Reports from the Research Laboratories

of the

Department of Psychiatry

University of Minnesota

Parametric versus Structural Variation in Psychological Theory and Research

by

DAVID T. LYKKEN

Report Number PR-75-3

June 2, 1975

Reports from the Research Laboratories

of the

Department of Psychiatry

University of Minnesota

Parametric versus Structural Variation in Psychological Theory and Research

by

David T. Lykken

Report Number PR-75-3

June 2, 1975

Parametric versus Structural Variation in Psychological Theory and Research

David T. Lykken
University of Minnesota

All ships within a given class of naval vessels, say destroyers, are built in accordance with the same general set of blueprints, the same 'species plan'. But any navy man will testify that each ship is an individual, that no two destroyers handle just the same way, that each has its crochets and idiosyncracies. These idiographic deviations from the nomothetic rule come about as a result of two kinds of causes. Firstly, all varieties of mechanisms - nautical, electronic, organismic - show parametric variation; i.e., deviation from whatever value is specified for each component part in the blueprint, wiring diagram or species plan. The particular behavior of each mechanism, while generally similar to that of other members of the class, will vary from the norm in consequence of such deviations. Secondly, the actual structure of a complex mechanism may differ in small ways from other members of the same class, either because of differences in the original blueprint or due to changes made later, by design or by the accidents of experience. One class of man-made mechanisms, the digital computer, is uniquely susceptible to structural modification. Each new computer program modifies or elaborates the essential structure of the computer, making it into a new and different special-purpose machine which will operate in a unique manner upon the data input. It is, of course, precisely this extraordinary modifiability of structure which impels the familiar analogy between the computer and the brain. The human nervous system is a mechanism which, while it (1) conforms generally to a common species plan, shows great individuality, due to (2) nomothetic, parametric variation, i.e., relatively fixed or stable differences between individuals, resulting largely from genetic factors; and (3) idiographic

structural variation, due primarily to the effects of learning.

Psychologists, by long tradition, tend to focus their research efforts more or less exclusively on one of the three categories of problems suggested above. Physiological and what are called "experimental" psychologists focus on the species plan; their objective is to elaborate a nomothetic account of the "generalized mind" or the average or typical case. At the other extreme, the clinician may attempt a psychodynamic, idiographic understanding of the individual case. Somewhere in the middle are the psychometrists or differential psychologists, the students of ability and temperament and behavior genetics, whose interest is in nomothetic trait variables, definable across individuals but with a focus on parametric variation rather than on structural regularities. For the traditional experimental psychologist, parametric variations - individual differences - are nuisance variables, a source of noise. For psychologists trained in the psychometric tradition, individual differences are the data base itself.

Psychophysiology, with origins that antedate both Galton and Wundt but only lately emerging from its chrysalis, is an interface between these two traditions. Experimentalists, often with a physiological perspective, move into psychophysiology on the hunt for general laws while others, renegade clinicians perhaps or refugees from the swamps of personality research, seek more modestly for objective measures of (important) dimensions of individual differences. This chapter will not be the first place where a plea will have been offered for a more effective, a more calculated and deliberate blending of the two points of view. For the radical experimentalist, science is the job of weaving the nomological net, the discovery of nomothetic laws relating the basic variables of a discipline. For the radical psychometrist, science

rather is the business of discovering what these basic variables are; for some, e.g. some factor analysts, there is even a Skinnerian disinterest in intra-organismic inter-relationships, based upon a simple faith in transfer functions, in the idea that one can predict as accurately as one might wish if one merely includes the right factor scores in one's specification equation.

The truth is, however, that the identification of the 'basic' variables and the study of the laws relating them are the two legs on which science moves forward; to try to hop along on one is to stumble and even a creature with two legs will be ambi-sinister if the two are driven by separate, uncoordinated brains. Galileo, the "father of modern physics", showed us the power of mathematical laws derived from quantitative experiments but one must not forget that laws state relations between variables and that, even in simple mechanics, it required the insight of genius to identify the right variables - the concept of mass, for example, which did not come easily to a 16th Century mind. In the behavior sciences, the problem of identifying 'basic' variables is particularly acute, not because of a shortage of candidates but rather the reverse - we are drowning in them. Each new test or measure, nearly every new experiment, defines one or more new variable; the dictionary offers names for more than 50,000 candidates for the field of personality alone. Considering only laws relating 3 variables at a time ($x = F(y,z)$), these 50,000 trait terms imply the empirical study of more than 10^{13} separate nomologicals. An experimentalist who would advocate forging ahead without prior winnowing of such a list would be more than radical, he would be mad.

But the task of identifying the 'short list' of basic variables -- "basic" in the sense that they might figure in the most orderly set of laws, deriving from a parsimonious and manageable theory -- typically requires the study of

individual differences. Even a task like choosing amongst competing methods of expressing the GSR (e.g., SRR, SCR, SCR/SCR_{max}, log SCR, the Autonomic Lability Score, etc.) requires the technology of the differential psychologist. An issue of considerable contemporary interest is the Lacey hypothesis (Lacey, 1967) to the effect that cardiac activity may be involved in the modulation of external stimulus input. The existence and nature of such an alleged mechanism in the mammalian nervous system is a fairly typical problem area for the experimentalist but the concerns and the techniques of the differential psychologist become immediately relevant also. To determine whether cardiac activity tends to increase when attention is deployed inward, for example, one must decide whether to use, as one's measure of the level of cardiac activity, heart rate or heart period. Although these two indices are perfectly reciprocally related, it is not a matter of indifference which one is used in experimental work. For example, imagine a test of the Lacey hypothesis based on the expectation that some measure of the degree of effectiveness of the internal deployment of attention ought to be correlated with the amount of increase in cardiac activity. It is easy to invent imaginary - but possible - data which would show a positive correlation, supporting the hypothesis, when cardiac activity is measured in terms of heart rate but a correlation of zero, refuting the same hypothesis, if the scientist chooses to measure activity in terms of heart period instead. As in the problem of the choice of electrodermal units, this methodological issue must be approached with the help of both experimental and individual differences techniques.

The same experiment illustrates the use of the individual differences approach to theoretical as well as to methodological issues. Of the many conceivable corrolaries of the Lacey hypothesis, some lead to traditional experi-

mental tests in which an independent variable (e.g. direction of attention) is manipulated while one observes the effect of this on an appropriate dependent variable (e.g. heart rate). But many other tests, like the one suggested above, depend for their existence and utility on the facts of individual differences. Even under the same conditions of measurement (i.e. with no attempt at manipulation), different Ss will show a range of degree of attention deployment inward as well as variation in cardiac response and it is of course this variation in both variables, related ex hypothesi, which makes possible a testable deduction from the theory.

This idea can be stated more generally along the following lines. Most theories in psychology are hypotheses concerning some portion of the species plan, conjectures about the nomothetic structure of the nervous system (this account assumes, of course, that all psychological theories are "about" the CNS in the sense that it is the CNS which mediates between input and output, between stimulus and behavior). An elementary theoretical proposition would be of the form $y = f(x)$, where x and y may be either observable or theoretical variables. A direct experimental test of this bit of theory would be to vary x and observe y, but such experimental manipulation is often difficult. However, most biological variables naturally vary from one organism to another and many variables change in value over time, due to a variety of reasons. This natural variation, within or between individuals, will often provide a substitute - or at least a supplementary - method for testing theories.

Improving Methods of Measurement Using Individual Difference Data

If X = a judge's estimate of Jones' weight, Y = a prediction of weight based on Jones' height and girth, and Z = the reading obtained when Jones steps on a scale, then X, Y and Z are three potential variables of, say, physiology.

Here we know that they are actually all measures of the same, underlying variable but this is not always so easy to see. Our problem, then, might be to determine, first, whether X, Y and Z represent three different dimensions or only one and, having answered this first question, to decide, second, which of these three is the best measure (i.e. of body weight).

Dimensional Analysis

A way to determine whether X, Y and Z all measure the same thing - albeit unequally well - would be to examine their inter-correlations. One must be careful about the linearity assumption, e.g. one should look at the bivariate scatter plots, since we know, for example, that SCR and SRR measure more-or-less the same thing but their product-moment correlation is attenuated since they are reciprocally rather than linearly related. Where the bivariate regressions are linear, the magnitude of intercorrelations should be evaluated in terms of the reliabilities of the individual variables. Body height and weight, for example, correlate about .80 yet, since we know that the reliability of these two variables is .95 or better, it is sensible to treat them as separate dimensions. On the other hand, the Taylor Manifest Anxiety Scale, the Repression-Sensitization Scale, and the Neuroticism subscale of the Eysenck Personality Inventory all inter-correlate between .7 and .8 while their reliabilities are in the same range; hence, we can conclude that all three are essentially measures of the same personality dimension, viz. stability vs. neuroticism or what Block (1965) calls "ego resiliency".

While it is important to avoid treating such measures as separate entities when virtually all their reliable variance is common variance (merely, e.g., because their inventors have given them romantically individual names), an equally serious sin is to assume that some other set of measures which are

claimed to relate to a common underlying variable really do so. Dimensional analysis provides protection from this danger also. Some years ago, for example, psychologists were busily turning out a spate of studies of the personality trait known as "rigidity", each author employing his favorite from among five or six tests alleged to measure this trait. When several investigators finally tried to intercorrelate these measures on the same Ss (Jenkins & Lykken, 1957, p. 99), with results ranging about zero, research interest in "rigidity" tailed off precipitously.

Western psychophysiologists have become increasingly interested in the Pavlovian concept of "strength of nervous system" (SNS), as developed by Nebylitsyn and others of the school of Teplov (Nebylitsyn & Gray, 1972). A number of alternative methods have been advanced for measuring this basic nervous parameter; rate of extinction-with reinforcement of the photochemical reflex, slope of the function relating reaction time to stimulus intensity, susceptibility to photic driving of the EEG, etc. Studies are appearing in which the correlates of some one of these measures are interpreted as correlates of SNS (ibid, chaps. 2, 4, 6, 18-21). Because the several putative indices of SNS are apparently only weakly correlated with each other (ibid, chap. 10), such interpretations seem premature. A more efficient research strategy would seem to be to concentrate first on increasing the common factor variance by trying to improve the measures, i.e. to find ways of reducing the extent to which each variable is affected by measurement error or by irrelevant factors. As discussed in the next section, such an increase in common variance can in fact be employed as a criterion in choosing among alternative methods of measurement.

Another example of dimensional analysis in psychophysiology concerns the problem of partitioning the frequency spectrum of the EEG. There is a

well-established convention of calling waves of less than 3 Hz "Delta" waves, frequencies from 3 to 7 Hz "Theta", 8 to 13 Hz "Alpha", and frequencies higher than 13 Hz "Beta" activity. It is clear that this conventional partition is not intended to be just an arbitrary convenience; Delta and Theta waves are believed to have different psycho-physiological significance and activity at 13 Hz is thought to have less in common with 14 Hz activity than it does with activity at 8 Hz. The warrant for these rather surprising claims lies in an accumulation of years of rather informal, 'clinical' observations of patterns of co-variation of the spectrum components. An increase in 11-13 Hz activity is in fact likely to be accompanied by an increase from 8-10 Hz (e.g. when the eyes are closed) and not by an increase in the Beta band.

With a view toward corroborating these now classical assumptions in a somewhat more systematic fashion, Auke Tellegen and I analyzed 3-minute samples of EEG obtained from 70 pairs of same-sex, adult twins. Each sample was spectrum-analyzed, yielding a magnitude spectrum of 40 ordinates at 0.5 Hz intervals from 0.25 to 19.75 Hz. The spectra were then 'standardized', i.e. reduced to unit area, a process which eliminates differences due to average EEG amplitude. The 40 spectrum ordinates were then intercorrelated across the 140 Ss and these data were factor analyzed. After rotation (Varimax), there appeared to be five - possibly six - significant factors which accounted for from 80 to 84 percent of the variance in these spectra. Three of the rotated factors were immediately identifiable with three of the classical frequency bands, Delta, Theta, and Beta; total activity in these bands correlated .92, .94 and .91 with Factors V, II and I, respectively. Factor III was positively loaded by activity from 7.25 to 8.75 Hz and negatively by activity from 10.25 to 12.25 Hz. Factor IV was positively loaded by spectrum activity from 11.75

to 14.25 Hz and negatively by activity from 9.25 to 10.25. Factor VI, although very weak, also showed this bipolar contrast between higher and lower portions of the Alpha band (9.25-9.75 vs 10.75-11.25 Hz). From a study of the original spectra themselves, we believed that three dimensions would be required to describe activity within the Alpha band, viz. the median frequency of the 'Alpha bump', which we call "Phi", the amount of activity in a 3-Hz band centered on Phi (i.e., on each S's own median Alpha frequency), and an index of the constancy of Alpha frequency, measured in terms of the kurtosis of the Alpha 'bump', a parameter that we call "Kappa". As a test of this inference, we intercorrelated our six a priori parameters (Delta, Theta, Beta, Phi, Alpha and Kappa) with the scores for the same 140 Ss on the six orthogonal Varimax factors, finding that the six factors accounted for from 92 to 100 percent of the variance in the first five parameters and 72 percent of the variance in Kappa. Obversely, from the multiple regressions of the factors on the parameters, we found that our six rational parameters accounted for 91, 95, 53, 49, and 94 percent of the variance in the first five factors, respectively.

To summarize, the factor analysis tells us that nearly all of the linear common variance in the EEG spectra can be contained within a factor space of only five orthogonal dimensions (Factor VI being very weak and ill-defined). Six (non-orthogonal) parameters (Delta, Theta, Alpha, Beta, Phi and Kappa), more easily translated into standard EEG parlance than the factor scores themselves, were able to account in their turn for about two-thirds of the total spectrum variance. Thus, the analysis both corroborates and extends the traditional practice. It provides an objective, quantitative basis for the definition of Delta, Theta and Beta. It also suggests, however, that Alpha and Phi should not be defined in terms of the entire classical Alpha band, but, rather, in terms

of the alpha 'bump' only. That is to say that not all of the spectrum from 8 to 13 Hz represents alpha activity in a dynamic sense but only that portion within an elevation or 'bump', seldom more than 3 Hz wide, whose center frequency, Phi, is characteristic of the subject and the situation and may vary at least from 8 to 12 Hz.

The "Orderliness" Criterion

Since science is a search for an orderly description of nature, 'good' scientific variables are those which tend to behave in an orderly fashion. Measures of relatively stable properties should not vary greatly over time; thus, retest reliability is a common test of the goodness of a measure, one example of a test based on the orderliness criterion. Identical twins provide a less widely recognized opportunity for the application of the same criterion to the improvement of methods of measurement. Genetically identical and having usually shared very similar environments, a pair of monozygotic (MZ) twins can be thought of as 'parallel forms' of the same individual. As the error variance is reduced in some psychological or physiological measure, the intra-class correlation of that measure within pairs of MZ twins should correspondingly increase. In the case of traits or properties that are thought to be genetically determined, an additional 'orderliness' criterion can be specified, namely that the ratio of the DZ to MZ intro-class correlations will tend toward 0.5.

As an illustration of this use of twin data in methodological research, the six EEG spectrum parameters referred to in the previous section were found to have intra-class correlations ranging above .80 in a sample of 40 pairs of MZ twins. While obviously not a necessary criterion of a 'good' measure, such a finding is sufficient evidence that the parameters in question are reliable and at least potentially meaningful.

More generally, when any variable, x, is known to be related to some other variable, y, then that measure, $f_i(x)$, which shows the most orderly relationship to y, thereby recommends itself as the best among competing methods of measuring x. For example, tonic skin conductance level (SCL) is known to increase generally with CNS arousal; the two-flash threshold (TFT) is known to decrease with arousal. In the set of studies reviewed by Lykken, Rose, Luther & Maley (1966), the average correlation between SCL and TFT was only - 0.44. When a range-corrected index was substituted for raw SCL as the measure of 'electrodermal arousal', this correlation increased sharply to -.67. In a later study (Lykken, 1972), evidence in favor of range correcting measures of tonic SCL and heart-rate, as well as phasic SCRs, was provided by showing that experimental manipulations which were expected to produce differences in all three variables gave substantially larger F-ratios (i.e. smaller residual error) when the range-corrected measures were employed. By a similar test, the phasic heart rate response (HRR) variable was shown not to be improved by range correction.

Individual Differences in Range of Response

Having mentioned the range correction procedure in the examples above, it may be appropriate here to explain the concept on which that procedure is based, since it is pertinent to the topic of this chapter. Suppose one were to ask Ss to indicate the subjective strength of a set of stimuli by squeezing a hand dynamometer, gently for weak stimuli and fiercely for strong ones. Analyzing such data, the most inexperienced investigator would quickly realize that a 20 Kgm squeeze by an amateur wrestler (who can do 100 Kgm without even trying) should not be considered equivalent to the same response from a dainty female S (for whom 25 Kgm is a supreme effort). Instead, one should express each response

from a given S as a proportion of the maximum response of which he is capable. Symbolically, if R_{ij} is the response produced by the ith subject to indicate the intensity of the jth stimulus, and if $R_{i(max)}$ is the largest response that S can emit, then the proper statistic to express his response to this stimulus, independently of individual differences in the range of possible variation of response amplitude, will be:

$$RC_{ij} = \frac{RR_{ij}}{R_{i(max)}}$$

By substituting these RC values for the original raw scores, one can immediately eliminate a large component of error variance, i.e., that portion of the variance across Ss in the raw scores, R_j, which is due to the variation in $R_{(max)}$.

(Note that an ANOVA would separate this component of raw score variance as "variance due to individuals", assuming that we had repeated measures on each S. But if we wanted to correlate, say, stimulus intensity with squeeze-rated subjective intensity we should use the range-corrected scores. Similarly, if we have only a single measure from each S, then one should be sure to obtain the $R_{(max)}$ also and compute RC even when using ANOVA.)

It may be less obvious - but it is equally true - that the same situation obtains with a great many of the dependent variables used by psychologists, including psychophysiologists. One S, for example, will nonchalantly emit SCRs of 5 micromhos each time he clears his throat or someone speaks his name while another, with identical electrode arrangements, could not deliver more than a 3 micromho response falling out of an airplane. Therefore, obviously, an SCR of, say, 1 micromho produced in an experiment is most unlikely to have the same psychological meaning for both individuals. Thus, it behooves an

experimenter using the GSR to deliver one strong, standard stimulus, preferably near the start of each experimental session, which will provide an estimate of $R_{(max)}$ that can be used to range-correct all subsequent GSRs recorded during that session. The maximum SCR of which an S is capable depends upon the density of sweat glands at the electrode site, the ambient temperature, the thickness and condition of the palmar epidermis, and other such factors which vary from person to person but normally have no relation whatever to the psychological processes one is using the GSR to measure; it is this irrelevant variance which the range-correction method is designed to reduce or eliminate.

Some dependent variables show "irrelevant" variation not only in $R_{(max)}$ but also in $R_{(min)}$ (the minimum possible GSR is of course zero for all Ss). Rating scale data, for example, show that some Ss frequently make use of the most extreme rating categories while others almost never do. If one is studying attitudes with an 11-point rating scale ranging from "Utterly Disagree" (1) to "Absolutely Agree" (11), a rating of '8' from an S who rates many other itmes '9', '10' or '11' cannot have the same meaning as an '8' from a more cautious S who never uses higher than a '9'. Such data can be range-corrected by expressing each raw score as a proportion of the range between that S's $R_{(max)}$ and his $R_{(min)}$:

$$RC_{ij} = \frac{R_{ij}}{R_{(max)} - R_{(min)}}$$

Tonic levels of heartrate and SCL are variables of this kind. In an experiment in which the S is severely stressed during one part of the session and allowed to relax to drowsiness during another part, one S may show an SCL while relaxed that is higher than another shows when he is most excited. Individual differences in range of heartrate variation are perhaps not quite so extreme

but they are great enough so that range-correction of tonic heartrate scores will usually prove to reduce error variance also (Lykken, 1972). If the experiment already includes enough variation in demand characteristics so that each S will be caused to display SCL or HRL values near both his minimum and his maximum capability, then one can merely take the smallest and largest values produced as estimates of $R_{(min)}$ and $R_{(max)}$, respectively. In other experiments it may be necessary to modify the plan of the session so as to insure that reasonable estimates will be available. I have found, for example, that one can conveniently elicit adequate estimates of $SCL_{(max)}$ by asking S to blow up a small balloon until it bursts. Since Ss are generally more reactive when they are first hooked up to the apparatus, it is wise to plan to estimate $R_{(max)}$ near the start of any session. Similarly, the only time one can be reasonably sure that most Ss will follow instructions to relax is at the end of the session when they can be assured that "we're almost finished now, there'll be no more tasks or stimuli, so just relax for a few minutes and when I wake you up you can leave."

Individual Differences and Psychological Theory

Individual difference data are useful not only for methodological purposes -- identifying 'basic' variables, improving experimental measures -- but also more generally in the testing of scientific generalizations and hypotheses.

Function Fluctuation. Every experimentalist should be alert to the possibility that his Ss may differ not only in their parametric values for x and y but also even in the nature of the function relating x to y; i.e., some 'laws' may hold for only certain sub-groups of the population. Thus, Venables (1963) reported moderate positive correlations (+.45 and +.61) between palmar skin potential (SPL) and the two-flash fusion threshold in two samples of normal Ss together with strong negative correlations (-.79 and -.72) in two groups of non-paranoid

schizophrenics. Clearly (assuming that these results are replicable facts of nature, although cf. Lykken & Maley, 1968) a similar study based on unselected psychiatric patients might have suggested that these putative measures of cortical and autonomic arousal were in fact unrelated altogether. More recently, Gruzelier & Venables (1972) have found that while some schizophrenics show greater than normal GSR reactivity to non-signal stimuli, coupled with poor habituation, others are grossly hypo-responsive, either habituating too fast or not responding at all. These two types of abnormal response are thought to be related to the psychiatric status or degree of decompensation of the patient. Again, a failure to be alert to the possibility of function fluctuation within the population studied might have prevented the discovery of this potentially important finding. Therefore, it is plain that one should always be careful to scrutinize the univariate and bivariate plots of the data, to determine whether there seem to be systematic differences between those Ss who fall on the positive and the negative diagonals, and the like -- always keeping in mind, of course, that with this additional 'massaging' of the data goes the obligation to test in separate studies any serendipitous findings that may seem to emerge. The worst procedure, on the other hand, is to run the usual mechanical correlations and ANOVAs and then ignore the actual data altogether, under the delusion that one has squeezed it dry.

Correlation vs Manipulation. As already mentioned, it is often possible to study the dependency of y on x in a correlational design, employing the natural variation in x across Ss as a substitute for manipulating x in the experiment. Where x is a fixed structural or constitutional trait variable, this approach may be the only one available. Since it is doubtful that many experimentalists are so moss-backed as to be unaware of this possibility, the emphasis here will

be on the limitations and potential dangers of the correlational approach. As an example, imagine a study in which y is some biochemical measure, x is a self-report measure of, say, anxiety, and x and y are found to be correlated. The usual conclusion would be that plasma-formaldehyde (y) is either a cause or a manifestation of fear; this suggests, in turn, that longitudinal or repeated-measures study of individual Ss would show x and y varying nicely together. But one has only to articulate this corallary to see that it is not necessarily so. In fact, the only time when the R-type or between-subject correlation will properly estimate the P-type or within-subject covariation is when both parameters of the within-subject regression line are quite constant across individuals. Thus, for example, assuming that the range of y can vary considerably across Ss, it is possible for x and y to be perfectly correlated within each individual over time and yet for the usual between-subject correlation to be zero! Alternatively, a high R-type correlation can be associated with P-type correlations ranging about zero. It is probable that most psychological and psychophysiological variables show both 'trait' (between-S) and 'state' (within-S) variation; the two trait components may be correlated and the state components not, or the state components may co-vary and not the trait components, or both of these, or neither. Therefore it follows that a complete study of the relationship of x and y will require both an R-type correlational design and either a P-type longitudinal study or else an experimental manipulation of either x or y (followed, of course, by a repetition of both, preferably in another laboratory, on the principle that the probability that a psychological finding will replicate seems to be inversely related to the intrinsic interest of that finding.)

Those of us who, like myself, are delighted to contemplate the extraordinary and manifold differences between people would be well advised to cultivate the

acquaintance of our dour colleagues for whom these differences are just a nuisance and a source of irritation, for the experimentalist is like the trout fisherman who keeps catching tasty bass and bream and throwing them away. Although he knows a great deal about his favorite variables and has ingenious and precise (and borrowable) methods for measuring and controlling them, it seldom occurs to him to wonder why it is that some Ss consistently score so differently from others, i.e. in what other ways these extremes might differ from each other (indeed, he commonly discards the deviates and forgets about them!). Needless to say, he would be most unlikely to speculate on how his variables might be related to stress or fatigue or extraversion or schizophrenia. Since he doesn't care for bass or individual differences there seems to be nothing immoral about using his methods or his discards for one's own (to him incomprehensible) purposes.

A recent and elegant example of successful poaching of this sort can be found in Sutton (1971) who made a foray into the very heartland of experimental psychology to seize upon the venerable Bunsen-Roscoe Law which states that, below some critical duration, the effect of a stimulus depends upon its total energy without regard to how that energy may be distributed in time; e.g., a 4 mSec. light pulse of intensity 5 will be equivalent to a 2 mSec. pulse of intensity 10. Sutton quite properly realized that the length of this critical duration, within which this time-intensity reciprocity holds, must like any other biological parameter show variation from one individual to another, the only question being whether the degree of such variation is sufficient to be of any interest. Specifically, he proposed that schizophrenics might have shorter critical durations than do non-schizophrenics and this hypothesis, which seems now to have been correct, led to the development of an ingenious test, perhaps the only test yet found on which the schizophrenic can

be shown to perform 'better' than the normal. Briefly, the test requires S to attempt to discriminate between two light flashes, one consisting of a single pulse of 4 mSec and the other a 'package' of two 2 mSec pulses separated by 2 mSec. Discrimination is measured through reaction time (RT) which, since RT decreases with increased stimulus intensity, allows one to determine whether the '2+2' package is being fully integrated so that it is equivalent to the single 4 mSec. pulse. If the normal critical duration allows full integration while the shorter duration for schizophrenics does not, then the schizophrenics will be able to respond consistently differently to the two stimuli while the normals will be unable to do so. In fact (Collins, 1971), all 10 schizophrenics performed 'better' on this test than any of the normals or non-schizophrenic patients used as controls, 7 of the 10 scoring more than 3 SDs above the normal mean.

A Concluding Caveat. We have attempted to show that individual differences are by no means just a nuisance, that they constitute a considerable part of the very raison d'etre of psychological science, and to illustrate a few of the ways in which the methods of the differential psychologist can be of value to the experimentalist. But it must be admitted, I think, that research employing correlations and the study of group differences (where the groups are segregated by their scores on some other variable and are not differently treated) seems to be more chancy and less reliable than that based on the traditional methods of experimental manipulation. I have not attempted an actual survey of the relevant literature but it seems a safe bet that, if one took a random sample of 100 recent articles from JEP or JCPP and another 100 purporting to show that variables X and Y are correlated or that Groups A and B have different mean values of Z, and then replicated each study (preferably by means of constructive

replication, cf. Lykken, 1968), one would find strong evidence that traditional experimental methods lead to more dependable conclusions about the state of Nature. Indeed, one way of distinguishing the experienced researcher from the neophyte (apart from counting gray hairs) is to show them a recent paper reporting high correlations between cortical evoked response latency and IQ or that schizophrenics have less blood-turpentine than normals, and then observe how excited they get. Experience in this field leads (or should lead) to an attitude of open-minded skepticism and one learns not to invest heavily in even the most provocative finding until one has seen it in one's own laboratory (or seen it twice, if the finding is one's own). In respect to one's own research strategy, an approach which recommends itself is the two stage process that begins with a 'pilot' study. It is important to remember that the only intended consumer of the results of this first stage is the researcher himself. The only purpose of the pilot study is to inform and to convince one's self and most of the cumbersome controls and precautions which have become rituals in the field can be set aside except for those which seem essential to uncover the truth of the matter. Then, if the pilot study shows something interesting, one can design the formal study, observing all the rituals, for the purpose of demonstrating to one's colleagues what one has already learned in Stage I. The benefits of this manner of proceding include (1) that, having invested less time and effort in the rough-and-dirty Stage I, one is better able to accept a negative finding in the pilot study for what it is rather than to twist and torture those data until they emit a publishable squeek; and (2) the elegant confirming findings of Stage II are more likely to be true since they are already a kind of replication of the previous result.

REFERENCES

Block, J. The challenge of response sets. New York: Appleton-Century-Crofts, 1965.

Collins, P. Reaction time measures in time intensity reciprocity between psychiatric patients and normals. Paper presented at meetings of the Eastern Psychological Association, 1971, New York.

Gruzelier, J. H. & Venables, P. H. Skin conductance orienting activity in a heterogeneous sample of schizophrenics. Journal of Nervous and Mental Disease. 1972, 155, 277-287.

Jenkins, J. J. & Lykken, D. T. Individual differences. In P. Farnsworth & Q. McNemar (Eds.), Annual review of psychology, Vol. 8. Palo Alto, Calif.: Annual Reviews, Inc., 1957.

Lacey, J. I. Somatic response patterning and stress: Some revisions of activation theory. In M. H. Appley & R. Trumbull (Eds.), Psychological stress: Issues in Research. New York: Appleton, 1967.

Lykken, D. T., Rose, R., Luther, B. & Maley, M. Correcting psychophysiological measures for individual differences in range. Psychological Bulletin, 1966, 66, 481-484.

Lykken, D. T. Statistical significance in psychological research. Psychological Bulletin, 1968, 70, 151-159.

Lykken, D. T. Range correction applied to hear rate and to GSR data. Psychophysiology, 1972, 9, 373-379.

Lykken, D. T. & Maley, M. Autonomic versus cortical arousal in schizophrenics and non-psychotics. Journal of Psychiatric Research, 1968, 6, 21-32.

Nebylitsyn, V. D. & Gray, J. A. (Eds.) Biological basis of individual behavior. New York: Academic, 1972.

Sutton, S. Fact and artifact in the psychology of schizophrenia. In M. Hammer, K. Salzinger & S. Sutton (Eds.), Psychopathology: Contributions from the biological, behavioral, and social sciences. New York: John Wiley, 1971.

Venables, P. H. The relationship between level of skin potential and fusion of paired light flashes in schizophrenic and normal subjects. Journal of Psychiatric Research, 1963, 1, 279-291.