statement is taken as an indication of a positive attitude. By way of contrast, in Thurstone scaling, nonendorsement of an item is viewed to have no implications for attitude; only the scale values of endorsed items (i.e., items the subject agrees with) are considered in computing the attitude score. Nonendorsement may or may not contribute to attitude scores in Guttman scaling, depending on whether or not the item has been reversed. For items with reversed scoring, nonagreement with a statement is treated as equivalent to agreement with its opposite, and thus it contributes to the attitude score. With respect to the semantic differential, a "neutral" response is treated as nonagreement with either member of the adjective pair, and thus nonagreements enter the computation of attitude scores.

In sum, disagreements or nonagreements influence attitude scores in Guttman, Likert, and semantic differential scales, but not in the Thurstone scale. Further, Table 3.3 shows that only the Guttman and Likert scales can assign negative values to beliefs, thereby implying that a person may hold disbeliefs which contribute to his attitude. That is, in these two scales the respondent may actively disagree with a statement. In contrast, the semantic differential does not assign negative values to belief, and although neutral responses contribute to the attitude score, they indicate lack of belief rather than disbelief.

Unipolar versus bipolar belief measures. Recall that in our conceptual framework, belief strength was defined as a person's location on a probability dimension linking an attitude object and some other concept. Although it is easy to conceptualize

lack of belief as a zero probability, the question of disbeliefs becomes problematic since probabilities cannot take on negative values. That is, the object is always linked to the concept with a probability that ranges from 0 to 1; a disbelief therefore has to be conceptualized as the probability that the object is linked to the negation of the concept. Thus, according to a probabilistic view of beliefs,

$$p(\bar{b}) = 1 - p(b), \tag{3.4}$$

where b is a given belief and $\bar{b}$ is the negation of b. This notion can be illustrated in a general way by representing a belief as $[(o)$ is $(x)]$ and its negation as $[(o)$ is $(\bar{x})]$.

For example, consider a person's disagreement with the belief statement "Chairman Mao is dead." Equation 3.4 suggests that this disagreement can be viewed as equivalent to an agreement with the statement "Chairman Mao is not dead." As a numerical example, imagine that the following measure of belief was obtained.

Communist China is aggressive

disagree		X						agree
	−3	−2	−1	0	+1	+2	+3	

Instead of treating this response as a disbelief, Eq. 3.4 suggests that within a probability formulation it should be viewed as an indication of the belief that Communist China is not aggressive, and the strength of this belief would be +2.[22]

This practice, however, may prove problematic within the context of an expectancy-value formulation. Clearly, the evaluation of an attribute that is related to the attitude object and the evaluation of its negation will not be the same. For example, assume that the concept "aggressive" is evaluated −3. In conjunction with the disbelief of −2 noted above, an expectancy-value formulation assigns an item score of $(-2)(-3) = 6$. However, this computation is not permissible in the framework of a probabilistic conception of beliefs, where beliefs cannot take on negative values. As shown earlier, within this framework the −2 disbelief would be treated as a +2 belief that "Communist China is not aggressive." To compute the $b \times e$ product, one would have to know the evaluation of "not aggressive," and one cannot assume that it is +3 (i.e., that it is merely the polar opposite of "aggressive").

The implications of these considerations for attitude measurement are as follows: All attitude scales derive a person's attitude from his beliefs about the attitude object and the evaluations of associated attributes. Each belief implies

22. Although on logical grounds a belief and its negation are mutually exclusive and exhaustive, and thus $p(\bar{b}) = 1 - p(b)$, psychologically this may not be true. For example, the concept "not aggressive" may not be perceived to cover everything that is the negation of "aggressive"; thus, psychologically, the concepts "passive," "peace loving," etc., may not be completely subsumed under the concept "not aggressive."

some degree of favorableness or unfavorableness toward the object. The most accurate inferences are likely to result when the following four pieces of information are available for each belief statement linking o to x.

1. $p(b_x)$ the probability that o is related to x
2. $p(b_{\bar{x}})$ the probability that o is related to $\bar{x}$, the negation of x
3. e_x the evaluation of x
4. $e_{\bar{x}}$ the evaluation of $\bar{x}$

As Table 3.4 shows, for each belief statement, products can be obtained of $p(b_x)e_x$ and $p(b_{\bar{x}})e_{\bar{x}}$. The sum of these two products is the best indicant of the subject's attitude implied by the statement in question. For example, Table 3.4 illustrates that a unipolar or probabilistic measure of the subject's belief that o is x and a measure of his evaluation of x lead to the inference of a negative attitude toward o in all three hypothetical cases. However, if the investigator also assessed the subject's belief that o is $\bar{x}$ and his evaluation of $\bar{x}$, and used both beliefs to infer the subject's attitude, he would have concluded that the subject had a positive attitude toward o. It is usually assumed that $p(b_x) = 1 - p(b_{\bar{x}})$, and that the evaluations of x and $\bar{x}$ are equally polarized with opposite signs (that is, $e_{\bar{x}} =$

Table 3.4 Computation of Item Scores

		Unipolar belief scale			Bipolar belief scale	
		p	e	pe	$p' = p - .5$	$p'e$
Case 1	x	.2	−3	−.6	−.3	.9
Symmetric probabilities: $p(b_{\bar{x}}) = 1 - p(b_x)$ and Symmetric evaluations: $e_{\bar{x}} = -e_x$	$\bar{x}$	.8	+3	2.4	.3	.9
	Σ			1.8		1.8
Case 2	x	.2	−3	−.6	−.3	.9
Symmetric probabilities: $p(b_{\bar{x}}) = 1 - p(b_x)$ and Asymmetric evaluations: $e_{\bar{x}} \neq -e_x$	$\bar{x}$	.8	+1	.8	.3	.3
	Σ			.2		1.2
Case 3	x	.2	−3	−.6	−.3	.9
Asymmetric probabilities: $p(b_{\bar{x}}) \neq 1 - p(b_x)$ and Symmetric evaluations: $e_{\bar{x}} = -e_x$	$\bar{x}$	.9	+3	2.7	.4	1.2
	Σ			2.1		2.1

$-e_x$), and under these assumptions, all necessary information could be derived from one measure of probability and one measure of evaluation (Case 1).

Unfortunately, such assumptions of symmetric probabilities and evaluations are not always warranted, and Table 3.4 shows that when either assumption is not met (Cases 2 and 3), appropriate measures of the subject's belief that o is $\bar{x}$ and his evaluation of $\bar{x}$ would lead to very different estimates of attitude than that based on deriving these measures under the assumptions of symmetry.

One way to avoid at least some of these problems is to treat beliefs in a bipolar fashion. That is, a response scale measuring beliefs can be scored either from − to +, indicating a range from disbelief to belief, or from 0 to +, implying a range from lack of belief to belief. Thus a unipolar probability scale ranging from 0 to +1 can be converted into a bipolar belief scale by subtracting the value of .5. The computations in Table 3.4 illustrate the effects of treating the probability measure in such a bipolar fashion. As we can see, when bipolar measures of beliefs are used, estimates based on both beliefs (o is x and o is $\bar{x}$) lead to the same inference about the direction of the subject's attitude toward o (positive) as estimates based on either belief considered alone. We can also see that when all four measures are available, using unipolar or bipolar belief scales will lead to the same results so long as the assumption of symmetric evaluations is met (Cases 1 and 3). When evaluations are asymmetric, however, different results are obtained by using unipolar and bipolar belief scales (Case 2).

Our comparison of standard attitude scales indicates that all procedures involve the multiplication of a belief and the evaluation of the associated attribute for each item. Since in all scaling methods only one b value and one e value are employed, it will clearly make a difference whether a unipolar or a bipolar belief scale is used.

Looking at Table 3.4, we can see that whenever a single belief is measured, use of a bipolar scale is preferable. Irrespective of the wording of the item, the same inference about the direction of attitude will be made. That is, one can usually assume that $p(b_{\bar{x}})$ and $p(b_x)$ will be negatively related (even if probabilistic symmetry cannot be assumed), and thus a respondent who has a high probability of endorsing a statement linking o and x will have a low probability of endorsing a statement linking o and $\bar{x}$. Since only one form of the statement will appear on a given scale, a unipolar scoring procedure will make very different inferences about the attitude, depending on the content of the particular item. For example, in Table 3.4 we can see that if x is evaluated −3 and $\bar{x}$ is evaluated +3, a .2 probability that o is x will be taken as an indication of a slightly unfavorable attitude (−.6), whereas a .8 probability that o is $\bar{x}$ will be taken as an indication of a very favorable attitude (+2.4). In contrast, with bipolar scoring, the response to either statement is indicative of the same attitude (+.9). The same logic applies when there is reason to assume asymmetric evaluations or probabilities. That is, Cases 2 and 3 also show that when only a single belief is measured (for example, o is x or o is $\bar{x}$), bipolar scoring leads to the same directional inference, but unipolar scoring does not. However, Table 3.4 suggests that when

there is reason to suspect that assumptions of symmetry are not met, both beliefs (b_x and $b_{\bar{x}}$) and their attribute evaluations (e_x and $e_{\bar{x}}$) should be obtained.[23]

Beliefs as indicants of attitude. Throughout the discussion above the argument has been that a person's attitude can be assessed by considering beliefs about the attitude object and evaluations of attributes associated with the object. All standard scaling methods obtain some measure of the person's beliefs about the object; a measure of attribute evaluation is not obtained from the person whose attitude is measured, but rather, it is assigned by the investigator, and it is assumed to be the same for all respondents. To meet this condition of uniform evaluations, many items are eliminated by the different scaling methods. Thus items that have positive attitudinal implications for one respondent but negative implications for another cannot be included on the standard scales. However, if the respondent's evaluation of each attribute were known, these items could be used to measure attitude.

Indeed, Fishbein (1967d) has argued that responses to any belief or intentional statement can serve as an indicant of a person's attitude, provided that his own evaluation associated with the statement and his belief or intention are known. Thus beliefs (b) and attribute evaluations (e) can be measured simultaneously for each respondent. It has been shown (e.g., Fishbein, 1963; 1967d) that, in accordance with an expectancy-value model, the summed products of b and e thus obtained can serve as a measure of attitude.

The considerations above point out that many beliefs that may serve as determinants of a person's attitude get eliminated from attitude scales by the standard scaling procedures. For example, it is often assumed that a person's attitude toward a political candidate is in part a function of the candidate's party affiliation. Thus the belief "Candidate X is a Democrat" should influence a person's attitude toward Candidate X. A belief statement of this kind, however, would be eliminated by the standard scaling methods since virtually everybody would agree with the statement, and thus its traceline would not meet scaling requirements. Nevertheless, the statement can be used to measure attitude, provided that the respondent's evaluation of "Democrat" is known. Thus a person who agreed with the statement and favorably evaluated "Democrat" would be assumed to hold a favorable attitude toward Candidate X. Similarly, a person who agreed with the statement but evaluated "Democrat" negatively would be viewed as having a negative attitude toward Candidate X.

Consistent with our discussion above, if the respondent assigned a low probability to the statement, or if he disagreed with it, inferences about his attitude

23. As Table 3.4 shows, when both beliefs are measured under conditions of evaluative symmetry (Cases 1 and 3), both forms of belief scoring will yield identical results. When evaluation is asymmetric (Case 3), however, unipolar measures do not yield the same results as bipolar measures. Since bipolar scoring is merely a convenience for dealing with the single-belief problem, we recommend the use of unipolar scales whenever both beliefs are assessed.

would require additional knowledge about his evaluation of "not a Democrat." In the absence of this information, evaluations of "Democrat" and "not a Democrat" are usually assumed to be symmetric, and therefore responses to the belief scale are scored in a bipolar fashion. However, the present example illustrates that the assumption of symmetric evaluations may not always be warranted. The possibility of asymmetric evaluations provides an additional reason for using several belief statements, rather than a single item, to measure attitude.

We argued earlier that responses to a single item can be viewed as consisting of a "true" score reflecting the attitude under consideration and some measurement error. Differences between the standard scaling procedures in terms of relative weights placed on beliefs and evaluations, inclusion or exclusion of disbeliefs and neutral items, etc. (see Table 3.3), can perhaps be regarded as measurement error. Across a large number of items, the errors will tend to cancel out, and the obtained attitude score will approximate the "true" attitude score. In fact, despite the existing differences, there is considerable evidence that the standard attitude scales tend to yield comparable results.

So far as obtaining a measure of attitude is concerned, therefore, the differences between the standard procedures are of relatively minor importance. These differences become relevant for an understanding of the formation of beliefs, attitudes, and intentions, for the study of changes in these variables, and for the prediction of overt behavior. For example, Likert and Guttman scaling arrive at the attitude score by *summing* $b \times e$ products, whereas an *average* is obtained in Thurstone scaling (see Table 3.3). Both a sum and an average have been used for semantic differential scales. Although the distinction between an averaging and an additive model has been of minor importance in attitude measurement, it has led to a theoretical controversy in research on attitude formation (see Chapter 6).

Response consistency. A final difference between the standard scaling procedures that is of theoretical significance concerns their implications for the notion of response consistency. In Chapter 1 we argued that attitudes are inferred from response consistency and that different interpretations can be given to this term. Item-selection procedures can be viewed as an attempt to examine responses to belief or intentional statements for a particular type of consistency. Thus Guttman scaling examines the degree to which cumulative response patterns are exhibited consistently by different individuals; that is, stimulus-response consistency is basic to the construction of Guttman scales.

In comparison, Likert and semantic differential scales examine the degree to which responses to one statement are consistent with responses to other statements. Use of a correlational selection procedure implies response-response consistency.

Finally, Thurstone scaling considers the degree to which a given individual responds consistently to items with the same scale value, as well as the degree to which he responds differently to items with different scale values. The consistency sought is between the evaluation associated with an item and responses to that item; this may be termed overall evaluative consistency.

The response consistency implied by the different item-selection procedures, however, is not necessarily the same as that implied by the final attitude scores. Specifically, with the exception of Guttman scores, the standard scaling procedures obtain attitude scores that reflect overall evaluative consistency. Two individuals with identical Thurstone, Likert, or semantic differential scores may exhibit completely different response patterns. Strictly speaking, the notion of response-response consistency implies that all of a person's responses indicate the same attitude toward the object in question. The absence of such consistency (e.g., if his response to one item indicated a favorable attitude and his response to another item indicated an unfavorable attitude), implies absence of an attitude. However, like Thurstone scaling, Likert and semantic differential scales result in attitude scores for all individuals, even when response-response inconsistency occurs.

In Guttman scaling, stimulus-response consistency is implied by the attitude score, just as it was implied by the item-selection procedure. That is, in a cumulative Guttman scale, attitude scores are associated with one—and only one—response pattern.

The lack of correspondence between the kinds of response consistency implied by item-selection procedures and by attitude scores in Likert and semantic differential scaling is reminiscent of a similar lack of correspondence discussed in Chapter 1. There we mentioned that investigators often obtain attitude scores that reflect overall evaluative consistency, but they assume that attitudes are predispositions to perform a particular response or a consistent set of responses; that is to say, the investigators assume stimulus-response or response-response consistency. The present discussion indicates that since most standard procedures result in overall evaluative attitude scores, those scores cannot be expected to predict individual responses.

In conclusion, although there are many important differences between standard attitude-scaling methods, they share two basic features. First, all four methods discussed infer attitude by considering a person's beliefs or intentions and their associated evaluations. Second, in all methods, those beliefs and associated evaluations are combined into an attitude score in accordance with an expectancy-value formulation. These conclusions support our conceptual framework, in which attitude is viewed as a function of an individual's beliefs and is expected to be related to his intentions. One criticism sometimes voiced with respect to the standard scaling procedures is that they frequently do not include belief items that are important determinants of the attitude in question. In fact, as we have indicated, the different item-selection procedures may eliminate just such belief statements. This problem can be overcome by obtaining independent measures of evaluations associated with a belief, as well as of belief strength. This more complete use of an expectancy-value notion obviates the need to eliminate any items in attitude measurement.

In the remainder of this chapter, we will see that other verbal attitude measures also appear to be based on an expectancy-value formulation involving beliefs

and associated evaluations. Further, we will see that certain measures of cognitive structure can also be obtained by considering these beliefs and/or their associated evaluations.

ALTERNATIVE MEASUREMENT TECHNIQUES

All measurement procedures discussed so far, including the standard scaling methods, can be viewed as nondisguised techniques. That is, the respondent is usually aware that his attitude is being assessed; no attempt is made to conceal the purpose of the measurement. One possible exception is the semantic differential, which is sometimes presented to the respondent as a measure of meaning, and the inclusion of nonevaluative scales may disguise its purpose to some degree.

Disguised Techniques

Most other attitude-measurement procedures which have been developed make an explicit attempt to disguise the purpose of the measuring instrument. The underlying assumption is that when the purpose of the instrument is not apparent, respondents are less likely to "distort" or "falsify" their responses, and thus a more valid measure of attitude can be obtained. Usually the respondent is led to believe that his responses will be interpreted as indicants of some variable other than attitude.[24] (For an excellent comprehensive review of disguised measurement techniques, see Kidder and Campbell, 1970.) The present discussion will be concerned primarily with those techniques that have demonstrated some degree of utility for the measurement of attitude. Numerous attempts to measure attitudes have been based on projective personality tests, such as the Thematic Apperception Test (TAT), Rorschach, and doll play, but these techniques will not be reviewed here, because there is little evidence for their reliability or validity as attitude measures (cf. Sechrest, 1968).

Hammond's Error-Choice Technique

One way of disguising the purpose of an attitude-measuring instrument is to lead the respondent to believe that he is to make some objective judgment, i.e., that there is some "correct" response. An example of this kind of disguise is given by Hammond's (1948) error-choice technique, in which respondents are led to believe that their factual knowledge concerning some object is being assessed. The underlying assumption is that attitudes will bias responses in a consistent way and, therefore, that such biased responses can be used to infer attitude. "Under the guise of an 'information test' items [are] presented with alternate choices for an-

24. Recent concern with the ethics of psychological research has led to criticism of such deceptive practices. Some more general questions of deception will be considered in Chapter 4.

swers, both choices being incorrect, either equidistant from the truth or truth being indeterminable. The subject [is] forced into error. This method is termed the 'error-choice technique'." (Hammond, 1948; p. 43)

For example, the following item, which has response alternatives equidistant from the truth, appeared on Hammond's original scale measuring "labor-management" attitudes.

Average weekly wage of the war worker in 1945 was

1. $37
2. $57

The correct figure would have been $47. Hammond assumed that prolabor subjects would be biased toward the first response and promanagement subjects toward the second.

On a second scale measuring attitude toward Russia, the following item had no determinably correct response.

Russia's removal of heavy industry from Austria was

1. Legal
2. Illegal

Note that this item is very similar to a dichotomous semantic differential scale. That is, agreement with the first alternative would be taken to imply a positive attitude toward Russia, and agreement with the second would be viewed as evidence for a negative attitude. The same argument can be made for the first item concerning labor-management attitudes; one response alternative would be viewed as indicating a favorable attitude ($e = +1$), the other an unfavorable attitude ($e = -1$).

For each item, the respondent indicates which of the two bipolar alternatives is associated with the attitude object ($b = 1$). Consistent with an expectancy-value formulation, the $b \times e$ products are computed for each item, and the attitude score is a sum of those products. To test the validity of each item as an indicant of attitude, as well as the validity of the total score, Hammond used the "known group comparison" technique. His respondents were union employees and businessmen. Most item scores and the total attitude score were found to discriminate between these two groups. Use of known group comparisons indicates that, as in Likert scaling or the semantic differential technique, items are selected in terms of their internal consistency. That is, each item must be shown to discriminate between subjects with positive and negative attitudes; thus a linear traceline is assumed. This suggests that in constructing an error-choice scale, one could select items by computing the correlation between item scores and the preliminary total score in lieu of using known group comparisons. Items with low correlations would be eliminated.

Cook's Plausibility Technique

Another way of disguising the purpose of an attitude measure involves the following procedure. Instead of indicating his agreement with a series of statements, the respondent judges each statement in terms of its validity, effectiveness, evaluative implication, or plausibility. For example, the task of judges in Thurstone scaling is to rate the favorableness of each item on an 11-place scale. A disguised measure of attitude would be obtained if these ratings were used to infer the judges' attitudes. In fact, as will be seen below, such judgments are used to infer involvement and other indices by a technique known as the "own categories" procedure.

Perhaps the best example of such a disguised technique in attitude measurement is Cook's measure of plausibility (Waly and Cook, 1965; Brigham and Cook, 1970). The underlying assumption is that a person's attitude will bias his judgment of a statement's plausibility. Specifically, it is assumed that "a statement will be considered a more plausible or effective argument by subjects who agree with the position which the statement supports than by subjects who disagree with that position" (Waly and Cook, 1965, p. 746). In developing a plausibility measure of attitude toward racial integration, Waly and Cook had their subjects rate the effectiveness of each of a set of arguments labeled "prosegregation" or "prointegration" on an 11-point scale ranging from *very ineffective* (−5) to *very effective* (+5). The following two items were among the arguments rated.

1. *Prosegregation:* Immediate desegregation damages the Negro by giving him responsibilities for which he does not have the necessary background and preparation.
2. *Prointegration:* If integration is adopted, race hatred will quickly disappear in southern communities which are now torn apart by the issue.

"Subjects' ratings on the prosegregation items were assigned scores from +5 = 0 to −5 = 10; that is, the more *ineffective* the argument was considered to be by the subject, the higher his score. Scores for the prointegration items were reversed, that is, the more *effective* the argument was considered to be by the subject, the higher the score." (Waly and Cook, 1965, p. 746) The item scores are summed to obtain a measure of attitude. The higher the attitude score, the more favorable the attitude toward integration.

This scoring procedure is very much like that used in Likert scaling. A given item is assumed to indicate either a favorable (+1) or an unfavorable (−1) attitude, and judgments of effectiveness are given scores of −5 to +5. As noted above, the assumption is that judging a statement as effective indicates agreement with it and judging it as ineffective indicates disagreement. Under this assumption, judgments of effectiveness can be viewed as measures of belief strength. Consistent with an expectancy-value formulation, $b \times e$ products can be obtained for each statement and summed to yield a measure of attitude. Indeed, this procedure yields a score which is equivalent to the score obtained by Waly and Cook.

Just as the error-choice technique was shown to be similar to the semantic differential, the plausibility technique can be viewed as similar to a Likert scale. Items are assumed to discriminate between individuals with positive and negative attitudes; i.e., they are assumed to have linear tracelines. This suggests that a scale based on plausibility judgments should be constructed by applying Likert item-selection criteria. Only discriminating items should be retained.

Recall that disguised techniques were developed in part in an attempt to avoid distortions or falsifications of responses used to infer attitudes. The basic assumption underlying these techniques is that if the subject is unaware that his attitude is being assessed, his responses are more likely to be valid indicants of his "true" attitude. The error-choice and plausibility techniques attempt to achieve this aim by leading subjects to believe that their factual knowledge is being measured or that the investigator is interested in judgments of plausibility. These responses are assumed to be systematically biased by the respondent's attitude. Clearly, the validity of the error-choice and plausibility techniques rests on the validity of that assumption.

Estimation of Others' Responses

A similar assumption is made when subjects are told that their knowledge of other people's opinions is being measured. Subjects can be asked to respond to any question as they think some other person or group of persons would respond. Based on psychodynamic theory, the assumption is that subjects project their own beliefs, attitudes, or intentions onto those other persons and thus, when they respond "as they think others would," they actually respond the way they themselves would. Unlike the previous disguised techniques, this one does not require development of a specific instrument. That is, a subject can be asked to indicate the way that "most people" would respond to any of the standard attitude scales, to the error-choice technique, or to any other verbal measure. Since estimates of others' responses are assumed to be equivalent to the subject's own responses, the normal procedures for computing attitude scores can be employed.

It is worth noting, however, that the assumption of equivalence between projected and own responses may hold for some items but not for others. It follows that an item may not be a valid indicant of attitude when estimates of other people's responses are obtained, even though it may be valid when the subject's own responses are used. For example, if subjects with favorable attitudes agree with an item but subjects with unfavorable attitudes disagree with it, the item meets Likert criteria and is considered a valid indicant of attitude. The same item, however, may fail to discriminate between subjects with favorable and unfavorable attitudes when estimates of other people's responses are involved. All subjects, irrespective of their attitudes, may estimate that most other people would agree (or disagree) with the item. Such an item, therefore, should be eliminated from the scale when subjects are asked to make estimates of others' responses.

Recall that item analyses can be done in two ways: by comparing responses of subjects with presumed-known attitudes (known groups comparison) or by comparing responses of subjects whose preliminary attitude scores are high or low (internal consistency). Clearly, when the known groups comparison method is used, items are selected in terms of a single "objective" external criterion. In contrast, the internal consistency criterion may differ for the selection of items based on subjects' own responses as opposed to estimates of other people's responses. Thus attitude scores based on disguised items will be similar to attitude scores based on nondisguised items when the known groups comparison criterion is used in item selection. The similarity will not necessarily exist when the internal consistency criterion is employed. This implies that an independent item analysis must be performed to construct an attitude scale that employs estimates of other people's responses. The failure to perform such item analyses may explain the relatively low validity of scales that have used the estimation procedure.

The "Bogus Pipeline" Technique

Another general procedure designed to measure a subject's "true" beliefs, attitudes, or intention has been suggested by Jones and Sigall (1971). The basic assumption underlying the "bogus pipeline" procedure is that "no one wants to be second-guessed by a machine. If a person could be convinced that we *do* have a machine that precisely measures attitudinal direction and intensity, we assume that he would be motivated to predict accurately what the machine is saying about him." (Jones and Sigall, 1971, p. 349)

With this technique, the respondent is attached to some machine that purportedly can measure his true response to any item by recording his implicit reactions to the item. Needless to say, subjects must be convinced that the machine can indeed do so. Subjects can then be asked to guess or predict the machine's readings of their responses to any item. Thus, in a sense, the subject's task is to predict his own "true" implicit responses. On the assumption that he does not want to be outguessed by the machine, the subject is expected to respond as truthfully as possible.

Under this procedure, a subject can be asked to predict his responses to any item or set of items, including standard attitude scales. Note that if a set of belief statements is to be used to infer the respondent's attitude, the same problems discussed with respect to estimates of other people's responses apply to predictions or estimates of one's own responses. That is to say, an item analysis must be performed to construct an attitude scale where attitudes are inferred from estimates of own responses.

Frequently, however, the bogus pipeline technique will be used to obtain a "direct" measure of the subject's true attitude. For example, the subject may be asked to predict the machine's reading of his response to the attitude object "desegregation" on a seven-point bipolar scale ranging from *good* to *bad*. As pointed

out at the beginning of this chapter, responses to this item may be viewed as relatively direct measures of attitude. Clearly, the bogus pipeline technique may elicit a different response from that which would be obtained if the subject were simply asked to rate "desegregation" on the seven-point *good–bad* scale.

The possibility of obtaining different results by using disguised and nondisguised techniques raises the question as to which type of instrument provides a more accurate measure of attitude. Proponents of disguised techniques argue that since these techniques are designed to avoid deliberate distortions and falsifications, they result in more valid measures of attitude. However, the question of validity is an empirical one and will be discussed in the following chapter.

Physiological Measures

In contrast to the disguised measurement techniques considered thus far, where an attempt is made to avoid distortion by concealing the instrument's purpose, physiological measures attempt to prevent distortion by assessing involuntary responses over which the individual has little or no control. A large number of physiological responses have been considered in the search for a valid, nonverbal indicant of attitude. Among them are the galvanic skin response (GSR) measuring electrical skin conductance, heart rate, palmar sweat, pupillary dilation and constriction, respiration, etc. Despite considerable research efforts, there is little evidence to indicate that any physiological measure can be used as a valid indicant of attitude. One problem is that most physiological measures appear to assess general arousal and therefore cannot be used to distinguish between positive and negative affective states. Since attitude has been defined in terms of an evaluative dimension ranging from favorableness to unfavorableness, these measures are by definition inadequate.

The pupillary response. One physiological measure, however, appeared for a time to be capable of distinguishing between positive and negative affective states, namely, the pupillary response. Specifically, Hess (1965) suggested that increase in pupil size (dilation) indicates a positive attitude toward the object viewed, whereas constriction in pupil size indicates a negative attitude. Further, he suggested that the amount of dilation or constriction corresponds to differences in attitudinal intensity. Unfortunately, the initial promise of this measure has not been fulfilled. In an excellent review of the pupillary response as a measure of attitude, Woodmansee (1970, p. 532) concluded that "studies which are relevant to the issue suggest that the pupil does not measure attitude or qualitatively different affective states. There is ample evidence, however, that the pupil, in its reflex dilation reaction, may be used to indicate arousal, attentiveness, interest, and perceptual orienting."

In conclusion, it would definitely be desirable to have a nonverbal measure of attitude not under the subject's control, but it appears unlikely that any known physiological reaction will serve this purpose.

Behavioral and Unobtrusive Measures

Various overt behaviors have been used to infer attitudes. For example, church attendance has been used to measure religiosity, choice behaviors are assumed to express preferences, and duration of eye contact has been assumed to reflect interpersonal attraction, as has amount of communication and volunteering to help. Unobtrusive measures have included return rate of ostensibly lost letters addressed to various persons or organizations, relative frequencies with which bumper stickers in support of various political candidates are displayed, and purchase frequencies of various brands of a given product.

As previously mentioned, when such behavioral observations are used to infer attitudes, they are comparable to verbal responses to items on an attitude scale. The implication of this notion is that overt responses must be submitted to the same scaling procedures that are applied to verbal items before they can be assumed to reflect attitude. It should be obvious that not all behaviors with respect to a given object are valid indicants of attitude toward that object. This issue will be considered again in Chapter 8 on the relation between attitude and behavior.

Multidimensional Scaling Procedures

All measurement techniques discussed thus far attempt to locate an individual on a single bipolar evaluative dimension with respect to some object. Although this approach is consistent with our definition of attitude, the reader should recall that other investigators have defined attitude as a complex multidimensional concept. Clearly, this multidimensional conception of attitude is incompatible with the measurement techniques we have discussed. Thus some investigators have turned to multidimensional scaling procedures, where an attempt is made to measure attitude in terms of the object's location in a multidimensional space.[25]

The major purpose of multidimensional techniques is to identify the relevant dimensions that underlie a person's judgments of a given object or set of objects. Thus, although one expects that different dimensions may be relevant for different subjects and for different objects, these dimensions should *not* vary as a function of the particular set of judgments the subject is asked to make.

The most frequently used multidimensional technique is factor analysis. We showed earlier in this chapter how factor analysis was used to construct the semantic differential measure of attitude. On the basis of Osgood's work on the semantic differential, Triandis (1964) developed the "behavioral differential" to investigate dimensions of interpersonal intentions. First, a large set of interpersonal behaviors was selected, such as "elect to political office," "go fishing with," "accept as an intimate friend," "work for," and "exclude from my neighborhood."

---

25. For example, some investigators have used the semantic differential and have argued that attitude toward a concept is defined by the person's location on all emerging dimensions, rather than only the evaluative dimension.

Subjects were asked to indicate on a nine-point scale whether they *would* or *would not* perform each behavior with respect to a variety of complex stimulus persons (e.g., a 50-year-old, Negro, Roman Catholic, male physician).

A factor analysis was performed on the matrix of correlations between behaviors, and five dimensions of interpersonal intentions were identified. These dimensions were interpreted as (1) formal social acceptance or respect; (2) marital acceptance; (3) friendship acceptance; (4) social distance; and (5) superordination-subordination. These results will be discussed in greater detail in Chapter 7, which deals with the formation of intentions. There we will see that different factors tend to emerge in different studies.[26]

Indeed, a major problem with the factor-analytic approach is that the emerging dimensions *are* dependent on the particular set of items used to elicit judgments. That is, factor analysis identifies the dimensions underlying a given set of judgments rather than all relevant dimensions; any dimension not represented by the particular set of judgments used cannot be identified. For example, if a semantic differential included only a set of evaluative scales, potency and activity dimensions would not emerge in a factor analysis of those scales. To avoid this difficulty, a number of context-free multidimensional scaling procedures have been developed. Most of these procedures are based on judgments of distance (or similarity) between objects. The basic principle involved can be exemplified as follows: Consider an investigator who is interested in identifying the relevant dimensions along which three objects, A, B, and C, are judged or compared. As a first step, he could ask subjects to judge the degree of similarity between the three possible pairs of objects (AB, AC, and BC) on a seven-point scale ranging from *similar* (0) to *dissimilar* (6). Figure 3.5 presents hypothetical responses of two subjects.

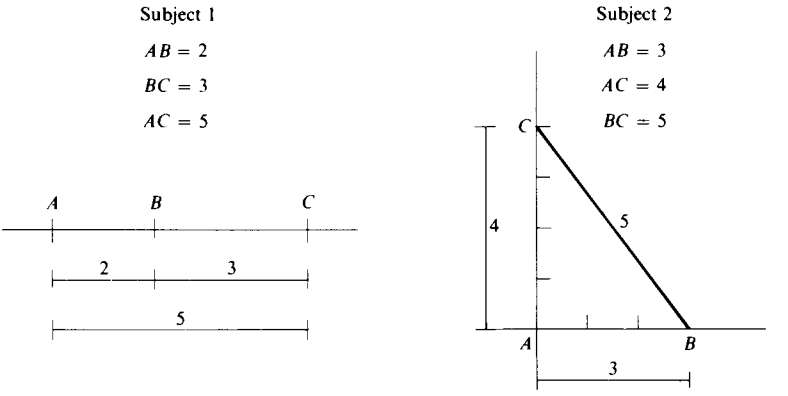

Fig. 3.5 Hypothetical similarity judgments and their use in multidimensional scaling.

26. It is worth noting, however, that irrespective of the label assigned to it, the first factor to emerge can usually be interpreted as an evaluative dimension.

The assumption is that the more similar a pair of objects, the closer their locations in multidimensional space. Thus similarity judgments can be interpreted as measures of the distance between objects in that space. The analytic problem of multidimensional scaling is the determination of the number of dimensions required to locate the objects' positions in the space in accordance with their distances from each other. In the example of Fig. 3.5, only one dimension is necessary to account for the first subject's distance estimates, but two dimensions are required for the second subject. This principle can be generalized to any number of objects, and the analytic procedure will determine the number of dimensions that are required.

The basic data used in multidimensional scaling are contained in a matrix of distances between a set of objects.[27] A procedure similar to factor analysis is employed, and as in factor analysis, a set of dimensions is identified. In addition, each object's location on these dimensions is obtained. Each dimension is given a psychological interpretation (i.e., it is named) by considering the kinds of objects that are located at opposite ends of the dimension. Figure 3.6 illustrates a hypothetical two-dimensional space for the judgment of politicians. The locations of

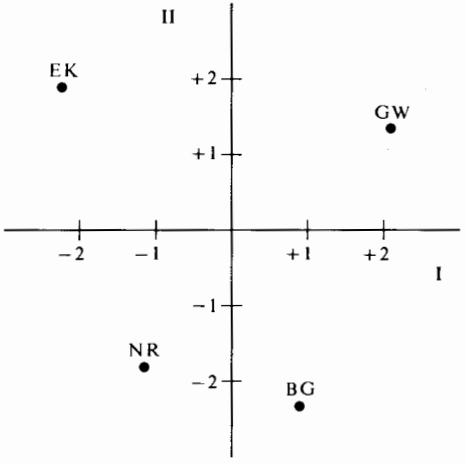

Fig. 3.6 Hypothetical two-dimensional space.

politicians in this space are determined by the values obtained in multidimensional scaling, shown in Table 3.5.

Since the two Democrats (Kennedy and Wallace) are located at one end of the second dimension and the two Republicans (Goldwater and Rockefeller) at the other, it seems reasonable to label this dimension "party affiliation." The first

27. In most applications, the entries in each cell of the matrix are mean distance ratings over a sample of subjects. If a given individual's set of dimensions is to be obtained, however, his own distance ratings must be analyzed.

Table 3.5 Hypothetical Results of Multidimensional Scaling

	Dimension I	Dimension II
Barry Goldwater	+ .9	−2.3
Nelson Rockefeller	−1.1	−1.8
Edward Kennedy	−2.2	+1.9
George Wallace	+2.1	+1.3

dimension might be labeled "liberalism-conservatism" since Kennedy and Rockefeller are located on one end and Goldwater and Wallace on the other. However, this dimension could also have been interpreted as a geographic dimension since Kennedy and Rockefeller are from the North while Goldwater and Wallace are from the South. Within our conceptual framework, all dimensions emerging in multidimensional scaling represent different beliefs about the objects. Returning to our example in Fig. 3.5, we can interpret the second dimension as beliefs about party affiliation. Similarly, the first dimension can be interpreted as beliefs about the politicians' liberalism or as beliefs about their regional origins. However, as in factor analysis, one of the emerging dimensions can usually be interpreted as evaluative. This is always the case when the analysis is performed at the level of the individual subject, but a clear evaluative dimension may not emerge when the analysis is based on data from a sample of respondents (see note 27). In our example, the first dimension can be interpreted as evaluative, since it could be argued that the politicians are ordered in terms of their attractiveness. This implies that some beliefs are related to the evaluative dimension while others are unrelated. However, as we pointed out in discussing factor analysis, for another set of politicians, liberalism-conservatism may not be related to the evaluative dimension (it might emerge as an independent dimension), while beliefs about party affiliation may be related.[28]

In conclusion, multidimensional scaling can be used as a general measure of attitudes toward a set of objects since, at least at the individual level, it locates those objects on an evaluative dimension. Moreover, it identifies the number of underlying belief dimensions that are used by a person in making judgments about the objects.[29] This suggests that different respondents may be compared in terms of the number of belief dimensions they employ in judging a given set of objects.

28. The same considerations apply to interpreting the dimensions of interpersonal intentions that have been obtained in studies using the behavioral differential. That is, although all emerging dimensions can be viewed as representing different kinds of interpersonal intentions with respect to a given set of persons, one dimension is usually evaluative in nature and can thus be used as a measure of attitude.

29. All emerging dimensions should be considered beliefs; however, in all individual analyses, and in many group analyses, one of the dimensions can also be interpreted as an evaluative dimension.

A measure of "cognitive complexity" can be obtained in this fashion. It should be clear that two individuals with similar attitudes toward one or more objects can differ in their degree of cognitive complexity. Measures of this kind may be useful in explaining why people with similar attitudes behave in different ways or differ in their acceptance of new information. Indeed, many measures of cognitive structure have been employed in attitude research.

Measures of Cognitive Structure

Associated with a person's attitude is a set of beliefs about the object of the attitude. In our consideration of learning theory we have suggested that these beliefs form a hierarchy in terms of the strength with which they are held. This hierarchy may be viewed as a person's belief system with respect to a given object or issue. Various measures of cognitive structure can be based on this belief system. To do so, however, requires that an adequate procedure be available for eliciting the person's belief hierarchy. Clearly, the technique employed should permit any belief in the hierarchy to be elicited, and it should not elicit beliefs that are not part of the person's belief system. Unfortunately, none of the techniques that have been used to elicit beliefs seems to completely satisfy these requirements.[30]

Perhaps the simplest and most direct procedure involves asking the subject to describe the attitude object, using a free-response format. For example, he could be asked to list "the characteristics, qualities, and attributes" (Zajonc, 1954) of an object such as "organized religion," or he could be asked to list the consequences of performing some behavior.

The number of beliefs elicited in a free-response situation has been used as an index of cognitive differentiation. Clearly, once the person's belief hierarchy has been obtained, he can be asked to rate his own beliefs in a variety of ways. Within the framework of an expectancy-value notion, the probability of each belief and the evaluation of the associated attribute would be assessed. It has been shown how these measures can be used to compute an attitude score, but it is also possible to derive measures of cognitive structure from the same measures. For example, an index of the intensity with which beliefs are held would be obtained by computing the average belief strength (i.e., probability). Two persons with the same attitude can have cognitive structures varying in intensity, and they may be differentially resistant to attitude change. A measure of total affect within the belief system can be obtained by summing the absolute values of the $b \times e$ products. The difference between this index and the absolute value of attitude could be considered a measure of ambivalence with respect to the object (K. J. Kaplan, 1972). Consistency among beliefs can be assessed by computing the ratio between the number of beliefs with positive (or negative) associated evaluations and the total number of beliefs. Table 3.6 illustrates these indices with respect to two persons' beliefs about Toyotas. It should be clear that many other indices of cognitive struc-

30. These procedures include sentence completion, word association, adjective check lists, and the "repertory grid." A review of some of these techniques can be found in Triandis (1971).

Table 3.6 Indices for Two Hypothetical Belief Systems Concerning Toyotas

Person 1				Person 2			
Beliefs elicited	b	e	$b \times e$	Beliefs elicited	b	e	$b \times e$
Reliable	+3	+2	6	Inexpensive	+3	+1	3
Inexpensive	+2	0	0	Small	+3	−1	−3
Inconvenient	+2	−3	−6	Made in Japan	+3	−2	−6
Small	+2	−1	−2	Economical	+1	+2	+2
Ugly	+1	−2	−2				
(1) Attitude $\Sigma\, b_i e_i$			−4				−4
(2) Differentiation: No. of beliefs			5				4
(3) Intensity $\lvert\Sigma\, b_i\rvert/(2)$			2				2.5
(4) Total affect $\Sigma\lvert b_i e_i\rvert$			16				14
(5) Ambivalence (4) − $\lvert(1)\rvert$			12				10
(6) Consistency $\frac{\text{No. of neg } (b \times e) \text{ products}}{(2)}$			0.6				0.5

ture can be obtained on the basis of the b and e values. Further, subjects could be asked to rate the importance of each belief or to estimate conditional probabilities between beliefs (e.g., subjects could indicate the probability that a car is inexpensive if it is made in Japan). These and other measures could then be used to compute additional indices of cognitive structure.[31]

Position Uncertainty, Belief Confidence, and Indifference

As we noted in our discussion of belief measures, subjects are often asked to place an object into one of several content categories. The categories may be ordered along some dimension, or they may be merely nominal categories. This procedure identifies the perceived object-attribute link. The indices of cognitive structure discussed above are based on subjective probabilities associated with the object-attribute link and on evaluations of the attribute category.

Position uncertainty. In addition to estimating the probability that the object has the indicated attribute, the subject can be asked to rate how confident or *certain* he is that the object has the attribute in question. There is considerable evidence that such certainty ratings are systematically related to probability estimates; the

31. Clearly, some of these indices can also be constructed without eliciting the respondent's own beliefs but rather by using some standard set of belief statements. Interested readers are referred to Zajonc (1960) and Scott (1969) for examples of indices of cognitive structure and their use in attitude research.

more polarized or extreme (either high or low) the person's probability estimate, the greater his certainty. Minimal certainty (or maximal uncertainty) is found when the probability estimate is at chance level (Beach and Wise, 1969; Wyer, 1973).

Assume a set of mutually exclusive and exhaustive attribute categories. The probabilities associated with these alternative categories must sum to 1.0. The probability defining the chance level is $1/n$, where n is the number of attribute categories. Maximal uncertainty with respect to a given object-attribute link thus occurs when the subjective probability of that link is $1/n$. With two response alternatives, maximal uncertainty is at the .50 level of probability (½). When four response categories are available, a probability of .25 represents chance level and thus maximal uncertainty.

Since position uncertainty decreases as subjective probabilities deviate from chance levels, there should be a U-shaped relation between probability and certainty. This relationship is shown in Fig. 3.7 for the dichotomous case. Support for this relationship has been reported by Beach and Wise (1969), who concluded that certainty ratings "are essentially another . . . version of the subjective probability estimates" (p. 441). Once an object has been assigned to a given attribute category, the subject's certainty in his judgment is essentially equivalent to his subjective probability that the object belongs in that category.

Thus far we have discussed an individual's uncertainty associated with his response category. It is also possible to consider uncertainty associated with the entire set of response categories. Clearly, this overall uncertainty will be re-

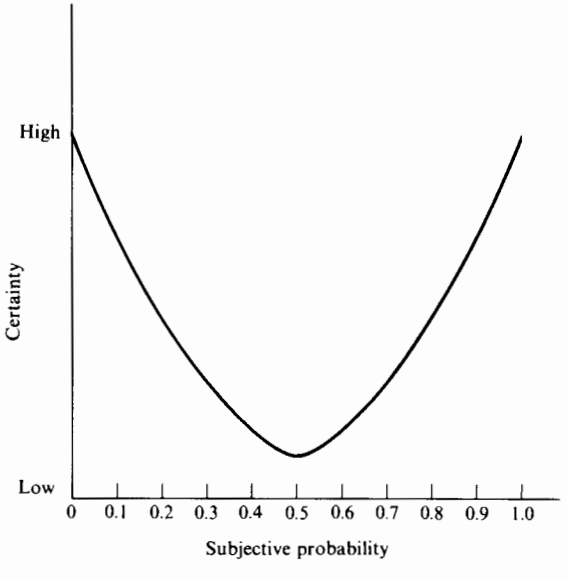

Fig. 3.7 Reaction between subjective probability and certainty.

lated to the subjective probabilities associated with each of the different response alternatives available to the subject. Consider, for example, a subject who rates "cornflakes" as *quite nutritious* on the following scale.

		X	
not at all nutritious	slightly nutritious	quite nutritious	extremely nutritious

We saw previously that he can also be asked to indicate his subjective probability that "cornflakes are quite nutritious." The higher this probability, the more certain he should be that his judgment is correct.

In addition, the subject can also be asked to indicate his subjective probabilities that cornflakes are associated with each of the other three response categories, i.e., that cornflakes are not at all nutritious, slightly nutritious, and extremely nutritious. The subjective probabilities associated with the four attribute categories can now be used to compute an index of overall uncertainty with respect to this attribute dimension. Generally speaking, the lower the variability among these subjective probabilities (i.e., the more similar they are to each other), the higher is the person's overall uncertainty.[32] In our example, maximal uncertainty is expected when cornflakes are equally likely to be rated extremely, quite, slightly, or not at all nutritious. Minimal uncertainty is expected when the respondent's subjective probability that cornflakes belong in one of the four categories is 1.0 and that they belong in any of the other three categories is 0. Wyer (1973) has shown that overall uncertainty associated with an attribute dimension is related to the person's uncertainty with respect to his own position on the dimension. When a person believes that the object may be associated with any one of the attribute categories (high overall uncertainty), his certainty about his own preferred category tends to be low. However, this relation between overall uncertainty and position uncertainty may not hold for any position other than the person's own. For example, a person whose overall uncertainty is moderate may assign a low probability to one of the content categories. With respect to the latter category, his position certainty might be quite high.

Belief confidence. Two persons who assign the same subjective probability to a given position on the content dimension may vary in their confidence associated with this probability estimate. It is thus possible to distinguish between position

32. More specifically, it is possible to use the information theory index of average uncertainty,

$$u = -\sum_{i=1}^{n} P_i \log_2 P_i,$$

where u is uncertainty, P_i is the subjective probability of response category i, and n is the number of response categories (cf. Wyer, 1973).

certainty and belief confidence.[33] Unlike position certainty, belief confidence may be unrelated to the subjective probability; that is, a person's belief confidence may be high or low irrespective of his subjective probability. However, it seems likely that belief confidence is related to stable individual difference variables. Thus some persons tend to be more confident in their beliefs than others. Some evidence indicating stable individual differences in belief confidence was provided by Ewing (1942).

Indifference. In addition to obtaining probability estimates for each response category, one can obtain evaluations of each category. To return to our example, subjects could be asked to rate "extremely nutritious," "quite nutritious," "slightly nutritious," and "not at all nutritious" on a *good–bad* scale.

The variance of the individual evaluations could be taken as one possible index of the person's *indifference.* When the alternative response categories are evaluated equally, the respondent may be said to be maximally indifferent to the attribute dimension in question (i.e., "nutritiousness" in the present example). Minimal indifference (or maximal concern) is indicated by a large variance of evaluations. Two persons with similar evaluations of a given attribute category may vary in terms of their indifference toward the attribute dimension. One interesting implication of these considerations concerns the effect of changing a person's location on a content dimension on his attitude toward the object of the belief. When the person is indifferent toward the attribute dimension, a change in belief from one content category to another may have little effect on his attitude.[34]

In sum, measures of beliefs and their associated evaluations can be supplemented by indices of position uncertainty, belief confidence, and indifference. As with other measures of cognitive structure, these indices may prove useful for an understanding of the effects of persuasive communications and other factors on the formation and change of beliefs, attitudes, and intentions.

All indices of cognitive structure discussed so far are based on measures of belief strength and attribute evaluations. The "own categories procedure" has

---

33. In statistical terms, belief confidence may be viewed as analogous to the confidence interval around a parameter estimate. Thus position certainty is equivalent to a probability judgment, and belief confidence refers to the confidence interval around this judgment.

34. The possibility of obtaining probability and evaluation ratings for each category on a content dimension again suggests that more than one $b \times e$ product can be computed for each content dimension. Earlier we considered the simplest case of a dichotomous dimension, and we saw that it is often necessary to consider probabilities and evaluations associated with both response categories. This notion can now be generalized to any number of response alternatives. It seems doubtful, however, that consideration of multiple categories would greatly increase the accuracy of attitude measurement. Further, there are obvious practical limitations to such an extension.

been used to derive a number of cognitive structure measures that do not rely on either of these two variables.

The Own Categories Procedure

On the basis of their work in social judgment, Sherif and Hovland (1961) questioned Thurstone's assumption that a judge's own attitude toward some object does not influence his judgments of the favorableness of statements concerning the object. Their attempts to demonstrate effects of attitudes on judgments eventually led to the development of an instrument known as the own categories procedure. Consistent with their expectations, Sherif and Hovland found that judgments of the favorableness of some statements were influenced by the judge's attitude. Specifically, judges with extreme attitudes rated some statements differently than did a random sample of judges. The practical implications of this phenomenon for Thurstone scaling, however, are relatively minor since the scale values used for computing attitude scores are based on a sample of judges representative of the population to be studied. Further, even when shifts in scale values do occur, they are relatively small.

In analyzing the responses of judges with extreme attitudes, Sherif & Hovland found that these judges made differential use of the response categories. Although all extreme judges tended to place relatively few items into neutral categories, judges with extremely positive attitudes placed a disproportionate number of items into unfavorable categories, and judges with extremely negative attitudes placed a disproportionate number of items into favorable categories. It thus appeared that involved subjects (with extreme attitudes) tended to use fewer categories than noninvolved subjects. The own categories procedure is based on this premise and has been used primarily to measure involvement with some attitude object or issue.

Instead of being asked to sort statements into 11 categories, the respondent is simply asked to "sort the pool of items into any number of piles, or categories, that seem to him necessary, so that items within each category seem to him to 'belong together' . . . he is told to place the items into categories as objectively as possible in terms of how 'favorable' or 'unfavorable' they are toward the persons, events, issues, or objects in question" (Sherif and Sherif, 1967, pp. 190–191). The number of categories used is taken as a measure of involvement; the fewer the categories, the greater the involvement. According to Sherif and Sherif, the number of items in each category can be used as an index of category width.

After they have sorted items into piles, subjects can be asked to indicate (a) the pile which comes closest to their own view of the object or issue (most acceptable pile), (b) other piles containing acceptable statements, (c) the pile which is most objectionable from their point of view (most objectionable pile), and (d) other piles containing objectionable statements. The number of statements in all acceptable piles is taken as a measure of the subject's *latitude of acceptance;*

the number of statements in all objectionable piles constitutes a measure of *latitude of rejection;* and the number of remaining items defines the *latitude of noncommitment.*

When Thurstone scale values are available for items used in the own categories procedure, the person's attitude is usually determined by averaging the scale values of all items in the most acceptable category. For all practical purposes, this attitude score is the same as a traditional Thurstone attitude score. Latitudes of acceptance, rejection, and noncommitment can now be defined in terms of the range of scale values in these regions.

A more direct procedure for obtaining measures of a person's latitudes of acceptance, rejection, and noncommitment involves use of an ordered set of items or any other representation of an evaluative dimension (Sherif, Sherif, and Nebergall, 1965; Sereno and Mortensen, 1969). Subjects are asked to check the most acceptable item or position on the scale,[35] other acceptable positions, the most objectionable position, and other objectionable positions.

We can see, then, that the own categories procedure is primarily a measure of involvement but also provides various indices of cognitive structure. Although it can be used to measure attitude, its primary impact has been in the area of persuasive communications (see Chapter 11), where the notion of latitudes has been used to explain differential effects of communications on subjects with the same attitude scores.

CONCLUSION

This chapter has reviewed some of the techniques used to measure beliefs, attitudes, and intentions. We have tried to show that all measurement ultimately rests on responses to single statements of belief or intention. The standard attitude-scaling methods use such responses to infer the person's location on a bipolar affective dimension vis-à-vis the object in question. Consistent with our definition of attitude, all traditional measurement techniques—whether direct, disguised, or physiological—attempt to arrive at a single attitude score which represents the person's evaluation of the attitude object. Standard attitude scaling is designed to select a set of beliefs or intentional statements which can be used to measure a person's attitude. Responses to these statements indicate strength of beliefs or intentions, and certain assumptions are made about associated attribute evaluations. Consistent with our expectancy-value model of attitude, the attitude score is based on these two pieces of information.

Measures of belief strength and attribute evaluations can also be used to arrive at various indices of cognitive structure. However, only few attempts have been made to develop and validate such measures of cognitive structure. Most

35. This response can be used as a measure of the person's attitude.

work on measurement techniques has concentrated on attitude, and relatively little attention has been paid to beliefs or intentions. Nevertheless, we have seen that techniques are available for the measurement of all three concepts. The next chapter will show that these techniques provide reliable and valid measures of beliefs, attitudes, and intentions.