

Prospects for the Development of Psychological Measurement

One of the theses argued in this book is that the theory of conjoint measurement opens the way to measurement in psychology, providing, of course, that psychology actually deals with quantitative variables. In the preceding chapters three theories of psychological measurement have been considered, and it was shown how they each entail the existence of structures of the kind treated by the theory of conjoint measurement. That is nothing but the first step down a long and difficult road. In this chapter let us consider some later steps in a general and schematic way.

THE NEED FOR STIMULUS CONTROL

The obvious second step, and one stressed throughout this book, is the need to develop technologies of stimulus control. In so far as stimuli are physical objects or events, and primary interest is in purely physical attributes of such stimuli, the technology exists. So the possibility of developing psychophysics using the theory of conjoint measurement is a practicable one (see Falmagne, 1986). However, in so far as the attributes of interest are nonphysical (for example, the favorability of attitude statements; the right or left wing tendencies of a political statement; the authoritarianism of a political candidate; even, the difficulty of an intellectual problem) there is little in the way of any technology of stimulus control.

So far in the history of psychological measurement there has been little attempt to achieve this kind of control. The aim has been to use techniques like Thurstone's scaling procedure, multidimensional scaling and to a lesser extent, Coombs' unfolding procedure, in the hope that they would identify the relevant stimulus

dimensions. This is a fatuous hope in the absence of evidence to show that these techniques are based on true theories. This approach shows the influence of operationism on attempts at psychological measurement. It is thought that these procedures must measure something and it is simply a matter of deciding what. As I have argued, this is not so. They may not measure anything. The experimental method must be applied to the underlying theories before that issue can be resolved. And to apply the experimental method, as shown in earlier chapters, the stimuli used must be carefully controlled.

My plan in this section is to give an illustration of how this might be approached by taking an area notorious for lack of such control, the area of attitude measurement. The stimuli in this case are attitude statements, and the relevant dimension is considered to be the favorability of such statements towards the object of the attitude. The aim is to develop procedures whereby statements may be constructed that vary on no other relevant variable than this. This involves clarifying precisely what properties of statements fix their value on the favorability variable, and what properties locate them on other possibly relevant variables. To do this one needs to characterize the favorability variable.

The concept of attitude is a poorly defined one and, yet, a consistent thread of meaning can be detected. It runs from Thurstone's (1931) definition of attitudes as "affect for or against a psychological object", through Katz's and Stotland's (1959) definition as "an individual's tendency or predisposition to evaluate an object or a symbol of that object in a certain way", to the more recent definitions of Fishbein, (1967), Maze, (1973) and Rokeach, (1968), who all agree that the concept of evaluation is central to that of attitude. This common thread is that an attitude towards some object or policy is necessarily evaluative and, thereby, has a location on a dimension running from pro (or favorable) to anti (or unfavorable). This I take to be the core of the concept and, thus, define an attitude towards person or policy x as an evaluation of x .

So it follows, then, that an attitude statement is an evaluative statement. If the concepts of good, bad, right, and wrong are taken as paradigmatic evaluative concepts then a necessary and sufficient condition can be stated for a statement's being evaluative.

A statement, S , is *evaluative* if and only if it implies that some object or policy is either good or not good, right or not right.

Applying this conception to the statements used in, so-called, attitude scales it is obvious that many of them are not evaluative. Consider, for example, some of the statements in Thurstone and Chave's (1929) study on attitudes towards the church. They included the following statement:

- (1) I believe the church is losing ground as education advances.

Such a statement, of itself, does not imply that the church is either good or not good. Implicit, of course, is the idea of a causal connection between education opening people's eyes and their seeing the teachings of the church as false, but

even that, itself, is nonevaluative. It becomes evaluative when conjoined with certain value judgments about education, enlightenment and believing falsehoods. Indeed, because value judgments of that kind vary amongst social groups, judgments about how favorable or unfavorable (1) is towards the church may vary from person to person. On the other hand, a statement such as

- (2) The church is an evil institution

(not, incidentally, one of Thurstone's and Chave's) is intrinsically evaluative, for it implies that the church is not good.

What is required, then, in order to test theories like Thurstone's, Coombs' and that of multidimensional scaling in the attitude area, are intrinsically evaluative statements that do not differ on other relevant dimensions. What the other relevant dimensions of a statement are, of course, is a matter to be investigated. I will assume that purely grammatical attributes are not relevant and that the only relevant features have to do with meaning. Then a necessary and sufficient condition for a statement's having nonevaluative content may be given.

A statement, S , possesses nonevaluative content if and only if it implies some nonevaluative statement.

Thus, for example, (2) above implies that the church is an institution and so possesses nonevaluative content. A unidimensional set of attitude statements could then be constructed by ensuring that the statements differ only with respect to their evaluative content. For example, consider the following statements,

- A. Bob Hawke is a perfect Prime Minister.
- B. Bob Hawke is an extremely good, but not perfect, Prime Minister.
- C. Bob Hawke is a good, but not extremely good, Prime Minister.
- D. Bob Hawke is a bad, but not very bad, Prime Minister.
- E. Bob Hawke is a very bad, but not absolutely deplorable Prime Minister.
- F. Bob Hawke is an absolutely deplorable Prime Minister.

They all contain precisely the same nonevaluative content (viz., Bob Hawke is a Prime Minister) and differ only with respect to their evaluative content. Furthermore, in so far as their evaluative content is concerned they are ordered from favorable to unfavorable (from A through F). This has been achieved by so wording them that if any one of them is true all the rest are false. Hence, they must each occupy a different value or position on the dimension of evaluations of Bob Hawke as Prime Minister. Such a carefully worded set of statements, differing only with respect to a single semantic dimension, may then provide a basis for testing theories like the ones considered in earlier chapters. More than that, such theories, if true, provide a basis for testing the hypothesis that such semantic dimensions are quantitative.

By way of illustrating these points, 103 psychology students at the University

of Sydney made pair comparison preference judgments on all 15 pairs of these six statements. The aim was to show that for carefully constructed sets of attitude statements, Coombs' unidimensional unfolding theory provides a basis for attitude measurement. Two subjects failed to complete the 15 pair comparison judgments and their data were excluded. Of the remaining 101 subjects, the judgments from ten were not transitive and a complete preference ordering of the six statements was not entailed. However, in most cases only one or two violations of transitivity were involved.

Given the order of the statements along the dimension of favorability, Coombs' theory predicts that the preference orders produced by the subjects should fall on a single path through the graph displayed in Fig. 7.7. 90 of the 91 preference orders produced were those given in that figure. The subgraph produced by this data is shown in Fig. 8.1. The dominant paths are shown by the solid arrows and as can be seen there are two dominant orderings of the interstimulus midpoints (Fig. 8.2). These both satisfy double cancellation (Fig. 8.3).

When the subjects whose data did not conform to Coombs' theory are taken into account the solution of best fit is that shown in Fig. 8.4 and it accounts for 1488 out of the 1515 pair comparison judgments made by the 101 subjects (i.e., 98.2% of the judgments). This result confirms the hypothesis that the subjects are able to discern the single relevant attribute in this case. That is, the method of constructing the stimuli was successful. Furthermore, satisfaction of double cancellation confirms the hypothesis that this attribute is quantitative, though clearly, *that* is an hypothesis that requires further research. This illustration shows that with careful construction, stimuli of a psychological kind can satisfy the boundary conditions for the application of Coombs' theory.

EXTENDING APPLICATIONS OF CONJOINT MEASUREMENT

Another step towards the achievement of psychological measurement is the extension of conjoint measurement theory to other attempts at psychological measurement. Some suggestions are included in Krantz et al. (1971) and Krantz (1972), showing how the theory may be applied to Stevens' theory of magnitude estimation and cross modality matching (Stevens, 1986). This is a significant application because of the prominent place of Stevens' methods within modern psychophysics. Magnitude estimation is one of the most widely used methods in the attempt at psychophysical measurement and it is, therefore, important that it be subjected to experimental examination. Two other very widely used methods in the quest for psychological measurement are the ability test and the rating scale. Indeed, it is no exaggeration to claim that between them they would account for the majority of attempts at psychological measurement. How conjoint measurement may be applied to the theory of abilities is explored in the following paragraphs.

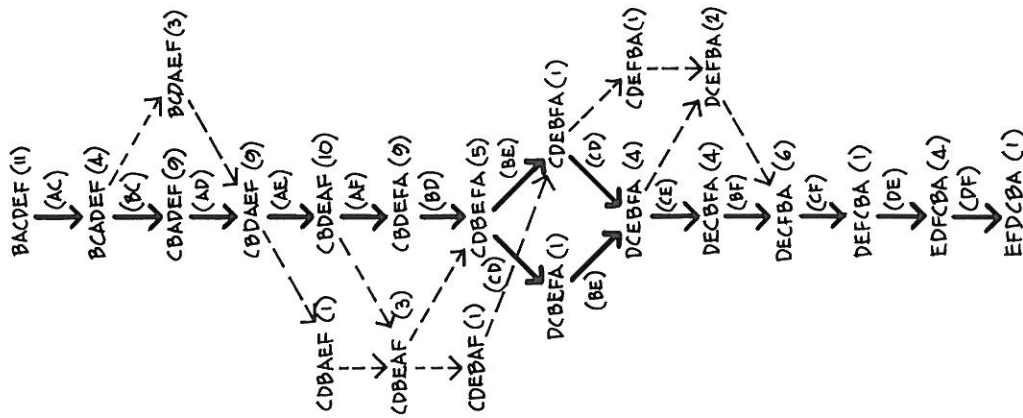


FIG. 8.1. The preference orders obtained from the 90 students who gave transitive pair comparison preference judgments to the statements evaluating Bob Hawke as Prime Minister and whose judgments were consistent with the order of the statements along the evaluative dimension. The solid arrows trace the dominant path through the graph shown in Fig. 7.7.

The suggestion that collections of intellectual tasks may be used to measure intellectual abilities is one that has permeated much of modern psychology, and found wide acceptance within schools, and other institutions with an interest in intellectual abilities. However widely accepted this suggestion is, it is at present no more than a hypothesis, and indeed, one that is widely accepted more because of its usefulness to those who believe it, than because of the evidence in support of it. Consequently, there is an urgent scientific need to extend conjoint measurement theory to the theories used in support of this hypothesis, so that its truth can be assessed.

The most important of such theories are the factor analytic theories stemming

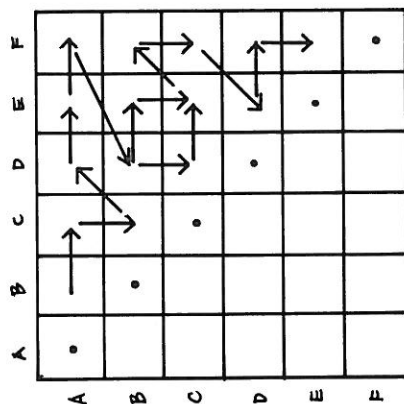


FIG. 8.2. The order on the midpoints entailed by the dominant path shown in Fig. 8.1.

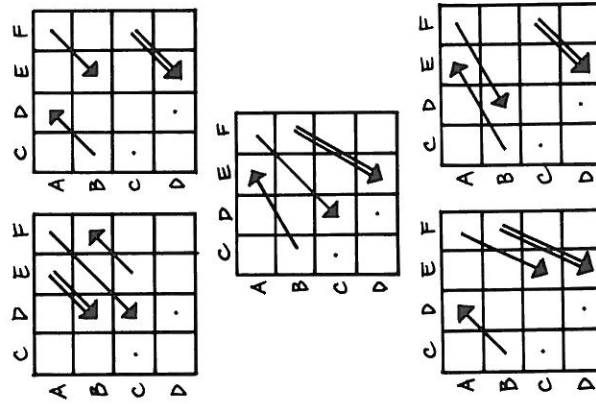


FIG. 8.3. Testing double cancellation on the midpoint order given in Fig. 8.2.

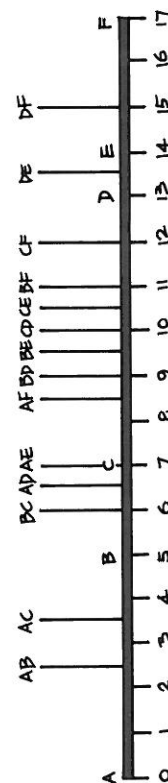


FIG. 8.4. A numerical scale consistent with the order on the midpoints in Fig. 8.2.

from Cattell, (1971), Spearman, (1904) and Thurstone, (1935). These theories postulate that a person's score on an ability test is a function of his abilities and the extent to which the test measures them. These abilities are hypothesized to be quantitative variables. It is thought that the factor analytic method provides a way of identifying them from the patterns of intercorrelations between test scores. However this method, whatever its merits, presumes that abilities are quantitative, and it is precisely this presumption that must be put to some kind of experimental test. Unfortunately, most factor analytic theories of ability are of such a level of complexity that they are not easily reduced to the kind of structure that conjoint measurement deals with. However, the least complex of them all, Spearman's so-called "two-factor" theory, does entail such a structure.

Spearman's theory was that if a test made up of intellectual tasks, all of *one kind*, is considered then a person's performance on such a test is determined by two variables, general ability (g), and the specific ability (s) suited to tasks of that kind. Each specific ability was thought of as a "mental engine" for solving problems of a particular kind, and general ability was thought of as the "mental energy" that drove those engines. People were thought to differ in both the competence of their "engines" and in the quantity of their "energy", and this accounted for their different performances on ability tests. The theory is today regarded as false because the factor analytic pattern it predicts does not generally occur. There are also other predictions, like that of zero tetrad differences, that are now also believed to be generally false. However, in Spearman's defense it must be said that the theory has never been satisfactorily tested. The issue of whether any existing test ever contains tasks of only *one* kind is one that has never been settled. To do so would require a theory specifying the different cognitive processes involved in solving intellectual tasks of various kinds. Given such a theory, genuinely unidimensional sets of test items may be able to be constructed. The issue again is one of stimulus control. Cognitive psychology is yet to furnish the necessary theory.

Mathematically, Spearman's theory may be expressed as

$$z_{ij} = g_i g_j + s_i s_j$$

(where z_{ij} is the z-score of person i on test j ; g_i is i 's measure of general ability; s_i is i 's measure of specific ability; g_j is the extent to which test j measures general ability; and s_j is the extent to which test j measures specific ability). If the test to be used is fixed then g_i and s_i become constants across all subjects. Thus, the j subscript may be ignored, whence,

$$z_i = g \cdot g_i + s \cdot s_i$$

For each subject, i , g_i and s_i are simply being multiplied by the constants g and s , so $g \cdot g_i$ and $s \cdot s_i$ may simply be regarded as rescalings of g_i and s_i . Letting $g \cdot g_i = G_i$, and $s \cdot s_i = S_i$, the equation becomes,

$$z_i = G_i + S_i$$

This equation in turn implies that for any two persons, i and k ,

$$z_i \geq z_k \text{ if and only if } G_i + S_i \geq G_k + S_k.$$

This proposition is a weaker theory than Spearman's (even though it is entailed by it) because it no longer implies that a person's z -score equals $G_i + S_i$, but merely that it be monotonic with that sum. Since z -scores and raw scores are monotonic with each other, this theory is equivalent to

$$X_i \geq X_k \text{ if and only if } G_i + S_i \geq G_k + S_k$$

(where X_i and X_k are i 's, and k 's raw scores on test j). Here we have a theory of the kind amenable to conjoint measurement analysis.

What such an analysis requires, of course, is that levels of the variables G and S be identifiable independently of test scores. Here is where the factor analytic theories of ability fail, for they use test scores as the sole criterion for identifying values of such variables. However, it is just here that the recent *rapprochement* between the areas of individual differences and cognitive psychology may help. An example of what I mean is Hunt's (1980) suggestion that Spearman's g may be identified with attentional resources, and his s with problem solving strategies. If each of these variables can be experimentally manipulated it may be possible to test Spearman's theory. Of course, the value of such a test would be contingent on the truth of Hunt's suggestion, and that may be false. However, this kind of approach is the only one that is going to show whether or not abilities are quantities and how they combine to determine performance. Progress can only be made in this area of psychology by linking the hypothetical quantitative variables of the factor analytic theories (the abilities), to experimentally controllable variables. This requires bold hypotheses similar to Hunt's and while such hypotheses may be false, it is better to show an hypothesis to be false than not to be able to test it at all.

This sketchy analysis of Spearman's theory shows the direction in which research must go if attempts at measurement in this area are to work. First, there needs to be a theory in terms of which tasks or items of a given *kind* can be identified. Two items will be of the same kind when they can be solved using the same family of cognitive processes. Second, there needs to be a theory relating abilities of different kinds to cognitive processes of different kinds (such as Hunt's). Only when these steps are taken can the issue of whether or not ability tests measure anything be experimentally evaluated.

EXTENDING THE THEORY OF CONJOINT MEASUREMENT

A third area where progress has been made in the quest for psychological measurement is in the mathematical development of the theory of conjoint measurement. As expounded in Chapter 4 the theory covers the case where two quantitative variables (A and X) combine noninteractively to produce a third (P). Already the

theory has been developed significantly beyond that point. Krantz and Tversky, (1971) and Krantz et al., (1971) explore certain cases where three quantitative variables (A , B , and C), combine to produce a fourth (D) and show how these cases may be distinguished from one another by the character of the order upon D . In particular they show how the following four composition rules in three variables may be distinguished:

- (1) $D = A + B + C$ (Additive)
- (2) $D = (A + B)C$ (Distributive)
- (3) $D = A \cdot B + C$ (Dual-distributive)
- (4) $D = A \cdot B \cdot C$ (Multiplicative).

The importance of this is that it allows experimental differentiation between rival theories which postulate different composition rules. For example, Hull's learning theory states that habit strength (H), drives strength (D), and incentive motivation (K), combine multiplicatively to produce reaction potential (E): $E = H \cdot D \cdot K$.

Spence's theory, on the other hand, has it that the distributive rule applies: $E = H(D + K)$.

Given that none of these variables can be measured, but that they may be ordered, it follows that conjoint measurement theory is necessary to distinguish between them experimentally. An experimental application of this development in conjoint measurement theory in another area of psychology is given by Coombs and Huang, (1970).

Extending this theme, Krantz (1968) shows how the four variable combination rule $A = B \cdot C + D \cdot E$ may be handled within the conjoint measurement framework. This is interesting because that is the general form Spearman's two-factor theory of abilities takes (i.e., without holding the test constant) and so it opens the possibility of a more complete test than permitted in two-variable conjoint measurement theory.

These extensions to the theory of conjoint measurement consider the effects of increasing the number of variables involved. There is another possible extension to the theory and that is to consider the effects of increasing the complexity of the composition rule with two variables. That is, to extend conjoint measurement theory to include interactive compositions. One such interactive composition rule often considered in the context of analysis of variance is $P = A + X + A \cdot X$ and, as noted by Krantz et al., (1971), a theory for this special case would be of interest.

DEEPENING THE UNDERSTANDING OF QUANTITY

Another area of theoretical activity over recent years has significantly advanced the understanding of what it means for a variable to be quantitative. This is research by R. D. Luce (1986, 1987) and L. Narens (1981, 1985) into the

mathematical foundations of measurement theory. Though highly technical, some of the major results can be explained in less technical terms.

The background to this work is in the theory of scale types. As expounded in earlier chapters of this book, since Stevens' (1946, 1951, 1959) liberalization of the representational theory, psychologists have accepted the view that there are four types of measurement scales: nominal, ordinal, interval, and ratio. These are normally distinguished from one another via the admissible scale transformations involved. However, as became increasingly obvious following Suppes' and Zinnes' (1963) axiomatic reformulation of Stevens' theory, the differences between different scale types must ultimately reside in some intrinsic property of the empirical relational systems involved. For example, an empirical relational system capable of a ratio scale representation, must differ in some fundamental way from an empirical relational system enabling only an ordinal scale representation. Attempts to characterize this difference led to the discovery of the properties called *homogeneity* and *uniqueness* (Narens, 1981). In turn, the concept of homogeneity has enabled the essential nature of quantity to be discerned more clearly.

The concepts of homogeneity and uniqueness have been defined via the concept of an *automorphism*. An automorphism is an isomorphism (or one-to-one mapping) between a relational system and itself. Let X be some nonempty set of empirical entities of some kind. Let $R_1, \dots, R_j$ be empirical relations upon X . Let f be a one-to-one function mapping each element of X onto some other element of X . Then f is an *automorphism* if and only if for any relation, R_i ($i = 1, \dots, j$), and ordered set of elements of X , $\langle x_1, x_2, \dots, x_n \rangle$,

$$R_i(x_1, x_2, \dots, x_n) \text{ if and only if } R_i(f(x_1), f(x_2), \dots, f(x_n))$$

(where " $R_i(x_1, x_2, \dots, x_n)$ " means that the set of objects, $x_1, x_2, \dots, x_n$, (in that order) stand in the relation R_i). That is, in words, a one-to-one function mapping elements of X onto other elements of X is an automorphism if whenever some elements of X stand in a particular relation, R_i , then so do the elements that f maps them into, and vice versa.

An example will make this concept clearer. Suppose that our empirical relational system consists of a set of rods of various lengths and that we have two relations, R_1 and R_2 , where $R_1(x, y)$ means that rod x is not shorter than rod y , and $R_2(x, y, z)$ means that rods x and y linearly concatenated end to end equal rod z in length. Now, suppose the function f maps each rod into another one twice its length, i.e., for any rods, x and y , $f(x) = y$ if and only if $R_2(x, x, y)$.

Then f is an automorphism if and only if for any rods x, y , and z

- (i) $R_1(x, y)$ if and only if $R_1(f(x), f(y))$, and
- (ii) $R_2(x, y, z)$ if and only if $R_2(f(x), f(y), f(z))$.

Obviously, (i) and (ii) are both true, because (i) if x is not shorter than y , then a rod twice x 's length will not be shorter than one twice y 's (and vice versa); and (ii) if the length of x concatenated with y equals the length of z , then the concatenation of rods twice the lengths of x and y will equal the length of a rod twice z and vice versa. So, in this case f would be an automorphism.

The importance of automorphisms is this. Consider any empirical relational system that is at least ordinally scalable. It may be represented numerically in many different ways, each different way being a distinct numerical scale, and each at least positively monotonically related to the others. Indeed, there are infinitely many distinct mappings of a structure that is at least ordinal into the real numbers. If we allow that X , the empirical set involved, is unbounded, dense and continuous (relative to the ordering relation involved), then corresponding to each numerical scale for this empirical relational system there exists an automorphism and vice versa. That is, there is an isomorphism between the set of all numerical scales representing some (unbounded, dense, continuous) empirical relational system and the set of all automorphisms on that system. Hence, any property that the scales possess (such as being ordinal, interval, or ratio scales) must have its analogue in the set of automorphisms for that system. Thus, what it is for an empirical relational system to be ratio scalable (or quantitative, in my terms) can be defined as a property of the automorphisms of such systems. All such systems must have a common automorphic structure. As we shall see later, R. D. Luce has precisely defined this structure and, hence, Luce's definition can be used to give a precise characterization of what it is for a variable to be quantitative.

Let me return to the claim made earlier in the last paragraph, that there is an isomorphism between the set of all numerical scales representing some (unbounded, dense and continuous) empirical relational system, and the set of all automorphisms on that system. Its truth can easily be illustrated for the case of the ratio scale.

Consider two numerical scales of length, say the Imperial scale (where the yard is the unit), and the metric scale (where the meter is the unit). We take any rod, x , and find the number that is its length in yards, call it $y(x)$, and then we find the rod z (I am assuming for simplicity that each rod is a different length and the set of rods is unbounded, dense and continuous), such that the number equal to z 's length in meters, $m(z)$, equals $y(x)$. Here we have an isomorphic mapping from lengths of rods to lengths of rods via numerical measurements. That is,

$$x \text{ maps to } y(x) = m(z) \text{ maps to } z.$$

Furthermore, this mapping must be an automorphism because it is simply the composition of two isomorphisms. So that procedure gives us one automorphism. Others may be obtained by varying the Imperial scale (or, alternatively, the metric scale) as the first step in establishing the automorphism. Select any length ratio scale from the infinite set possible. Find the numerical value of x on that scale, call it $I(x)$. Then we have a new automorphism because,

x maps to $I(x) = m(z)$ maps to z .

And for each numerical scale there will be an automorphism and vice versa.

If for any empirical relational system that is at least ordinal, unbounded, dense and continuous, there is an isomorphism between scales of the same kind (i.e., isomorphisms into the real numbers) and automorphisms (i.e., isomorphisms onto itself), then it simply remains to specify the property or properties of automorphisms that correspond to scale type. Narens (1981) identified two properties of automorphisms that jointly characterize the scale type of measurements of such empirical relational systems. These he called *n-point homogeneity* and *n-point uniqueness*. He defined these properties as follows.

Let $< X, \geq, R_1, R_2, \dots >$ be a relational system, $\geq$ a simple order on X (which is unbounded, dense, and continuous) and $R_1, R_2, \dots$ other relations on X . Let A be a set of automorphisms of this system. A satisfies *n-point homogeneity* (for some nonnegative integer n) if and only if for each $x_1, \dots, x_n$ and $y_1, \dots, y_n$ in X (where $x_1 > x_2 > \dots > x_n$ and $y_1 > y_2 > \dots > y_n$) there exists an automorphism a in A such that $a(x_i) = y_i$ (for $i = 1, \dots, n$). A satisfies *n-point uniqueness* if and only if for each a and b in A , if a and b agree at least at n distinct elements of X then $a = b$. Now, if A is the set of all automorphisms of this relational structure, then this relational structure satisfies *n-point homogeneity* if and only if A is *n-point homogeneous*, and this relational structure satisfies *n-point uniqueness* if and only if A is *n-point unique*.

We can easily see that as an extensive structure length satisfies 1-point homogeneity but not 2-point homogeneity. No matter what elements x_1 and y_1 are of the length variable, there will be an automorphism, a , for which $a(x_1) = y_1$. For, suppose on the meter scale that $m(x_1) = r \cdot m(y_1)$ (where r is some positive real number), then a will be the function mapping each length into another, r times as large and therefore for any lengths x, y , and z , $x > y$ if and only if $m(x) > m(y)$, if and only if $r \cdot m(x) > r \cdot m(y)$, if and only if $a(x) > a(y)$; and relative to the concatenation relation $x \circ y \sim z$, if and only if $m(x) + m(y) = m(z)$, if and only if $r \cdot m(x) + r \cdot m(y) = r \cdot m(z)$, if and only if $a(x) \circ a(y) \sim a(z)$.

Hence, a is an automorphism. Therefore, length is 1-point homogeneous. However, it is not 2-point homogeneous, for letting $x_1 > x_2$ and $y_1 > y_2$ be any lengths whatsoever, we know values can be chosen such that

$$m(x_1) = r_1 \cdot m(y_1),$$

$$m(x_2) = r_2 \cdot m(y_2), \text{ and}$$

$$r_1 \neq r_2.$$

For example, suppose $r_1 = 2$ and $r_2 > 2$. Suppose $x_2 \circ x_2 \sim x_1$, then it cannot be that $a(x_2) \circ a(x_2) \sim a(x_1)$, if and only if $y_2 \circ y_2 \sim y_1$, for y_2 is more than double x_2 , while y_1 is twice x_1 . So there will be values of the length variable for which 2-point homogeneity fails.

As well, as an extensive structure, length is 1-point unique. If $a(x_1) = y_1$ and $b(x_1) = y_1$, then $a = b$. It is, of course, also 2-point unique: if $a(x_1) = y_1$, $a(x_2) = y_2$, and $b(x_1) = y_1$, $b(x_2) = y_2$, then of course $a = b$.

If some relational system is *n-point homogeneous* for every positive integer n , then it is said by Narens to be *∞ -point homogeneous* (see also Narens & Luce, 1986). If some relational system is not *n-point unique* for any finite n , it is said to be *∞ -point unique*.

It follows then, Narens (1981) shows, that relational systems giving rise to interval scales (*linear structures*) are 2-point homogeneous and 2-point unique; and those giving rise to ordinal scales (*monotonic structures*) are ∞ -point homogeneous and ∞ -point unique. Furthermore, a quite surprising result of Narens' research is that there are no simply ordered, unbounded, dense and continuous structures that are both *n-point homogeneous* and *n-point unique* for $2 < n < \infty$. This research demonstrates, first, the great structural gulf between scalar and linear structures (both of them I call quantitative structures) and monotonic (or ordinal) structures. Second, it shows why representationalists have identified so few scale types for structures that are at least ordinal.

Luce (1986, 1987) builds upon Narens' concept of homogeneity, showing precisely what structure the automorphisms of a ratio scalable structure must have. Considering what he calls the class of *ordered relational structures* ($< X, \geq, R_1, R_2, \dots >$, where X is a nonempty set, $\geq$ is a simple order, and $R_1, R_2, \dots$ are relations upon X), Luce distinguishes between automorphisms that do not fix any points and those that do. An automorphism, f , fixes a point when for some x in X , $f(x) = x$. There will always be the automorphism that fixes every point (i.e., the automorphism i , such that for all x in X , $i(x) = x$). This is called the *identity automorphism*. The identity automorphism together with all automorphisms that do not fix any points (relative to a relational structure) Luce designates *translations*. What Luce shows is that an ordered relational structure's being ratio scalable is logically equivalent to three conditions being satisfied by the set of all of its translations.

Let $O = < X, \geq, R_1, R_2, \dots >$ be any ordered relational structure. Let T be the set of all translations upon O . T satisfies conditions (1), (2), and (3) below, if and only if O is isomorphic to a real unit structure (i.e., ratio scalable):

1. T is a (mathematical) group;
2. T is simply ordered and Archimedean;
3. T is one-point homogeneous (what Luce calls, simply, *homogeneous*).

Let me explain each of these conditions in detail and, also, what Luce means by a real unit structure.

A mathematical group is a special kind of algebraic structure. It consists of a set, A , a binary operation, $*$, on A , and an element, e , of A (the identity element) such that,

- (i) for all a in A , $a * e = e * a = a$;
- (ii) for all a, b , and c in A , $(a * b) * c = a * (b * c)$; and
- (iii) for each a in A there exists b in A such that $a * b = e$.

So for T to be a mathematical group there must be an operation on it and an identity element within it satisfying (i), (ii), and (iii). The operation is that of function composition. Let s and t be any two elements of T . s maps each element of X into another and so does t . Relative to s and t we can define another element of T , r , as follows: for any x in X ,

$$r(x) = y \text{ if and only if } s(x) = z \text{ and } t(z) = y.$$

That is, r is the composition of s and t (written: $r = s * t$). This identifies the necessary operation. The identity element is just the identity translation, i , mentioned above. So condition (1) is actually saying that $< T, *, i >$ is a group, i.e.,

- (i) for all t in T , $t * i = i * t = t$;
- (ii) for all r, s , and t in T , $(r * s) * t = r * (s * t)$; and
- (iii) for each t in T there exists s in T such that $t * s = i$.

Condition 2 states that there is a simple order on the elements of T . This may be defined relative to the order on X for some range of X 's values. That is, for any s and t in T ,

$$s \geq t \text{ if and only if for some fixed } y \text{ in } X, \\ s(x) \geq t(x) \text{ for all } x \text{ in } X \text{ such that } x \geq y.$$

However, if O is ratio scalable then the translations in T do not cross (i.e., if for any x in X , $s(x) \geq t(x)$ then for all x in X , $s(x) \geq t(x)$) (see Luce, 1987), and so the order on elements of T applies across all values of X .

The simple order on T is Archimedean in this sense: if any s and t in T are such that $s > i$, then for some positive integer, n , $s^n > t$ (where s^n means n successive applications of s i.e., $s^1(x) = s(x)$ and $s^{n+1}(x) = s(s^n(x))$, or $s^1 = s$ and $s^{n+1} = s^n * s$).

If T satisfies (1) and (2) then it is an Archimedean ordered group and is isomorphic to a subgroup of the real numbers (by Hölder's theorem (see Krantz et al., 1971 p. 53). Furthermore, if T is 1-point homogeneous, then O has a numerical representation in which each element of T maps into multiplication by a positive constant. That is, O is ratio scalable and T corresponds to the class of positive similarities transformations. Luce refers to any O having this structure as a *unit structure*.

By a *real unit structure* Luce means a numerical relational system whose domain, R , is a subset of the positive real numbers, whose relations are $\geq$ and

$R_1, R_2, \dots$ and it is such that there is some T' (a proper subset of the positive reals) where

- (i) T' is a group (with multiplication as the operation);
- (ii) T' maps R into R (i.e., for each r in R and t' in T' , $t'r$ is in R);
- (iii) T' restricted to R is the set of translations of this structure.

So a real unit structure is a unit structure of real numbers.

The significance of Luce's work for the theory of measurement proposed in this book is that it specifies what it is for a variable to be quantitative: a variable whose values are simply ordered is quantitative if and only if its translations are 1-point homogeneous and constitute an Archimedean simply ordered group.

A second advantage of Luce's formulation is that it enables *ratios of values* of a quantitative variable to be defined in a more general way than was done in Chapter 2. There ratios were defined via additivity. Of course, all quantitative variables must be additive (in the sense that additive relations are taken to hold between values of the variable). However, additive relations may not be directly evident (i.e., extensive measurement may not be possible). In such cases it may be beneficial to have a definition of ratios that is independent of additivity. Apart from that, it is a conceptual advance to be able to define ratios relative to the empirical relational system.

The way in which ratios may be defined using Luce's result is as follows. Let $O = < X, \geq, R_1, R_2, \dots >$ be an ordered relational structure whose translations are 1-point homogeneous and constitute an Archimedean simply ordered group. Then for any w, x, y , and z in X , $w/x \geq y/z$ if and only if $s \geq t$, (where s and t are those translations in T such that $s(x) = w$ and $t(z) = y$). Each translation in T is really an equivalence class of ratios. That is, for any t in T , $t = \{x/y | t(y) = x\}$.

This makes sense once it is recognized that within a unit structure each translation "expands" or "contracts" the structure by some fixed proportion. There must then be some interpretation of conditions (1), (2), and (3), above (for O to be a unit structure) that applies directly to ratios. Such an interpretation would not only be of theoretical interest. It could be applied to those instances where data consists of (or may be interpreted as consisting of) an order on ratios of elements of X (as, for example, in some interpretations of Stevens' method of magnitude estimation).

Because an ordering on ratios is a special case of an ordering on the Cartesian product of X with itself, it follows that an ordering on ratios is ratio scalable (i.e., is a unit structure or quantitative) if and only if it is a weak order satisfying double cancellation, solvability and the Archimedean condition. Here then we have an interpretation of Luce's result about translations at the level of ratios, and it turns out to be a special instance of conjoint measurement. Given the role of both ratios and conjoint measurement in this book, this is an interesting conclusion. (Note, of course, that the interval scale is the special case where X is a set of differences).

As already mentioned, this observation about ratios could be applied to the method of magnitude estimation or, more directly, to ordinal judgments of relative magnitude. One magnitude estimation procedure involves the experimenter assigning some number, say 100, to a standard stimulus, and then instructing the subject to assign numbers to other stimuli in such a way that the ratios of these numbers to 100 express the magnitude of the stimuli to that of the standard (relative to some specified attribute). Order relations between such numerical ratios could easily be used to test cancellation conditions on the relative magnitudes (or ratios) of the stimuli. Alternatively, subjects could be instructed to order stimulus ratios directly. Either way the hypothesis that human judgments about some attribute are quantitative could be tested.

CONCLUSION

As must be obvious by now, the science of psychology is still a long way from developing procedures that may legitimately be described as measurement procedures. However, despite Pythagorean cries to the contrary, this does not impugn psychology's status as a science. Its status as a science rests solely on its use of the methods of observation, hypothesis, and deduction. It needs no more to be given that status and in this it is no different to any other science. Psychological measurement will only ever be a reality if some psychological variables are quantitative and that issue is an empirical one and must be settled by experimental research. What has been considered in this book are the kinds of experimental evidence that will support (or, alternatively, falsify) the hypothesis that some psychological variable is quantitative. The theory of conjoint measurement provides a method whereby such evidence may be collected.

In this chapter a number of developments that will enhance the quest for measurement in psychology have been looked at. The primary need is for the development of techniques of stimulus control. This is necessary not only for any application of conjoint measurement theory but also for that special application, the construction of standard sequences. It is this, more than anything else, that will establish psychological measurement. In the absence of such sequences legitimate doubts must remain.

The other developments considered relate to extending the theory and applications of conjoint measurement. This is necessary if the full range of quantitative theories used within psychology are to be brought within its scope.

Beyond these needs there lies another: the need to struggle against the mistaken belief that psychological measurement already exists. This illusion is based on the false views of measurement prevalent within psychology. That entrenched illusion is now the major obstacle in the way of those who would sooner test the reality, than accept the myth of psychological measurement.

Appendix 1

The Theory of Order

Quantification in psychology is distinguished from that in physics by the extent to which it has been attempted from a purely ordinal base. Order, of course, is one of the characteristics of quantity. Within psychology it has sometimes been assumed that where there is order, there must also be quantity. Thus psychologists have postulated a number of quantitative theories to explain ordinal effects such as the ones considered in this book. Although this approach has not yet been resoundingly successful in establishing psychological measurement on a sure footing, it has given rise to some interesting theories, some widely used techniques (especially multidimensional scaling), and one important discovery, the theory of conjoint measurement. Through this theory it has been discovered that measurement can be achieved upon a purely ordinal base.

THE CONCEPT OF ORDER

An order on a class of entities is a relation that lines them up so that each entity is earlier or later than each other. Consider, for example, the points on the line in Fig. 1. They are ordered from left to right as *A, B, C, D, E, F, G, and H*. Any order is a relational system and there are a number of different ways that it can be described. One of the simplest is to characterize the order in terms of the relations holding between the pairs of entities. One such relation in Fig. 1 is that

FIG. 1. An illustration of an order: the eight points, *A, B, C, D, E, F, G, and H* are ordered from left to right along a line.

