

# THE AMERICAN JOURNAL OF SOCIOLOGY

VOLUME XXXIII

JANUARY 1928

NUMBER 4

## ATTITUDES CAN BE MEASURED<sup>1</sup>

L. L. THURSTONE  
University of Chicago

### ABSTRACT

The object of this study is to devise a method whereby the distribution of attitude of a group on a specified issue may be represented in the form of a frequency distribution. The base line represents ideally the whole range of opinions from those at one end who are most strongly in favor of the issue to those at the other end of the scale who are as strongly against it. Somewhere between the two extremes on the base line will be a neutral zone representing indifferent attitudes on the issue in question. The ordinates of the frequency distribution will represent the relative popularity of each attitude. This measurement problem has the limitation which is common to all measurement, namely, that one can measure only such attributes as can be represented on a linear continuum, such attributes as volume, price, length, area, excellence, beauty, and so on. For the present problem we are limited to those aspects of attitudes for which one can compare individuals by the "more and less" type of judgment. For example, we say understandingly that one man is more in favor of prohibition than another, more strongly in favor of the League of Nations than another, more militaristic than some other, more religious than another. The measurement is effected by the indorsement or rejection of statements of opinion. The opinions are allocated to different positions on the base line in accordance with the attitudes which they express. The ordinates of the frequency distribution are determined by the frequency with which each of the scaled opinions is indorsed. The center of the whole problem lies in the definition of a unit of measurement for the base line. The scale is so constructed that two opinions separated by a unit distance

<sup>1</sup> This is one of a series of papers by the staff of the Behavior Research Fund, Illinois Institute for Juvenile Research, Chicago. Series B No. 110.

The original manuscript for this paper has enjoyed a great deal of friendly criticism, some of which turns on matters of terminology and some on the assumptions which are here stated. In order to keep this paper within reasonable length, the description of the detailed psychophysical methods used and the construction of several attitude scales are reserved for separate publication. This paper concerns then only an outline of one solution to the problem of measuring attitude.

on the base line seem to differ as much in the attitude variable involved as any other two opinions on the scale which are also separated by a unit distance. This is the main idea of the present scale construction. The true allocation of an individual to a position on an attitude scale is an abstraction, just as the true length of a chalk line, or the true temperature of a room, or the true spelling ability of a child, is an abstraction. We estimate the true length of a line, the true temperature of a room, or the true spelling ability of a child, by means of various indices, and it is a commonplace in measurement that all indices do not agree exactly. In allocating an individual to a point on the attitude continuum we may use various indices, such as the opinions that he indorses, his overt acts, and his past history, and it is to be expected that discrepancies will appear as the true attitude of the individual is estimated by different indices. The present study is concerned with the allocation of individuals along an attitude continuum based on the opinions that they accept or reject.

---

### I. THE POSSIBILITY OF MEASURING ATTITUDE

The purpose of this paper is to discuss the problem of measuring attitudes and opinions and to offer a solution for it. The very fact that one offers a solution to a problem so complex as that of measuring differences of opinion or attitude on disputed social issues makes it evident from the start that the solution is more or less restricted in nature and that it applies only under certain assumptions that will, however, be described. In devising a method of measuring attitude I have tried to get along with the fewest possible restrictions because sometimes one is tempted to disregard so many factors that the original problem disappears. I trust that I shall not be accused of throwing out the baby with its bath.

In promising to measure attitudes I shall make several common-sense assumptions that will be stated here at the outset so that subsequent discussion may not be fogged by confusion regarding them. If the reader is unwilling to grant these assumptions, then I shall have nothing to offer him. If they are granted, we can proceed with some measuring methods that ought to yield interesting results.

It is necessary to state at the very outset just what we shall here mean by the terms "attitude" and "opinion." This is all the more necessary because the natural first impression about these two concepts is that they are not amenable to measurement in any real sense. It will be conceded at the outset that an attitude is a complex affair which cannot be wholly described by any single numerical index. For the problem of measurement this statement is analogous to the observation that an ordinary table is a complex

affair which cannot be wholly described by any single numerical index. So is a man such a complexity which cannot be wholly represented by a single index. Nevertheless we do not hesitate to say that we measure the table. The context usually implies what it is about the table that we propose to measure. We say without hesitation that we measure a man when we take some anthropometric measurements of him. The context may well imply without explicit declaration what aspect of the man we are measuring, his cephalic index, his height or weight or what not. Just in the same sense we shall say here that we are measuring attitudes. We shall state or imply by the context the aspect of people's attitudes that we are measuring. The point is that it is just as legitimate to say that we are measuring attitudes as it is to say that we are measuring tables or men.

The concept "attitude" will be used here to denote the sum total of a man's inclinations and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions about any specified topic. Thus a man's attitude about pacifism means here all that he feels and thinks about peace and war. It is admittedly a subjective and personal affair.

The concept "opinion" will here mean a verbal expression of attitude. If a man says that we made a mistake in entering the war against Germany, that statement will here be spoken of as an opinion. The term "opinion" will be restricted to verbal expression. But it is an expression of what? It expresses an attitude, supposedly. There should be no difficulty in understanding this use of the two terms. The verbal expression is the *opinion*. Our interpretation of the expressed opinion is that the man's *attitude* is pro-German. An opinion symbolizes an attitude.

Our next point concerns what it is that we want to measure. When a man says that we made a mistake in entering the war with Germany, the thing that interests us is not really the string of words as such or even the immediate meaning of the sentence merely as it stands, but rather the attitude of the speaker, the thoughts and feelings of the man about the United States, and the war, and Germany. It is the attitude that really interests us. The opinion has interest only in so far as we interpret it as a symbol of

attitude. It is therefore something about attitudes that we want to measure. We shall use opinions as the means for measuring attitudes.<sup>2</sup>

There comes to mind the uncertainty of using an opinion as an index of attitude. The man may be a liar. If he is not intentionally misrepresenting his real attitude on a disputed question, he may nevertheless modify the expression of it for reasons of courtesy, especially in those situations in which frank expression of attitude may not be well received. This has led to the suggestion that a man's action is a safer index of his attitude than what he says. But his actions may also be distortions of his attitude. A politician extends friendship and hospitality in overt action while hiding an attitude that he expresses more truthfully to an intimate friend. Neither his opinions nor his overt acts constitute in any sense an infallible guide to the subjective inclinations and preferences that constitute his attitude. Therefore we must remain content to use opinions, or other forms of action, merely as indices of attitude. It must be recognized that there is a discrepancy, some error of measurement as it were, between the opinion or overt action that we use as an index and the attitude that we infer from such an index.

But this discrepancy between the index and "truth" is universal. When you want to know the temperature of your room, you look at the thermometer and use its reading as an index of temperature just as though there were no error in the index and just as though there were a single temperature reading which is the "correct" one for the room. If it is desired to ascertain the volume of a glass paper weight, the volume is postulated as an attribute of the piece of glass, even though volume is an abstraction. The volume is measured indirectly by noting the dimensions of the glass or by

<sup>2</sup> Professor Faris, who has been kind enough to give considerable constructive criticism to the manuscript for this paper, has suggested that we may be measuring opinion but that we are certainly not measuring attitude. It is in part a terminological question which turns on the concept of attitude. If the concept of attitude as here defined is not acceptable, it may be advisable to change the terminology provided that a distinction is retained between (1) the objective index, which is here called the statement or opinion, and (2) the inferred subjective inclination of the person, which is here called the attitude variable.

immersing it in water to see how much water it displaces. These two procedures give two indices which might not agree exactly. In almost every situation involving measurement there is postulated an abstract continuum such as volume or temperature, and the allocation of the thing measured to that continuum is accomplished usually by indirect means through one or more indices. Truth is inferred only from the relative consistency of the several indices, since it is never directly known. We are dealing with the same type of situation in attempting to measure attitude. We must postulate an attitude variable which is like practically all other measurable attributes in the nature of an abstract continuum, and we must find one or more indices which will satisfy us to the extent that they are internally consistent.

In the present study we shall measure the subject's attitude as expressed by the acceptance or rejection of opinions. But we shall not thereby imply that he will necessarily *act* in accordance with the opinions that he has indorsed. Let this limitation be clear. The measurement of attitudes expressed by a man's opinions does not necessarily mean the prediction of what he will do. If his expressed opinions and his actions are inconsistent, that does not concern us now, because we are not setting out to predict overt conduct. We shall assume that it is of interest to know what people *say* that they believe even if their conduct turns out to be inconsistent with their professed opinions. Even if they are intentionally distorting their attitudes, we are measuring at least the attitude which they are trying to make people believe that they have.

We take for granted that people's attitudes are subject to change. When we have measured a man's attitude on any issue such as pacifism, we shall not declare such a measurement to be in any sense an enduring or constitutional constant. His attitude may change, of course, from one day to the next, and it is our task to measure such changes, whether they be due to unknown causes or to the presence of some known persuasive factor such as the reading of a discourse on the issue in question. However, such fluctuations may also be attributed in part to error in the measurements themselves. In order to isolate the errors of the measurement instrument from the actual fluctuation in attitude, we must calculate

the standard error of measurement of the scale itself, and this can be accomplished by methods already well known in mental measurement.

We shall assume that an attitude scale is used only in those situations in which one may reasonably expect people to tell the truth about their convictions or opinions. If a denominational school were to submit to its students a scale of attitudes about the church, one should hardly expect intelligent students to tell the truth about their convictions if they deviate from orthodox beliefs. At least, the findings could be challenged if the situation in which attitudes are expressed contains pressure or implied threat bearing directly on the attitude to be measured. Similarly, it would be difficult to discover attitudes on sex liberty by a written questionnaire, because of the well-nigh universal pressure to conceal such attitudes where they deviate from supposed conventions. It is assumed that attitude scales will be used only in those situations that offer a minimum of pressure on the attitude to be measured. Such situations are common enough.

All that we can do with an attitude scale is to measure the attitude actually expressed with the full realization that the subject may be consciously hiding his true attitude or that the social pressure of the situation has made him really believe what he expresses. This is a matter for interpretation. It is something probably worth while to measure an attitude expressed by opinions. It is another problem to interpret in each case the extent to which the subjects have expressed what they really believe. All that we can do is to minimize as far as possible the conditions that prevent our subjects from telling the truth, or else to adjust our interpretations accordingly.

When we discuss opinions, about prohibition for example, we quickly find that these opinions are multidimensional, that they cannot all be represented in a linear continuum. The various opinions cannot be completely described merely as "more" or "less." They scatter in many dimensions, but the very idea of measurement implies a linear continuum of some sort such as length, price, volume, weight, age. When the idea of measurement is applied to scholastic achievement, for example, it is necessary to force the

qualitative variations into a scholastic linear scale of some kind. We judge in a similar way such qualities as mechanical skill, the excellence of handwriting, and the amount of a man's education, as though these traits were strung out along a single scale, although they are of course in reality scattered in many dimensions. As a matter of fact, we get along quite well with the concept of a scale in describing traits even so qualitative as education, social and economic status, or beauty. A scale or linear continuum is implied when we say that a man has more education than another, or that a woman is more beautiful than another, even though, if pressed, we admit that perhaps the pair involved in each of the comparisons have little if anything in common. It is clear that the linear continuum which is implied in a "more and less" judgment may be conceptual, that it does not necessarily have the physical existence of a yardstick.

And so it is also with attitudes. We do not hesitate to compare them by the "more and less" type of judgment. We say about a man, for example, that he is more in favor of prohibition than some other, and the judgment conveys its meaning very well with the implication of a linear scale along which people or opinions might be allocated.

## 2. THE ATTITUDE VARIABLE

The first restriction on the problem of measuring attitudes is to specify an attitude variable and to limit the measurement to that. An example will make this clear. Let us consider the prohibition question and let us take as the attitude variable the degree of restriction that should be imposed on individual liberty in the consumption of alcohol. This degree of restriction can be thought of as a continuum ranging from complete and absolute freedom or license to equally complete and absolute restriction, and it would of course include neutral and indifferent attitudes.

In collecting samples from which to construct a scale we might ask a hundred individuals to write out their opinions about prohibition. Among these we might find one which expresses the belief that prohibition has increased the use of tobacco. Surely this is an opinion concerning prohibition, but it would not be at all serviceable for measuring the attitude variable just mentioned. Hence it



would be irrelevant. Another man might express the opinion that prohibition has eliminated an important source of government revenue. This is also an opinion concerning prohibition, but it would not belong to the particular attitude variable that we have set out to measure or scale. It is preferable to use an objective and experimental criterion for the elimination of opinions that do not belong on the specified continuum to be measured, and I believe that such a criterion is available.

This restriction on the problem of measuring attitudes is necessary in the very nature of measurement. It is taken for granted in all ordinary measurement, and it must be clear that it applies also to measurement in a field in which the multidimensional characteristics have not yet been so clearly isolated. For example, it would be almost ridiculous to call attention to the fact that a table cannot be measured unless one states or implies what it is about the table that is to be measured; its height, its cost, or beauty or degree of appropriateness or the length of time required to make it. The context usually makes this restriction on measurement. When the notion of measurement is applied to so complex a phenomenon as opinions and attitudes, we must here also restrict ourselves to some specified or implied continuum along which the measurement is to take place.

In specifying the attitude variable, the first requirement is that it should be so stated that one can speak of it in terms of "more" and "less," as, for example, when we compare the attitudes of people by saying that one of them is more pacifistic, more in favor of prohibition, more strongly in favor of capital punishment, or more religious than some other person.

Figure 1 represents an attitude variable, militarism-pacifism, with a neutral zone. A person who usually talks in favor of preparedness, for example, would be represented somewhere to the right of the neutral zone. A person who is more interested in disarmament would be represented somewhere to the left of the neutral zone. It is possible to conceive of a frequency distribution to represent the distribution of attitude in a specified group on the subject of pacifism-militarism.

Consider the ordinate of the frequency distribution at any



point on the base line. The point and its immediate vicinity represent for our purpose an attitude, and we want to know relatively how common that degree of feeling for or against pacifism may be in the group that is being studied. It is of secondary interest to know that a particular statement of opinion is indorsed by a certain proportion of that group. It is only to the extent that the opinion is representative of an attitude that it is useful for our purposes. Later we shall consider the possibility that a statement of opinion may be scaled as rather pacifistic and yet be indorsed by a person of very pronounced militaristic sympathies. To the extent that the

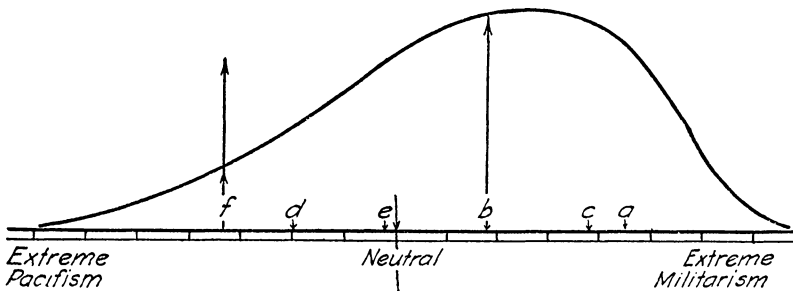


FIG. 1

statement is indorsed or rejected by factors other than the attitude-variable that it represents, to that extent the statement is useless for our purposes. We shall also consider an objective criterion for spotting such statements so that they may be eliminated from the scale. In our entire study we shall be dealing, then, with opinions, not primarily because of their cognitive content but rather because they serve as the carriers or symbols of the attitudes of the people who express or indorse these opinions.

There is some ambiguity in using the term attitude in the plural. An attitude is represented as a point on the attitude continuum. Consequently there is an infinite number of attitudes that might be represented along the attitude scale. In practice, however, we do not differentiate so finely. In fact, an attitude, practically speaking, is a certain narrow range or vicinity on the scale. When a frequency distribution is drawn for any continuous variable, such as stature, we classify the variable for descriptive pur-

poses into steps or class intervals. The attitude variable can also be divided into class intervals and the frequency counted in each class interval. When we speak of "an" attitude, we shall mean a point, or a vicinity, on the attitude continuum. Several attitudes will be considered not as a set of discrete entities, but as a series of class intervals along the attitude scale.

### 3. A FREQUENCY DISTRIBUTION OF ATTITUDES

The main argument so far has been to show that since in ordinary conversation we readily and understandably describe individuals as more and less pacifistic or more and less militaristic in attitude, we may frankly represent this linearity in the form of a unidimensional scale. This has been done in a diagrammatic way in Figure 1. We shall first describe our objective and then show how a rational unit of measurement may be adopted for the whole scale.

Let the base line of Figure 1 represent a continuous range of attitudes from extreme pacifism on the left to extreme militarism on the right.

If the various steps in such a scale were defined, it is clear that a person's attitude on militarism-pacifism could be represented by a point on that scale. The strength and direction of a particular individual's sympathies might be indicated by the point *a*, thus showing that he is rather militaristic in his opinions. Another individual might be represented at the point *b* to show that although he is slightly militaristic in his opinions, he is not so extreme about it as the person who is placed at the point *a*. A third person might be placed at the point *c* to show that he is quite militaristic and that the difference between *a* and *c* is very slight. A similar interpretation might be extended to any point on the continuous scale from extreme militarism to extreme pacifism, with a neutral or indifference zone between them.

A second characteristic might also be indicated graphically in terms of the scale, namely, the range of opinions that any particular individual is willing to indorse. It is of course not to be expected that every person will find only one single opinion on the whole scale that he is willing to indorse and that he will reject all

the others. As a matter of fact we should probably find ourselves willing to indorse a great many opinions on the scale that cover a certain range of it. It is conceivable, then, that a pacifistically inclined person would be willing to indorse all or most of the opinions in the range  $d$  to  $e$  and that he would reject as too extremely pacifistic most of the opinions to the left of  $d$ , and would also reject the whole range of militaristic opinions. His attitude would then be indicated by the average or mean of the range that he indorses, unless he cares to select a particular opinion which most nearly represents his own attitude. The same sort of reasoning may of course be extended to the whole range of the scale, so that we should have at least two, or possibly three, characteristics of each person designated in terms of the scale. These characteristics would be (1) the mean position that he occupies on the scale, (2) the range of opinions that he is willing to accept, and (3) that one opinion which he selects as the one which most nearly represents his own attitude on the issue at stake.

It should also be possible to describe a group of individuals by means of the scale. This type of description has been represented in a diagrammatic way by the frequency outline.

Any ordinate of the curve would represent the number of individuals, or the percentage of the whole group, that indorses the corresponding opinion. For example, the ordinate at  $b$  would represent the number of persons in the group who indorse the degree of militarism represented by the point  $b$  on the scale. A glance at the frequency curve shows that for the fictitious group of this diagram militaristic opinions are indorsed more frequently than the pacifistic ones. It is clear that the area of this frequency diagram would represent the total number of indorsements given by the group. The diagram can be arranged in several different ways that will be separately discussed. It is sufficient at this moment to realize that, given a valid scale of opinions, it would be possible to compare several different groups in their attitudes on a disputed question.

A second type of group comparison might be made by the range or spread that the frequency surfaces reveal. If one of the groups is represented by a frequency diagram of considerable

range or scatter, then that group would be more heterogeneous on the issue at stake than some other group whose frequency diagram of attitudes shows a smaller range or scatter. It goes without saying that the frequent assumption of a normal distribution in educational scale construction has absolutely no application here, because there is no reason whatever to assume that any group of people will be normally distributed in their opinions about anything.

It should be possible, then, to make four types of description by means of a scale of attitudes. These are (1) the average or mean attitude of a particular individual on the issue at stake, (2) the range of opinion that he is willing to accept or tolerate, (3) the relative popularity of each attitude of the scale for a designated group as shown by the frequency distribution for that group, and (4) the degree of homogeneity or heterogeneity in the attitudes of a designated group on the issue as shown by the spread or dispersion of its frequency distribution.

This constitutes our objective. The heart of the problem is in the unit of measurement for the base line, and it is to this aspect of the problem that we may now turn.

#### 4. A UNIT OF MEASUREMENT FOR ATTITUDES

The only way in which we can identify the different attitudes (points on the base line) is to use a set of opinions as landmarks, as it were, for the different parts or steps of the scale. The final scale will then consist of a series of statements of opinion, each of which is allocated to a particular point on the base line. If we start with enough statements, we may be able to select a list of twenty or thirty opinions so chosen that they represent an evenly graduated series of attitudes. The separation between successive statements of opinion would then be uniform, but the scale can be constructed with a series of opinions allocated on the base line even though their base line separations are not uniform. For the purpose of drawing frequency distributions it will be convenient, however, to have the statements so chosen that the steps between them are uniform throughout the whole range of the scale.

Consider the three statements *a*, *c*, and *d*, in Figure 1. The statements *c* and *a* are placed close together to indicate that they

are very similar, while statements  $c$  and  $d$  are spaced far apart to indicate that they are very different. We should expect two individuals scaled at  $c$  and  $a$  respectively to agree very well in discussing pacifism and militarism. On the other hand, we should expect to be able to tell the difference quite readily between the opinions of a person at  $d$  and another person at  $c$ . The scale separations of the opinions must agree with our impressions of them.

In order to ascertain how far apart the statements should be on the final scale, we submit them to a group of several hundred people who are asked to arrange the statements in order from the most pacifistic to the most militaristic. We do not ask them for their own opinions. That is another matter entirely. We are now concerned with the construction of a scale with a valid unit of measurement. There may be a hundred statements in the original list, and the several hundred persons are asked merely to arrange the statements in rank order according to the designated attitude variable. It is then possible to ascertain the proportion of the readers who consider statement  $a$  to be more militaristic than statement  $c$ . If the two statements represent very similar attitudes we should not expect to find perfect agreement in the rank order of statements  $a$  and  $c$ . If they are identical in attitude, there will be about 50 per cent of the readers who say that statement  $a$  is more militaristic than statement  $c$ , while the remaining 50 per cent of the readers will say that statement  $c$  is more militaristic than statement  $a$ . It is possible to use the proportion of readers or judges who agree about the rank order of any two statements as a basis for actual measurement.

If 90 per cent of the judges or readers say that statement  $a$  is more militaristic than statement  $b$  ( $p_{a>b} = .90$ ) and if only 60 per cent of the readers say that statement  $a$  is more militaristic than statement  $c$  ( $p_{a>c} = .60$ ) then clearly the scale separation ( $a - c$ ) is shorter than the scale separation ( $a - b$ ). The psychological scale separation between any two stimuli can be measured in terms of a law of comparative judgment which the writer has recently formulated.<sup>3</sup>

<sup>3</sup> For a more detailed discussion of this law see my article "The Law of Comparative Judgment," *Psych. Rev.* (July, 1927). For the logic of the psychological S-scale see "Psychophysical Analysis," *Amer. J. Psych.* (July, 1927).

The detailed methods of handling the data will be published in connection with the construction of each particular scale. The practical outcome of this procedure is a series of statements of opinions allocated along the base line of Figure 1. The interpretation of the base-line distances is that the apparent difference between any two opinions will be equal to the apparent difference between any other two opinions which are spaced equally far apart on the scale. In other words, the shift in opinion represented by a unit distance on the base line seems to most people the same as the shift in opinion represented by a unit distance at any other part of the scale. Two individuals who are separated by any given distance on the scale *seem* to differ in their attitudes as much as any other two individuals with the same scale separation. In this sense we have a truly rational base line, and the frequency diagrams erected on such a base line are capable of legitimate interpretation as frequency surfaces.

In contrast with such a rational base line or scale is the simpler procedure of merely listing ten to twenty opinions, arranging them in rank order by a few readers, and then merely counting the number of indorsements for each statement. That can of course be done provided that the resulting diagram be not interpreted as a frequency distribution of attitude. If so interpreted the diagram can be made to take any shape we please by merely adding new statements or eliminating some of them, arranging the resulting list in a rough rank order evenly spaced on the base line. Allport's diagrams of opinions<sup>4</sup> are not in any sense frequency distributions. They should be considered as bar-diagrams in which are shown the frequency with which each of a number of statements is indorsed. Our principal contribution here is an improvement on Allport's procedure. He is virtually dealing with rank orders, which we are here trying to change into measurement by a rational unit of measurement. Allport's pioneering studies in this field should be read by every investigator of this problem. My own interest in the pos-

<sup>4</sup> Floyd H. Allport, and D. A. Hartman, "Measurement and Motivation of Atypical Opinion in a Certain Group," *American Political Science Review*, XIX (1925), 735-60.

sibility of measuring attitude by means of opinions was started by Allport's paper, and the present study is primarily a refinement of his statistical methods.

The unit of measurement for the scale of attitudes is the standard deviation of the dispersion projected on the psychophysical scale of attitudes by a statement of opinion, chosen as a standard. It is a matter of indifference which statement is chosen as a standard, since the scales produced by different standard statements will have proportional scale values. This mental unit of measurement is roughly comparable to, but not identical with, the so-called "just noticeable difference" in psychophysical measurement.

A diagram such as Figure 1 can be constructed in either of at least two different ways. The area of the frequency surface may be made to represent the total number of votes or indorsements by a group of people, or the area may be made to represent the total number of individuals in the group studied. Allport's diagrams would be made by the latter principle if they were constructed on a rational base line so that a legitimate area might be measured. Each subject was asked to select that one statement in the list most representative of his own attitude. Hence at least the sum of the ordinates will equal the total number of persons in the group. I have chosen as preferable the procedure of asking each subject to indorse all the statements with which he agrees. Since we have a rational base line, we may make a legitimate interpretation of the area of the surface as the total number of indorsements made by the group. This procedure has the advantage that we may ascertain the range of opinion which is acceptable to each person, a trait which has considerable interest and which cannot be ascertained by asking the subject to indorse only one of the statements in the list. The ordinates of the frequency diagram can be plotted as proportions of the whole group. They will then be interpreted as the probability that the given statement will be indorsed by a member of the group. In other words, the frequency diagram is descriptive of the distribution of attitude in the whole group, and at each point on the base line we want an ordinate to represent the relative popularity of that attitude.



## 5. THE CONSTRUCTION OF AN ATTITUDE SCALE

At the present time three scales for the measurement of opinion are being constructed by the principles here described.<sup>5</sup> These three scales are planned to measure attitudes on three different variables, namely, pacifism-militarism, prohibition, and attitude toward the church. All three of these scales are being constructed first by a procedure somewhat less laborious than the direct application of the law of comparative judgment, and if consistent results are obtained the method will be retained for other scales.

The method is as follows. Several groups of people are asked to write out their opinions on the issue in question, and the literature is searched for suitable brief statements that may serve the purposes of the scale. By editing such material a list of from 100 to 150 statements is prepared expressive of attitudes covering as far as possible all gradations from one end of the scale to the other. It is sometimes necessary to give special attention to the neutral statements. If a random collection of statements of opinion should fail to produce neutral statements, there is some danger that the scale will break in two parts. The whole range of attitudes must be fairly well covered, as far as one can tell by preliminary inspection, in order to insure that there will be overlapping in the rank orders of different readers throughout the scale.

In making the initial list of statements several practical criteria are applied in the first editing work. Some of the important criteria are as follows: (1) the statements should be as brief as possible so as not to fatigue the subjects who are asked to read the whole list. (2) The statements should be such that they can be indorsed or rejected in accordance with their agreement or disagreement with the attitude of the reader. Some statements in a random sample will be so phrased that the reader can express no definite indorsement or rejection of them. (3) Every statement should be such that acceptance or rejection of the statement does indicate something regarding the reader's attitude about the issue

<sup>5</sup> Three attitude scales are now in course of preparation by Mr. E. J. Chave, of the Divinity School, University of Chicago, on attitudes toward the church; by Mrs. Hattie Smith on attitudes about prohibition; and by Mr. Daniel Droba on attitudes about pacifism-militarism. The latter two will be published as Doctor's dissertations.

in question. If, for example, the statement is made that war is an incentive to inventive genius, the acceptance or rejection of it really does not say anything regarding the reader's pacifistic or militaristic tendencies. He may regard the statement as an unquestioned fact and simply indorse it as a fact, in which case his answer has not revealed anything concerning his own attitude on the issue in question. However, only the conspicuous examples of this effect should be eliminated by inspection, because an objective criterion is available for detecting such statements so that their elimination from the scale will be automatic. Personal judgment should be minimized as far as possible in this type of work. (4) Double-barreled statements should be avoided except possibly as examples of neutrality when better neutral statements do not seem to be readily available. Double-barreled statements tend to have a high ambiguity. (5) One must insure that at least a fair majority of the statements really belong on the attitude variable that is to be measured. If a small number of irrelevant statements should be either intentionally or unintentionally left in the series, they will be automatically eliminated by an objective criterion, but the criterion will not be successful unless the majority of the statements are clearly a part of the stipulated variable.

When the original list has been edited with these factors in mind, there will be perhaps 80 to 100 statements to be actually scaled. These statements are then mimeographed on small cards, one statement on each card. Two or three hundred subjects are asked to arrange the statements in eleven piles ranging from opinions most strongly affirmative to those most strongly negative. The detailed instructions will be published with the description of the separate scales. The task is essentially to sort out the small cards into eleven piles so that they *seem* to be fairly evenly spaced or graded. Only the two ends and the middle pile are labelled. The middle pile is indicated for neutral opinions. The reader must decide for each statement which of five subjective degrees of affirmation or five subjective degrees of negation is implied in the statement or whether it is a neutral opinion.

When such sorting has been completed by two or three hundred readers, a diagram like Figure 2 is prepared. We shall discuss

it with the scale for pacifism-militarism as an example. On the base line of this diagram are represented the eleven apparently equal steps of the attitude variable. The neutral interval is the interval 5 to 6, the most pacifistic interval from 0 to 1, and the most militaristic interval from 10 to 11. This diagram is fictitious and is drawn to show the principle involved. Curve *A* is drawn to show the manner in which one of the statements might be classified by the three hundred readers. It is not classified by anyone below the value of 3, half of the readers classify it below the value 6, and all of them classify it below the value 9. The scale value of the

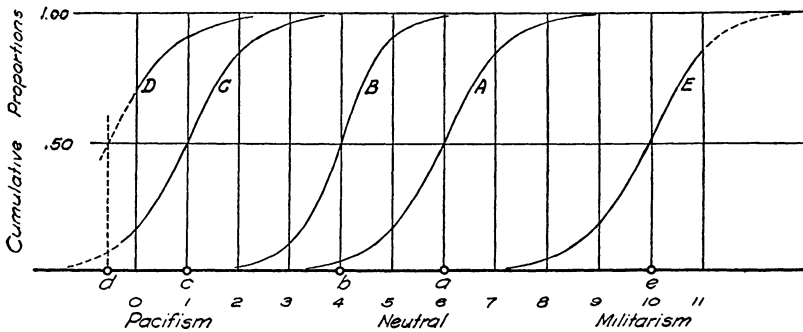


FIG. 2

statement is that scale value below which just one half of the readers place it. In other words, the scale value assigned to the statement is so chosen that one half of the readers consider it more militaristic and one half of them consider it less militaristic than the scale value assigned. The numerical calculation of the scale value is similar to the calculation of the limen by the phi-gamma hypothesis in psychophysical measurement.

It will be found that some of the statements toward the ends of the scale do not give complete ogive curves. Thus statement *C* is incomplete in the fictitious diagram. It behaves as though it needed space beyond the arbitrary limits of the scale in order to be completed. Its scale value may, however, be determined as that scale value at which the phi-gamma curve through the experimental proportions crosses the 50 per cent level, which is at *c*. Still other statements may be found, such as *D*, which have scale

values beyond the arbitrary range of the scale. These may be assigned scale values by the same process, though less accurately.

The situation is different at the other end of the scale. The statement *E* has a scale value at *e*, but owing to the limit of the scale at the point 11 the experimental proportion will be 1.00 at that point. If the scale continued beyond the point 11 the proportions would continue to rise gradually as indicated by the dotted line. The experimental proportions are all necessarily 1.00 for the scale value 11, and hence these final proportions must be ignored in fitting the phi-gamma curves and in the location of the scale values of the statements.

#### 6. THE VALIDITY OF THE SCALE

a) *The scale must transcend the group measured.*—One crucial experimental test must be applied to our method of measuring attitudes before it can be accepted as valid. A measuring instrument must not be seriously affected in its measuring function by the object of measurement. To the extent that its measuring function is so affected, the validity of the instrument is impaired or limited. If a yardstick measured differently because of the fact that it was a rug, a picture, or a piece of paper that was being measured, then to that extent the trustworthiness of that yardstick as a measuring device would be impaired. Within the range of objects for which the measuring instrument is intended, its function must be independent of the object of measurement.

We must ascertain similarly the range of applicability of our method of measuring attitude. It will be noticed that the *construction* and the *application* of a scale for measuring attitude are two different tasks. If the scale is to be regarded as valid, the scale values of the statements should not be affected by the opinions of the people who help to construct it. This may turn out to be a severe test in practice, but the scaling method must stand such a test before it can be accepted as being more than a description of the people who construct the scale. At any rate, to the extent that the present method of scale construction is affected by the opinions of the readers who help to sort out the original statements into a

scale, to that extent the validity or universality of the scale may be challenged.

Until experimental evidence may be forthcoming on this point, we shall make the assumption that the scale values of the statements are independent of the attitude distribution of the readers who sort the statements. The assumption is, in other words, that two statements on a prohibition scale will be as easy or as difficult to discriminate for people who are "wet" as for those who are "dry." Given two adjacent statements from such a scale, we assume that the proportion of "wets" who say that statement *a* is wetter than statement *b* will be substantially the same as the corresponding proportion for the same statements obtained from a group of "drys." Restating the assumption in still another way, we are saying that it is just as difficult for a strong militarist as it is for a strong pacifist to tell which of two statements is the more militaristic in attitude. If, say, 85 per cent of the militarists declare statement *A* to be more militaristic than statement *B*, then, according to our assumption, substantially the same proportion of pacifists would make the same judgment. If this assumption is correct, then the scale is an instrument independent of the attitude which it is itself intended to measure.

The experimental test for this assumption consists merely in constructing two scales for the same issue with the same set of statements. One of these scales will be constructed on the returns from several hundred readers of militaristic sympathies and the other scale will be constructed with the same statements on the returns from several hundred pacifists. If the scale values of the statement are practically the same in the two scales, then the validity of the method will be pretty well established.<sup>6</sup> It will still be necessary to use opinion scales with some discretion. Queer results might be obtained with the prohibition scale, for example, if it were presented in a country in which prohibition is not an issue.

*b) An objective criterion of ambiguity.*—Inspection of the curves in Figure 2 reveals that some of the statements of the fictitious diagram are more ambiguous than others. The degree of

<sup>6</sup> The neutrality point would not necessarily be represented by the same statement for both militarists and pacifists, but the scale separations between all pairs of statements should be practically the same for the two conditions of standardization.

ambiguity in a statement is immediately apparent, and in fact it can be definitely measured. The ambiguity of a statement is the standard deviation of the best fitting phi-gamma curve through the observed proportions. The steeper the curve, the smaller is the range of the scale over which it was classified by the readers and the clearer and more precise is the statement. The more gentle the slope of the curve, the more ambiguous is the statement. Thus of the two statements *A* and *B* in the fictitious diagram the statement *A* is the more ambiguous.

In case it should be found that the phi-gamma function does not well describe the curves of proportions in Figure 2, the degree of ambiguity may be measured without postulating that the proportions follow the phi-gamma function when plotted on the attitude scale. A simple method of measuring ambiguity would then be to determine the scale distance between the scale value at which the curve of proportions has an ordinate of .25 and the scale value at which the same curve has an ordinate of .75. The scale value of the statement itself can also be defined, without assuming the phi-gamma function, as that scale value at which the curve of proportions reaches .50. If no actual proportion is found at that value, the scale value of the statement may be interpolated between the experimental proportions immediately above and below the .50 level. In scaling the statements whose scale values fall outside the ten divisions of the scale, it will be necessary to make some assumption regarding the nature of the curve, and it will probably be found that for most situations the phi-gamma function will constitute a fairly close approximation to truth.

*c) An objective criterion of irrelevance.*—Before a selection of statements can be made for the final scale, still another criterion must be applied. It is an objective criterion of irrelevance. Referring again to Figure 1, let us consider two statements that have identical scale values at the point *f*. Suppose, further, that these two statements are submitted to the group of readers represented in the fictitious diagram of Figure 1. It is quite conceivable, and it actually does happen, that one of these statements will be indorsed quite frequently while the other statement is only seldom indorsed in spite of the fact that they are properly scaled as implying the same degree of pacifism or militarism. The conclusion is

then inevitable that the indorsement that a reader gives to these statements is determined only partly by the degree of pacifism implied and partly by other implied meanings which may or may not be related to the attitude variable under consideration. Now it is of course necessary to select for the final attitude scale those statements which are indorsed or rejected primarily on account of the degree of pacifism-militarism which is implied in them and to eliminate those statements which are frequently accepted or rejected on account of other more or less subtle and irrelevant meanings.

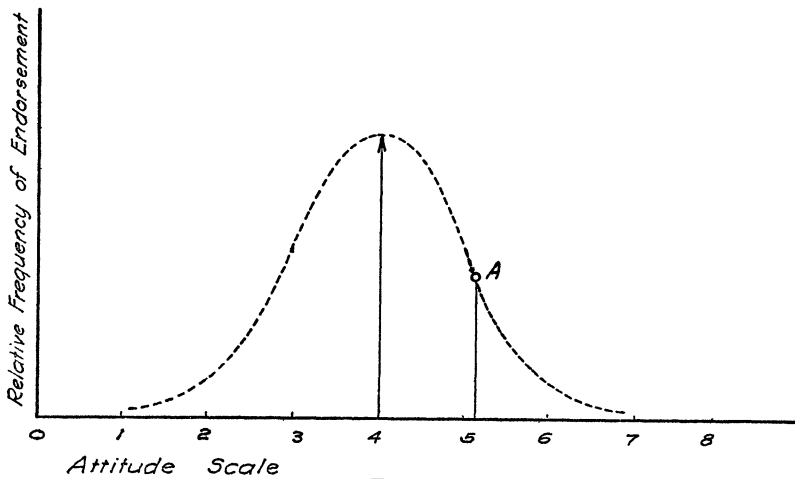


FIG. 3

An objective criterion for accomplishing this elimination automatically and without introducing the personal equation of the investigator is available. It is essentially as follows: Assume that the whole list of about one hundred statements has been submitted to several hundred readers for actual voting. These need not be the same readers who sorted the statements for the purpose of scaling. Let these readers be asked to mark with a plus sign every statement which they indorse and to reject with a minus sign every statement not to their liking.

If we want to investigate the degree of irrelevance of any particular statement which, for example, might have a scale value of 4.0 in Figure 3, we should first of all determine how many readers indorsed it. We find, for example, that 260 readers indorsed it. Let



this total be represented on the diagram as 100 per cent, and erect such an ordinate at the scale value of this statement. We may now ascertain the proportion of these 260 readers who *also* indorsed each other statement. If the readers indorse and reject the statements largely on the basis of the degree of pacifism-militarism implied, then those readers who indorse statements in the vicinity of 4.0 on the scale will not often indorse statements that are very far away from that point on the scale. Very few of them should indorse a statement which is scaled at the point 8.0, for example. If a large proportion of the 260 readers who indorse the basic statement scaled at 4.0 should also indorse a statement scaled at the point 8.0, then we should infer that their voting on these two statements has been influenced by factors other than the degree of pacifism that is implied in the statements. We can represent this type of analysis graphically.

Every one of these other statements will be represented by a point on this diagram. Its  $x$ -value will be the scale value of the statement, and its  $y$ -value will be the proportion of the 260 readers who indorsed it. Thus, if out of the 260 readers who indorsed the basic statement there were 130 who also indorsed statement No. 14, which has a scale value of, say, 5.0, then statement No. 14 will be represented at the point  $A$  on Figure 3.

If the basic statement, the degree of irrelevance of which is represented in Figure 3, is an ideal statement, one which people will accept or reject primarily because of the attitude on pacifism which it portrays, then we should expect the one hundred statements to be represented by as many points hovering more or less about the dotted line of Figure 3. The diagram may of course be more contracted or spread out, but the general appearance of the plot should be that of Figure 3. If, on the other hand, the basic statement has implications that lead to acceptance or rejection quite apart from the degree of pacifism which it conveys, then the proportion of the indorsements of the statements should not be a continuous function of their scale distance from the basic statement. The one hundred points might then scatter widely over the diagram. This inspectional criterion of irrelevance is objective and it can probably be translated into a more definite algebraic form so as to eliminate entirely the personal equation of the investigator.

Two other objective criteria of irrelevance have been devised. They will be described in connection with the attitude scales now being constructed.

#### 7. SUMMARY OF THE SCALING METHOD

The selection of the statements for the final scale should now be possible. A shorter list of twenty or thirty statements should be selected for actual use. We have described three criteria by which to select the statements for the final scale. These criteria are:

1. The statements in the final scale should be so selected that they constitute as nearly as possible an evenly graduated series of scale values.

2. By the objective criterion of ambiguity it is possible to eliminate those statements which project too great a dispersion on the attitude continuum. The objective measure of ambiguity is the standard deviation of the best fitting phi-gamma curve as illustrated in Figure 2.

3. By the objective criteria of irrelevance it is possible to eliminate those statements which are accepted or rejected largely by factors other than the degree of the attitude-variable which they portray. One of these criteria is illustrated in Figure 3.

The steps in the construction of an attitude scale may be summarized briefly as follows:

1. Specification of the attitude variable to be measured.

2. Collection of a wide variety of opinions relating to the specified attitude variable.

3. Editing this material for a list of about one hundred brief statements of opinion.

4. Sorting the statements into an imaginary scale representing the attitude variable. This should be done by about three hundred readers.

5. Calculation of the scale value of each statement.

6. Elimination of some statements by the criterion of ambiguity.

7. Elimination of some statements by the criteria of irrelevance.

8. Selection of a shorter list of about twenty statements evenly graduated along the scale.

## 8. MEASUREMENT WITH AN ATTITUDE SCALE

The practical application of the present measurement technique consists in presenting the final list of about twenty-five statements of opinion to the group to be studied with the request that they check with plus signs all the statements with which they agree and with minus signs all the statements with which they disagree. The score for each person is the average scale value of all the statements that he has indorsed. In order that the scale be effective toward the extremes, it is advisable that the statements in the scale be extended in both directions considerably beyond the attitudes which will ever be encountered as mean values for individuals. When the score has been determined for each person by the simple summation just indicated, a frequency distribution can be plotted for the attitudes of any specified group.

The reliability of the scale can be ascertained by preparing two parallel forms from the same material and by presenting both forms to the same individuals. The correlation between the two scores obtained for each person in a group will then indicate the reliability of the scale. Since the heterogeneity of the group affects the reliability coefficient, it is necessary to specify the standard deviation of the scores of the group on which the reliability coefficient is determined. The standard error of an individual score can also be calculated by an analogous procedure.

The unit of measurement in the scale when constructed by the procedure here outlined is not the standard discriminial error projected by a single statement on the psychological continuum. Such a unit of measurement can be obtained by the direct application of the law of comparative judgment, but it is considerably more laborious than the method here described. The unit in the present scale is a more arbitrary one, namely, one-tenth of the range on the psychological continuum which covers the span from what the readers regard as extreme affirmation to extreme negation in the particular list of statements with which we start. Of course the scale values can be determined with reliability to fractional parts of this unit. It is hoped that this unit may be shown experimentally to be proportional to a more precise and more universal unit of measurement such as the standard discriminial error of a single statement of opinion.

It is legitimate to determine a central tendency for the frequency distribution of attitudes in a group. Several groups of individuals may then be compared as regards the means of their respective frequency distributions of attitudes. The differences between the means of several such distributions may be directly compared because of the fact that a rational base line has been established. Such comparisons are not possible when attitudes are ascertained merely by counting the number of indorsements to separate statements whose scale differences have not been measured.

In addition to specifying the mean attitude of each of several groups, it is also possible to measure their relative heterogeneity with regard to the issue in question. Thus it will be possible, by means of our present measurement methods, to discover for example that one group is 1.6 more heterogeneous in its attitudes about prohibition than some other group. The heterogeneity of a group is indicated perhaps best by the standard deviation of the scale values of all the opinions that have been indorsed by the group as a whole rather than by the standard deviation of the distribution of individual mean scores. Perhaps different terms should be adopted for these two types of measurement.

The tolerance which a person reveals on any particular issue is also subject to quantitative measurement. It is the standard deviation of the scale values of the statements that he indorses. The maximum possible tolerance is of course complete indifference, in which all of the statements are indorsed throughout the whole range of the scale.

If it is desired to know which of two forms of appeal is the more effective on any particular issue, this can be determined by using the scale before and after the appeal. The difference between the individual scores, before and after, can be tabulated and the average shift in attitude following any specified form of appeal can be measured.

The essential characteristic of the present measurement method is the scale of evenly graduated opinions so arranged that equal steps or intervals on the scale *seem* to most people to represent equally noticeable shifts in attitude.