PSYCHOMETRIKA—VOL. 21, NO. 4
DECEMBER, 1956

THE SCALE GRID: SOME INTERRELATIONS OF DATA MODELS*

CLYDE H. COOMBS†
UNIVERSITY OF MICHIGAN

Perhaps an appropriate subtitle for this paper would be "Some Speculations on the Interrelations of Psychological Methodologies." The *Scale Grid* is a name I have given to a model which presumes to define the underlying continuities between such diverse areas as psychophysics, objective testing, attitude studies including questionnaire and interview techniques, learning experiments, rating scale methods, essay examinations, and projective instruments. The intent of the Scale Grid is to make explicit the fundamental similarities and differences of the methodologies among these various areas of psychological research. The increasing abundance of models and methods in all of these areas, with their associated nomenclature and specialized vocabularies, makes a unification of them increasingly desirable.

In some of these areas of psychological research, serious and intensive efforts have been made to construct models on a genotypic level to explain and predict manifest behavior. One thinks here, for example, of the area of signal detection in psychophysics and of objective test performance. In other areas the models are less explicit and tend to be on a literary level. However, even in these latter areas one may look at the methods of analyzing data. Because there is always a model, at least implied, some of the elements of the models are evident. One may look at this universe of models, explicit and implicit, abstract certain universal elements, and try to characterize them.

Our starting point will be to determine the primitive datum in psychology. What we want is an abstract definition which will hold for every type of psychological observation. Let us begin by taking some examples, seeing what the basic abstract datum is in each case, and then formulating a general definition. In some psychophysical experiments, for example, individuals judge which one of several stimuli is the greatest. In view of what the experimenter subsequently does with the data, it is evident that he thinks of each stimulus as a point; the judgment of the individual is inter-

*Presidential Address to the Psychometric Society, September 4, 1956.

†I wish to express my deep appreciation of the courtesies, assistance, and critical audience furnished me in the preparation of this paper by Professor H. C. J. Duijker and Professor A. DeGroot of the University of Amsterdam, while I was a Fulbright Research Fellow there in 1955-56.

preted as an order relation on these points. Over many replications, the model may deal with a distribution of points for each stimulus and be a probability or actuarial model, but this is not the type of distinction which is at all important to us now. Whether the models are deterministic or actuarial is not relevant to the fundamental distinctions between methodologies I want to make. In fact, I want only to make the point now that in some types of experiments the manifest behavior is interpreted as an order relation between a pair of points, both of which are identified with stimuli.

A different case arises, however, when an individual takes a mental test and passes some items and fails others. The use to which the behavior is put also suggests that it is being interpreted as a relation on a pair of points, but here one point is identified with an item and the other point with the individual. The point associated with the individual represents a measure of his ability; the point associated with the item represents its difficulty. The behavior of the individual in passing or failing the item is interpreted as an order relation on this pair of points. I am not here concerned with the numerical scores on a test or even how a theory arrives at such a score from the basic datum. These are differences on a higher level, and I am here concerned only with differences on most the primitive quantitative level.

When an individual is given an attitude scale and asked to indicate which items he will indorse, again the behavior is interpreted as a relation on a pair of points—one a stimulus, the other an individual. Here the relation is on a psychological distance between the point associated with an individual and the point associated with the stimulus. If the point associated with the stimulus is "near," in a sense defined by the model, the point associated with the individual, he indorses the item, otherwise not. So the behavior is interpreted as indicating whether the distance between two points is greater or less than a certain amount.

When an individual is asked to place a stimulus on rating scale, again the behavior is interpreted as a distance between the point associated with the stimulus and a point associated with a response category which is just another stimulus. Consider, for example, an individual who is asked to rate a stimulus, say a picture, as to whether it is superior, good, or poor. The picture is conceived of as being a point on the scale for this individual; the three points on the rating scale, superior, good, and poor, are also three stimulus points. From the latter points the individual selects the one nearest the point corresponding to the picture. So we see that rating scale behavior is interpreted as a relation between points. The same analysis holds if the rating scale is an ordered set of numbers or the real line. In fact counting is also rating scale behavior. If we ask an individual how many students there are in a class it does not matter whether he guesses or counts as far as the basic datum is concerned. The response, e.g., "35," is interpreted as

a relation between one stimulus, the size of the class as perceived, and another stimulus, a real number.

When an individual is asked whether he observed a light increment or not, the behavior is interpreted as a relation between a point identified with the individual, a threshold, and a point identified with the stimulus, the magnitude of the increment. For a final example, consider an individual asked to judge which of two pairs of stimuli is more similar. Here the behavior is usually interpreted as a relation between two distances; if each distance is interpreted as a point, then behavior implies a relation on a pair of points.

All of these examples illustrate one important fact: behavior is made into data by interpreting it as an order relation between points or a relation on distances—both may more generally be regarded as a relation on a pair of points. An important distinction must be drawn between behavior and data. A datum is defined in this paper as a relation between points. That this is not a new idea is evident from a half-page note by Madison Bentley (1) in which he speaks of Stumpf, Wundt, Ebinghaus, Mach, G. E. Müller, and others who took the view that psychological measurement is a distance measurement, which is just a special case of a relation between pairs.

We have been speaking here as if there were just a single distinction between behavior and data. Actually a threefold distinction should properly be made. We may use the term *behavior* to refer to anything observable about the individual, *raw data* to refer to that which is selected for analysis, and *data* to refer to the interpretation of the *raw data* as an abstract relation between points. The first step in going from behavior to raw data, deciding what to observe, is a many-faceted problem which lies outside the scope of this paper. We are here concerned exclusively with the raw data and how it is interpreted as data in the sense defined above. I shall pursue this distinction between the *raw data* and the *data*, illustrating it in detail shortly.

In principle one could put any data in a matrix as follows: If the data were a relation on a pair of stimuli then a square matrix with rows and corresponding columns identified with stimuli (cf. Figure 1) could nicely accommodate the data. Each cell would contain an entry indicating the relation between that corresponding pair of stimuli. Another experiment, where one member of the pair of points was identified with an individual and the other with a stimulus would require that the matrix of Figure 1 be expanded as in Figure 2. Thus an experiment in which one point of a pair of points was identified as a stimulus and the other as an individual would be entered in the left portion; another experiment in which the behavior was interpreted as a relation on a pair of points, both of which were identified as stimuli, would be entered in the right portion.

Now if we go a little further and consider an experiment where the members of the pair of points are both identified with individuals, the

matrix of Figure 2 becomes as in Figure 3. According to this figure, the behavior observed in some experiments is interpreted as a relation on pairs of points in which both points may be identified with stimuli, both points may be identified with individuals, or one point with a stimulus and one point with an individual. When the data are relations between stimulus

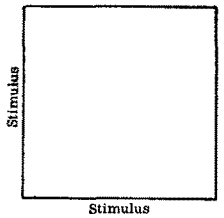

FIGURE 1

A Data Matrix

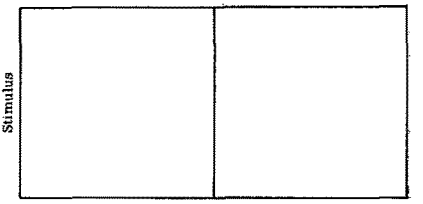

FIGURE 2

A Data Matrix

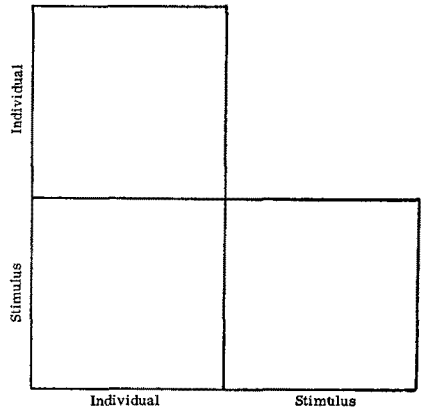

FIGURE 3

A Data Matrix

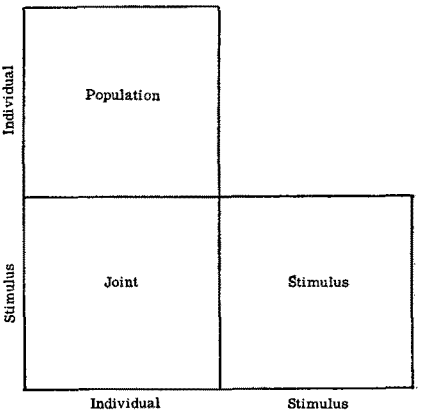

FIGURE 4

A Data Matrix

points, the analysis yields the order of these points on a psychological attribute or the location of these points in a multidimensional space; the individuals who made the judgments or responses are not located as points in the space. So we might call such a space a *Stimulus* space. Correspondingly, we can talk about a space in which only individuals are located as a *Population* space, and a space with both stimuli and people as a *Joint* space. We have the kinds of spaces in which psychological data are analyzed classified in Figure 4.

We now have the beginning of what I call the *Scale Grid*. We could move in either of two directions: developing the model and putting a little more meat on the skeleton, or constructing a psychological interpreta-

tion in order to bring out the implications of the grid. I find it rather difficult not to do both as they are so closely interdependent. So I shall first develop some of the theoretical ideas which will be most related to the interpretations which will follow. We shall consider some typical experiments which are mapped into Joint spaces and some which are mapped into Stimulus spaces; then we shall see what the difference is between them. The characterization of this difference will constitute one dimension of the Scale Grid.

Model Underlying The Scale Grid.

Behavior on a mental test is interpreted as a relation on a pair of points, one of which is associated with the individual and the other associated with the item or stimulus. Such behavior is mapped into data which, when analyzed, yields a Joint space with both stimuli and individuals located in it. What is the primitive operation here? The test item was conceived of as having a certain difficulty and the individual was, in effect, asked to compare his ability level with the difficulty level of the item. In a psychophysical study of the thresholds of individuals, the same is true, e.g., the individual is asked whether he perceives the stimulus, *yes* or *no*, and the behavior is interpreted as a relation on a pair of points, one of which is associated with the threshold of the individual, the other with a stimulus magnitude.

On the other hand, in a psychological study designed to measure heaviness of weights, length of lines, brightness of lights, or what have you, what is the primitive operation? The stimuli are conceived of as points on an attribute continuum, and the individual is asked which of the stimuli is greater. The behavior is interpreted as data on pairs of points both of which are associated with stimuli. Analysis of the data locates the stimuli on a scale, but no attempt is made to locate the individual as a point on the scale. The result is a Stimulus space.

Suppose I have some attitude statements about the church. I want to scale the items and then measure people's attitudes with them. The first step is to scale the items, so we ask individuals to evaluate the items as to which is more pro-church. A data matrix is constructed which is analyzed to yield a Stimulus space, say a one-dimensional scale, with the items located as points on a continuum. We note that the individuals were asked to evaluate the items with respect to where the items were on the continuum and not with respect to any *point* on this continuum which corresponded to the individual. Having scaled the items, we turn around and ask the individual which items he indorses. When we analyze these data, we end up with the individuals located on the same continuum, because this time the experimenter gave the individuals a different task to perform. The individuals were each asked to evaluate the items with respect to some point on the continuum corresponding to his own attitude toward the church, so now we are in a Joint space.

I could go on with examples from conditioning experiments or studies in perception, etc., but will not take the time for it. The important thing is to see the essential difference between behavior which is mapped into a Joint space and behavior which is mapped into a Stimulus space. In all the experiments, there are always both individuals and stimuli—what is it that determines whether an experimenter maps his experiment into a Stimulus space or into a Joint space?

If you go back and look at experiments with this question in mind, it becomes obvious that the experimenter puts his experment in a Joint space or a Stimulus space according to whether he regards the individual as having evaluated the stimuli with respect to a point corresponding to himself, the individual, or whether the individual evaluated the stimuli with respect to an attribute. I have called these two kinds of tasks, task A and task B, respectively. Task A may be described as evaluative, having to do with the relation of stimuli to the individual himself. Task B may be described as substantive, having to do with the nature of the stimuli per se.

I formalized this distinction between task A and task B in the following manner. In all experiments, both the individuals and the stimuli are points in a space, but in task A the points associated with the individuals are independent of the points associated with the stimuli. Whereas in task B, where the individual is evaluating stimuli with respect to an attribute, the point associated with an individual is completely dependent upon the points associated with the stimuli he is evaluating. I will not try to go further with this now, but essentially what we have is one dimension of the Scale Grid with just its two extremes represented—complete independence of the individual's point from those of the stimuli and complete dependence of his point on those of the stimuli, corresponding to task A and task B, respectively.

I have taken a good deal of time just to give an intuitive notion of one dimension of the Scale Grid. Let me briefly say just a few words about a second dimension. We can ask exactly the same questions about the difference between a Joint space and a Population space; we would find an equivalent answer if we just reverse the roles of stimuli and individuals in the argument and analysis just made. The reasoning is not difficult but is too detailed for an address. Let me merely state the conclusions. In a Joint space the points associated with the stimuli are completely independent of the points associated with the individual, whereas in a Population space the points associated with the stimuli are completely dependent on the points of the individuals responding to them. There is a duality between Stimulus spaces and Population spaces: in Stimulus spaces the points associated with individuals are dependent upon the points associated with the stimuli judged, whereas the reverse holds for Population spaces.

We have here two of the dimensions of the Scale Grid. I have constructed

two others which can be used to characterize the data within these major areas of Joint spaces, Stimulus spaces, and Population spaces; but I will say nothing further about them here as they are not relevant to the interpretations I wish to bring out. I will only say that something of their nature is described in my early monograph on theory of psychological scaling (**2**).

If we take the two dimensions which we already have, they suggest a fourth type of space, called a *Field* space, in which the points associated with stimuli and those associated with individuals are completely mutually dependent. This gives us a two-dimensional Scale Grid as illustrated in Figure 5.

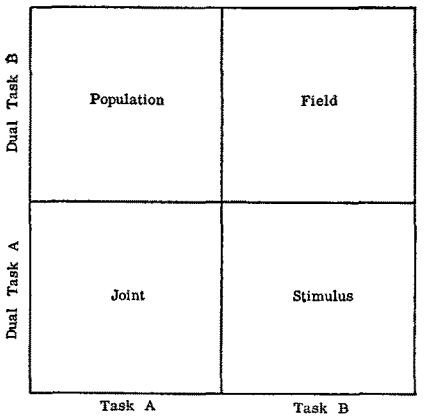

FIGURE 5

The Scale Grid

Some questions naturally arise as to just what might be the significance of all of this and just what this Field space type of behavior is. While we know what goes into a Stimulus, Population, or Joint space, this Field space was a consequence of our analysis of the others, and it is not immediately obvious what significance it has. In order to answer this question, I shall suggest a psychological interpretation of the Scale Grid.

Psychological Interpretation of the Scale Grid

I shall first point out what I consider to be the psychological processes involved in collecting data in a Stimulus space. Then by virtue of the duality between stimuli and individuals, a dual interpretation for Population spaces will be made and certain implications of duality pointed out. From these two kinds of spaces we shall move in one direction and get Joint spaces and in the other direction to get Field spaces.

Consider what is involved when we collect data in a Stimulus space. An individual, the subject, is asked to make judgments about stimuli with respect to an attribute (task B). He is given a set of weights and asked about their

felt-heaviness, a set of tones and asked about their pitch or loudness, etc. The objects of judgment in this situation have many measures; each object has a measure on each of its many attributes—e.g., color, size, heaviness, form, volume, aesthetic quality, etc.

So we have stimuli corresponding to points having several components, and individuals instructed to select one of these components and evaluate the stimuli with respect to that attribute. Let me say it again and contrast the difference—the stimuli are points on many attributes, the individual comes with an attribute in mind but no scale position of his own from which to evaluate the stimuli. In an exaggerated sense, for a Stimulus space the stimuli have provided the points, and the individual has provided the attribute. These then are the respective functions of stimuli and individuals in generating data in a Stimulus space. The behavior observed is ultimately converted to measures of the stimuli on the attribute. The behavior observed may run the gamut of paired comparisons, rating scales, or free-answer protocols—these differences are not relevant here. The important thing is that the behavior observed is interpreted as data which are relations on the stimuli with respect to the attribute. The analysis then leads to measures of the stimuli only.

Let us now exercise the duality relation and consider what the process must be for a Population space. When we reverse the roles of stimuli and people we must have a group of individuals, each possessing measures on many attributes just as the stimuli had in a Stimulus space. An individual thus corresponds to a point with many components. Then we must have stimuli which, carrying through the analogy, must be instructed to select one of these components and evaluate the individuals with respect to that attribute. In the Population space the individuals provide the points, and the stimulus provides the attribute. These then are the respective functions of stimuli and individuals in a Population space. What kinds of behavior do psychologists observe in which these are the respective functions of the stimuli and the subjects? I would say that certain questionnaires, certain interest and neurotic inventories, and essay examinations represent the kinds of behavior that are typically mapped into Population spaces. There is a variety of possible methods of observing such behavior, but, speaking category-wise, the most typical are rating scales and free-answer protocols.

The rating scale method is the questionnaire item with a number of ordered alternatives—an example from a questionnaire used on soldiers during the war is the following:

Do you ever get so blue and discouraged that you wonder whether anything is worth while?

a) Hardly ever
b) Not so often
c) Pretty often
d) Very often

Such a procedure is formally equivalent to asking an individual to judge weights as being light, medium, or heavy. But one experiment is in a Population space, the other in a Stimulus space.

Free-answer protocols are illustrated by the essay examination, the open-ended questionnaire, and the interview. The individual is asked a question which in principle specifies an attribute, e.g., How do you feel about the farm policy? The individuals who are asked this question are playing the role of stimuli being evaluated with respect to an attribute. The protocols which emerge are analyzed for relations between the individuals, which constitute measures of them on this attribute. This is formally equivalent to the use of individual's evaluations of stimuli with respect to an attribute, which leads to information about where the stimuli are on the attribute selected by the individual.

One immediate consequence of this duality between Stimulus spaces and Population spaces is that any method for collecting or analyzing data constructed for either one of these spaces immediately becomes a potential method for collecting or analyzing data for the other. Thus we have, for example, Thurstone's Law of Comparative Judgment constructed for analyzing the judgment of individuals about stimuli and arriving at a Stimulus scale. Immediately there is implied the dual method of having stimuli make paired comparison judgments about individuals—as yet I have seen no good way of getting stimuli to do this. However, there are variations of this basic method of Thurstone's: the Method of Successive Intervals and the Method of Equal Appearing Intervals are used for constructing Stimulus spaces which do transfer completely to Population spaces. To transfer the Method of Successive Intervals to Population spaces you need to have stimuli sort people into piles. I suggest this is exactly what is done by those questionnaire items with multiple alternatives, e.g., from strongly agree to strongly disagree. Abstractly, we can look upon such behavior as stimuli sorting individuals into piles. In the Method of Successive Intervals the instructions to the subject are dualistically equivalent to the writing and editing of an item for a questionnaire or essay examination. I find it strikingly curious that we frequently tend to use five degrees of indorsement or five ordered steps in the alternatives, whereas in the Method of Successive Intervals we have individuals sort items into as many as eleven piles. Whether there is a profound reason behind this, or whether it is unjustified adherence to tradition, I am not sure.

Just as Thurstone's methods for Stimulus scales are transferable to Population spaces, so also are methods designed for the analysis of data in Population spaces alternative methods for Stimulus spaces, e.g., Lazarsfeld's methods of latent structure analysis could be used for scaling stimuli in Stimulus spaces by reversing the subscripts which identify stimuli and individuals, and obtaining the appropriate kind of judgments from individuals.

We see here, in fact, the relation of certain methodologies of the psycho-

physicist studying Stimulus spaces to those of the social psychologist studying Population spaces. What the first makes people do to the stimuli, the latter makes stimuli do to people. Their methods of doing research, collecting, and analyzing data are formally isomorphic with the roles of stimuli and people reversed. Surely, with reference to methodology, whatever one develops suggests a dual development for the other. Each delineates an attribute on which the objects of judgment are to be evaluated. In psychophysics the experimenter does this through his instructions to the subject; in questionnaires and essay examinations the experimenter does it through his careful writing of items. All the various experimental controls developed in one context, again, in principle, transfer to the other context with the reversal of roles between stimuli and people.

The greater status in measurement of psychophysics is due, at least in part, to the fact that an individual can compare two stimuli directly, whereas a stimulus cannot compare two individuals directly. We are much happier with the judgment of an individual as to which of two attitude statements he prefers to indorse than we are with the judgment of which of two individuals indorses a given statement more strongly. The reason for this is very simple. When we ask an individual which of two stimuli he prefers, we assume he has an implicit standard of measurement that is an ordered scale applicable to both stimuli. If we wish to compare two individuals as to which indorses a stimulus more strongly, we have to assume not only that they each have an implicit interval scale but also that the scales have the same origin and the same unit of measurement. Thus, if individual A says he would pay \$10 for a picture and individual B \$5, how do we know but that A has less value for money than B? Once we have made the assumption of an interpersonally comparable interval scale, there is no sense in reducing the data to a paired comparison—that would be throwing away information already assumed. This argument can be summed up by saying that the implicit standards of judgment of one person are presumably more stable over the two stimuli than the standards of two people over one stimulus. This might well be the consideration that underlies using fewer alternatives for an item in a questionnaire than the number of piles used in the Method of Successive Intervals.

When one looks at the differences between areas in this context one finds no justification for quarrels nor for differences in respectability. One area can use a system just as logically precise as the other, but the basic data observed in one area may be a weaker relation than is observed in the other area.

With Stimulus and Population spaces mutually described and related we turn briefly to Joint spaces. Here both stimuli and individuals come together, jointly specifying what the attributes will be; both have their own measures on these attributes. For example, consider an individual working an arithmetic problem. The arithmetic problem is represented by

a point with measures on one or more components. The individual is similarly represented by a point with measures on these components. The response of the individual to the stimulus will be information about the relation of these two points in the Joint space.

By defining what this information is in different ways, one gets Guttman scalogram theory, test theory, one of Lazarsfeld's models, or my unfolding technique for the analysis of preferences. All of these are just different models for what a response on the phenotypic level means in terms of distances between pairs of points in the Joint genotypic space. [The relation of these various spaces to the classification of methodologies contained in (**3**) should be pointed out. The methods of collecting data which apply to Joint spaces are classified in Quadrants I and II, and the methods which apply in Stimulus spaces or in Population spaces are classified in Quadrants III and IV.]

In Joint spaces all the psychophysical methods for analyzing experiments concerned with thresholds, as distinct from those concerned with measuring only the stimulus magnitudes, are present. The individual in such experiments is regarded as having a threshold on an attribute, such as his sensitivity to light or his ability to discriminate pitch. This characteristic of the individual corresponds to a point in the genotypic space, which we have called his ideal on that attribute. The stimulus then is an increment of light or a difference between two tones, and the individual is asked whether he observes it. The stimulus is then also represented by a point in the space. The response of the individual is a formal relation on the pair of points in exactly the same manner as passing or failing an arithmetic item.

The data obtained from neurotic inventories and interest inventories are typically mapped into Joint spaces. For example, an individual is asked a question like "Are you shy?" which he is to answer *yes* or *no*. The individual is presumed to prossess and recognize his particular amount of shyness—this corresponds to a point in the genotypic space, which is his ideal. The question "Are you shy?" with the alternatives *yes* or *no* also corresponds to a point in the genotypic space which is that amount of shyness the individual feels he should have to say *yes*. This amount of shyness is formally equivalent to the difficulty of an arithmetic problem, the increment of light, or the difference between two tones in the preceding examples. Again the individual's response to the question is interpreted as a relation between the respective points.

That the data obtained from individual's indorsements or preferences between attitude statements may also be mapped into Joint spaces is too obvious to need further description. Most learning experiments are mapped into Joint spaces. A conditioning experiment, for example, is like an objective test given backwards: a combination of unconditioned stimulus and conditioning stimulus may be thought of as an item, and eliciting a conditioned response is equivalent to passing an item. Then the most difficult item in

the test is presented first, i.e., the first presentation of conditioning stimulus and unconditioned stimulus, and the individual usually fails it. As learning takes place each successive presentation is essentially an easier item until items are so easy that the individual passes them all. It is interesting to note that the conditioning test has a different method of scoring from the objective mental test, e.g., the number of items taken to reach a certain number of items passed successively. One wonders why most objective tests should have a different convention. I do not object to different conventions, I just like to know what the logic behind them is.

So we have all these superficially different kinds of behavior: objective tests, certain psychophysical experiments, neurotic inventories, interest questionnaries, attitude scale studies, and conditioning experiments. All tend to develop their own methodologies and their own vocabulary—but all are formally isomorphic and hence their methodologies transferable from one to the other. A model for analysis of one of these kinds of data with a particular distance function, for example, immediately raises the question whether it does not also constitute a theory about each of the other seemingly different kinds of behavior. There are, of course, differences in the characteristics of the data one gets in these areas. In some areas experimentally independent replication is possible, in others not; in some areas the stability of a point associated with an individual or with a stimulus is greater or less than in other areas. But fundamentally these differences are quantitative, not qualitative, and the methodological contributions to any one area are in principle transferable to all the others.

Before going on to a psychological interpretation of Field spaces, I should digress for a moment and clear up a possible source of confusion. I have covered Stimulus, Population, and Joint spaces using repeated illustrations. There is a danger of certain misconceptions arising from the illustrations, which we must try to avoid. When I have illustrated one of these spaces, I have tried to follow the most conventional ways of analyzing such behavior, but the implication should *not* be drawn that the theory says there is only one kind of quantitative data or only one space into which any particular behavior can be mapped.

The act of a psychologist in putting his experimental data into one of these spaces (Joint, Stimulus, or Population) represents an optional decision on his part. The sense in which these decisions are optional is what I now want to make clear. The same behavior may be put into more than one of the spaces, thus reflecting different points of view or problems in the mind of the experimenter. The distinction between behavior and data, which was made earlier, is the relevant principle here. It is sometimes easy to see how the same behavior may be interpreted separately as two different kinds of data and consequently be put into different spaces. We have become so accustomed to certain conventions in the converting of manifest behavior

into data that we sometimes neglect any mapping but the conventional one. While data is obtained from behavior by interpreting the behavior as a relation between points, it is up to the interpreter to decide what to identify as points and to define the properties of the relation.

Consider a study on nationality preferences. Let each subject make paired comparison judgments as to which nationality he prefers. In Thurstone's well-known study (**5**) such data were analyzed by the Law of Comparative Judgment and a scale obtained with the stimuli ranging on a one-dimensional continuum from most preferred to least preferred by the group as a whole. When the experimenter does this, he is regarding "preferability" as an attribute of stimuli and is saying that the individuals made task *B* judgments, substantive judgments, about the stimuli. The behavior is interpreted as an order relation on pairs of points, both of which are stimuli. He is saying the behavior belongs in a Stimulus space, and proceeds to construct a stimulus scale.

On the other hand, one could take the identical experiment and put it in a Joint space. In doing this, one would be assuming that the individuals were also points in the space and that their behavior is to be interpreted as an order relation on distances of the stimuli's points from the individual's point. Thus, the behavior is being put into a Joint space instead of a Stimulus space, and analysis of the data by multidimensional unfolding would yield a solution with both stimuli and individuals in the space.

There is nothing intrinsically correct about one of these procedures or wrong about the other. In the first instance, analyzing the data in a Stimulus space, one's problem is essentially that of amalgamating the preferences of individuals to arrive at a single preference scale, which in some sense best represents all the individuals. This is the problem of social utility or social choice. In the second instance, analyzing the data in a Joint space, one's problem is that of discovering the latent attributes underlying nationality preferences from which an individual's preferences could be derived. It might be parenthetically remarked that these two solutions would bear a certain interesting relation to each other, this relation has been developed in two previous publications (**2**, **4**).

Here we have taken an example of behavior and made the transition into two different kinds of data, analysis of which yields different results. We usually overlook this step that we take between behavior and data because this step, at least in some areas of research, is so conventional and immediate. Everyone can usually agree on what is or is not the right answer to an arithmetic problem; when an individual says this weight is heavier than that one, everyone usually agrees he means this weight is higher on an attribute of felt-heaviness than that one. But when we have an individual's answer to an essay examination question or his answer to an open-ended questionnaire item, we speak of "coding" them. This is the process of con-

verting the behavior to data by processing it through the mind of another person to get statements of magnitude or relations—these data are what are analyzed. It is important to note that what one analyzes is always data, not behavior.

This distinction between behavior and data now becomes an even more important and relevant distinction as we turn to a psychological interpretation of Field spaces. To arrive at this interpretation we move along two dimensions of the Scale Grid simultaneously. In going from Joint to Stimulus spaces, the point associated with the individual became dependent on the stimuli being judged. There ceased to be a unique point characterizing the individual. Another way of looking at it is that the stimulus ceased to define the attribute with respect to which the judgments were made. In passing from Joint to Population spaces, the point associated with a stimulus became dependent on the individuals being judged. There ceased to be a unique point characterizing the stimulus. Another way of looking at this is that the individual ceased to define the attribute with respect to which the judgments were made.

In Joint spaces both stimuli and individuals are points and jointly define the attribute. In Stimulus spaces the stimuli are independent points, and the individuals are instructed to define the attribute. In Population spaces the individuals are independent points and the stimuli are instructed to define the attribute. Putting these together for Field spaces, we have the points for stimuli and individuals mutually dependent with neither instructed to define an attribute.

If you wanted to observe such behavior what would you do? You would present an individual with a stimulus that was of such an ambiguous nature it would not arouse any common attribute space in individuals. At the same time the individual would be totally uninstructed to respond with respect to any particular attribute space. This is my definition of what would be a perfect projective test situation. The behavior that is observed is associated with a point in a psychological space with which both the individual and the stimulus are identified.

It is to be noted that in all the other types of spaces (Joint, Stimulus, and Population) the attribute space is at least implicitly defined by the stimuli and/or by the instructions to the subject. Consider all the care given to selecting and wording items properly so that they will ask exactly the right question. This is nothing more than trying to limit the attribute space generating behavior. Exactly the same objective underlies the care in the communication of instructions to the observer in a psychophysical experiment. This care is taken to insure that he will ask the same question of every stimulus. It is then assumed that these precautions have succeeded, and the behavior is interpreted as information about a pair of points, or the distance between them, or a pair of distances. This mapping, done by definition, is what translates behavior into data.

In a Stimulus space, the individual was instructed to ask of the stimulus how heavy it was, or how esthetically pleasing it was, etc. An attribute was explicit, and so the behavior could be interpreted as magnitudes on an attribute and thus made into data. Analysis of such data leads to conclusions about *interstimulus differences*. In a Population space the roles are reversed: a stimulus comes to the individual and asks him how he feels about the farm policy. Again an attribute is explicit. The behavior is interpreted as magnitudes on an attribute, analysis of which leads to conclusions about *interindividual differences*.

In a Joint space both interpretations are possible because the behavior is interpreted as data on a relation between individuals and stimuli. In a Field space the point associated with the individual has merged with the point associated with the stimulus. The behavior is information about this point in a psychological space, a point in which the subject and the stimulus are inextricably identified.

The care given in Population spaces to selecting items for a questionnaire or essay examination in order to ask every individual the same question, and the care given in Stimulus spaces to phrasing instructions to the subject in order that he evaluate all the stimuli on the same attribute is now exercised in Field spaces so that precisely these effects will *not* occur. Every effort is made to insure a setting in which the stimulus will *not* suggest a particular attribute space, and every effort is made in the instructions to the subject *not* to suggest a particular attribute space. Herein lies both the strength and weakness of Field spaces. The protocol that emerges now constitutes a stimulus to be evaluated—so it is a stimulus to be located in a Stimulus space. In order to convert this protocol, this behavior, into data certain problems need to be solved. One is: what is the attribute or attributes which underlie the behavior? This is a new problem which had not previously arisen for any of the other spaces. Here now we have a protocol which is to be converted into a measure on some attributes. The first problem is: which attributes? This problem arises because there was not deliberately built into the stimulus nor into the individual constraints or instructions which would provide a simple answer.

It is immediately obvious that behavior in this area does not lead to interindividual comparisons, because there has been no instruction to the stimulus, no built-in device by virtue of which it can be assumed that a stimulus has evaluated each individual on the same attribute. If one person exhibits guilt feelings and another does not, one cannot conclude the latter person has less guilt feelings unless one can assume the stimulus was such as to make every individual reveal his guilt feelings. In this case, of course, such data could be put back in Population spaces, and interindividual comparisons would be possible.

It may help one to recognize and understand this problem if we point

out that it is like having the answer of an individual to an essay examination question when you do not know what the question was. Consider, then, having the answers of several individuals, each to an unknown question; the problem is to decide which individual's answer represents more of someing than another's. It seems to me the problem is meaningless if the question answered by an individual has been left up to him to select, hence each individual has perhaps selected a different question. One could say, well I can evaluate their relative command of English, or their vocabulary level, or their handwriting if it is a written protocol. This is entirely correct, of course, and amounts to saying this is the common attribute which the stimulus aroused in all individuals, hence they may legitimately be compared. This puts the behavior into a Population space, not a Field space. What I am talking about are those aspects of the behavior which are not attribute controlled and hence belong in a Field space.

Understandably enough, there have been instruments constructed, called projective instruments, which seek to avoid this particular problem. For instance, there are test instruments in which a picture suggests an attribute; the individual is asked to write a story which is presumed to reveal where he is on that attribute. Examples are Proshansky's Labor TAT, Johnson's Anglo-Spanish TAT, and toy play with negro and white dolls. If these instruments succeed in their purpose, then we have the stimulus coming to the individual with an attribute in mind and asking the individual where he is on it. These instruments then, if successful, are formally the same thing as a rather subtle essay examination or an interview by a laborer, a Mexican, or a Negro. The data that are obtained pertain to a Population space rather than a Field space, and interindividual comparisons are logically permissible. Such instruments, however, are not projective instruments in the sense of belonging in Field spaces. A further question then arises as to whether or not these instruments accomplish their purpose. If they fail to arouse the attribute which the experimenter subtly built into the stimulus then there is serious danger of drawing false conclusions.

Let us assume that the first problem is solved or can be solved—that we can look at the protocol and say what the attributes are which underlie the behavior. Then a second problem arises, which is the most fascinating and perhaps the most important of all: what does it mean that the individual selected these particular attributes to exhibit out of all of those possible in his repertoire? I think that no solution yet exists, but ultimately this problem must be answered in order to interpret projective instruments. This problem lies in the area of the psychology of the individual. Because the attributes were left up to the individual, their selection is a reflection of his internal dynamics. Because each individual is answering different questions, the behavior cannot be taken to reflect interindividual comparisons on a common

attribute. On the contrary, and therein lies both its importance to psychology and its weakness as data, the behavior reflects *intra-individual* comparisons.

A protocol in a Field space reflects a point in a psychological space. When we know enough to interpret the protocol as a measure of that point in a known attribute space, then we shall be able to make comparisons between the points. I suspect that these will be comparisons on some hyperabstract attributes which will reflect intra-individual dynamics. The problems which must be solved to reach this stage are what I would regard as our ultimate measurement problems. Field spaces are a maximally significant domain of behavior. It is the area that reflects intra-individual differences to a degree that no other area does. There are fascinating and important problems for psychologists here. It is my thought and hope that the Scale Grid will help to delineate more clearly the basic measurement problems involved.

REFERENCES

1. Bentley, M. Early and late metric uses of the term "distance." *Am. J. Psychol.*, 1950, **63**, 619.
2. Coombs, C. H. A theory of psychological scaling. *Eng. Res. Inst. Bull. No.* 34, Ann Arbor, Mich.: Univ. Mich. Press, 1952.
3. Coombs, C. H. Theory and methods of social measurement. In L. Festinger and D. Katz (eds.), Research methods in the behaviorial sciences. New York: Dryden Press, 1953.
4. Coombs, C. H. Social choice and strength of preference. In R. M. Thrall, C. H. Coombs, and R. L. Davis (eds.), Decision processes. New York: Wiley, 1954.
5. Thurstone, L. L. An experimental study of nationality preferences. *J. Gen. Psychol.*, 1928, **1**, 405-424.

Manuscript received 9/4/56