Methodological Issues in Objective Research
with Psychiatric Patients

Samuel Sutton, Ph.D.

Biometrics Research, New York State Department of Mental Hygiene
and

Department of Psychiatry, Columbia University

In 1956, Horwitt recounted a rather sorry history of various assertions, based on research data, of biochemical differences between schizophrenic patients and normals which have proved to be artifacts of differences in diet, differences in psychological state, or differences which arose from comparing institutionalized patients with non-institutionalized normals. Two of the differences which have been reported turned out to arise from differences in the quantities of orange juice and coffee used by the two populations being compared (Kety, 1959).

A comparable consideration is long overdue for differences between patients and normals in objective psychological measures. Here, the primary sources of artifact are probably not diet or institution-wide infections, but a number of psychological states associated either with the illness or with the effects on personality of the institutional environment. In the last several decades, thousands of reports have appeared asserting sensory, perceptual, conceptual, or psychomotor differences between psychiatric patients and normals. It is conceivable that patients are different from normals in all of these functions, but one's suspicions are aroused when the general finding is that the patient performs more poorly than the normal, whatever the measure used. His reaction time is slower; his sensory threshold is less sensitive; his perceptual and conceptual performance is less adequate. Are these, in fact, truly psychomotor, sensory, perceptual, or conceptual differences, or are they secondary side effects of the fact that patients are less motivated, less cooperative, or less attentive than normals during the testing session? Such factors would tend to reduce the efficiency of patient performance on any function tested. Additionally, there is a number of attitudinal states associated with various illnesses. Thus, the conservativeness and rigidity of the depressive patient or the suspiciousness and need for certainty of the paranoid patient, may interact dramatically with threshold measurements in which the degree of conservativeness versus risk-taking will influence the sensitivity of the measure obtained. Similarly, in size constancy experiments such attitudes would give results in the direction of overconstancy. The kinds of questions and doubts I am raising are not logically different from the kinds of criticisms made in the last two decades about the many reports of biochemical differences between psychotic patients and normals.

For certain purposes, the problem is not serious; it is useful to know that reaction time speeds up with clinical improvement (King, 1959) and it might be feasible to use such objective measures as additional indicators of the state of the patient. However, when we wish to use our measures to infer that the nervous system of patients is somehow different from that of normals, i.e., when we wish to use our psychological data to develop inferences about the organic, as opposed to a functional basis of some mental illness, we must know whether flicker fusion thresholds are in fact poorer, or that reaction times are in fact slower, and so on. For this purpose we need to know the source of the poorer performance and should not interpret as a sensory defect what is, in fact, a byproduct of the overall psychological state of the patient. Similarly, it would be incorrect to interpret as a sensory defect behavior which arises from the patient's preoccupation with his hallucinations.

The issues raised here may also have a different kind of importance than the issue of controlling for artifact. In fact, it may be that for the illness under study, poor performance on all tasks implicates in some fundamental way a general state variable such as increased or decreased arousal level, increased neural
noise, reduced neural inhibitory control, etc. There are a large number of such hypotheses which, if found to be true, could account for a generalized decrement of performance and increase in intra-individual variability. The problem here is how to design experiments so as to support one of these hypotheses as opposed to the others, as well as to rule out trivial explanations of the findings.

Still another issue requires clarification. As scientists are materialists, in one sense every performance reflects the state of the nervous system. However, what is meant here by central nervous system differences or organic etiology requires explanation. The issue is best illustrated with a physiological measure. If a particular group of patients show a lower percent alpha in their EEG -- if the record is dominated by low voltage fast activity -- it may mean one of several things. If this is a manic group in a high state of excitement, the predominance of low voltage fast activity may simply be reflecting this state -- since we know that low-voltage fast EEG activity is correlated with focussed attention and mental activity. If this is what the finding means, I would argue that it does not add significantly to our knowledge about physiological etiology in mental illness. On the other hand, if one could show that the low voltage fast activity were primary -- arising perhaps from an excess of some brain metabolite -- then we would have the kind of information which would be relevant to establishing a physiological dysfunction as an etiological agent in the illness. The heightened psychological activity might then be a reflection, rather than the cause of the kind of EEG we are recording. Of course, in this hypothetical example, further steps would be necessary to establish the source of the deviant metabolite. If it were a result of deviant psychological processes, we would be back in the same bind.

The key to the search for a physiological source or etiological agent for mental illness is temporal sequence. One might think of general paresis as a model. While many of the symptoms of the illness may be dependent on the psychological and cultural conditions, the disease does not develop without the prior invasion of the brain by the spirochete. In this example, if some individuals were infected and did not develop general paresis, it would not eliminate the spirochete as an etiological agent; it would then be necessary but not sufficient and such a finding would make it necessary to specify other required conditions. However, the inverse would be more damaging to the hypothesis. If one could ever get general paresis without the presence of the spirochete, this would weaken the hypothesis that the spirochete is the key etiological agent. In this case, the spirochete might be a sufficient but not a necessary cause.

The problem is very similar to the ones involved in the evaluation of the effects of drugs -- which is the subject matter of this study section. To the extent that we wish to use objective psychological measures to evaluate the impact of the drug on the nervous system, it serves no purpose to interpret poorer flicker fusion thresholds or slower reaction times found under the drug condition as a primary alteration of the sensory, perceptual, or motor systems, if, in fact, they are the result of sleepiness or mood alteration. Thus, any solutions which are worked out for more valid comparisons between patients and normals on basic psychological functions should have implications and perhaps direct application to the evaluation of the effects of drugs.

The present status of our knowledge with respect to comparisons of patients and normals on objective measures can unfortunately be summarized by saying that
it is sadly deficient with respect to these problems. It is certainly true that when chronic schizophrenics are tested they give slower reaction times than normals do, but at present we are unable to interpret the meaning of such a finding. It is unfortunate that the majority of studies reporting differences in performance between patients and normals cannot be given precise and specific interpretation. It is the purpose of this paper to examine what strategies may be adopted which would begin to meet the test of the criticisms that I have indicated thus far. I do not intend to undertake a critical review of available findings, but rather to try to summarize the available strategies in psychological research for obtaining findings which could constitute the early steps in the search for a physiological etiology in mental illness.

**Objectivity versus Validity**

One direction of solution must be dismissed at the outset. Objectivity in the classical Watsonian sense, e.g., the use of reaction time as opposed to a verbal response, does not get us much further to our goal. The problem I have posed is one of validity -- what is it that is being measured? What do obtained differences mean? If our problem in interpreting findings with patients is due to motivational or attentional variables, then reaction time is at least as susceptible to influence by these variables as any verbal response -- perhaps more so. A similar caveat must be entered with respect to physiological measures such as the EEG, the GSR, the EMG, the heart rate, etc. All of these measurements are highly influenced by the psychological state of the subject. In a recent critique of human evoked potential research (Sutton, 1969), I have been led by the findings to assert that the most significant source of variability in evoked potential data arises from failure to exert adequate experimental control over the psychological state of the subject. The evoked potential is enormously sensitive to the subject's psychological state.

Actually, in threshold situations, verbal or choice responses are often much better than a measure such as reaction time. These at least permit accuracy indicators as one can not only ask a subject whether a light is flickering, but also inquire of the subject which of several lights is flickering. When the question is in the yes-no form, one must accept the subject's response on faith. Of course, one can use reliability and variability measures to add some conviction to the obtained value, but one is in a much better position with the addition of an accuracy indicator. This permits one to correct for responses such as a statement that the subject perceives flicker, but identifies the wrong stimulus as flickering. As Dember has phrased it, it is the difference between asking a child whether he knows how much two and two is, and asking him for the answer.

As will be developed later, it is not my intent to dismiss objective measures such as reaction time or physiological measures. Rather I will later specify what strategies are necessary to make such responses yield valid data for our purposes.

**Detection and Forced Choice Techniques in Threshold Work**

Dr. Rapoport at this conference has covered the advantages of using a detection theory procedure for measuring thresholds. I only want to add one point about
the temporal forced choice method. Detection theory has formalized the distinction between sensory and criterion measures (Swets, 1964). In traditional psychophysics, except for the use of catch trials, the experimenter is at the mercy of the subject's degree of caution in making a psychophysical judgment. In the method of limits, if the subject feels that he must be quite sure before he is willing to say he detects the stimulus, thresholds will be quite high. On the other hand, if he is willing to take risks, thresholds will be lower. One would like to obtain an estimate of the subject's threshold performance which is independent of such criterion factors. This would obviously be of service in research on psychopathology where anxiety and rigidity are among the attitudinal sources of artifact that we would like to overcome. Detection theorists have worked out the rationale and calculations which permit the obtaining of a separate measure for the sensory factor (d') and for the criterion factor (β). They have shown that under certain conditions d' can be constant despite experimental manipulation of β (the degree of cautiousness of the subject).

In the detection theory yes-no procedure, the stimulus in each trial is either present or absent, and the subject may say either present or absent. The temporal forced choice procedure may be even more useful for work in psychopathology. Here every trial consists of three observation intervals and the stimulus is only present in one of the three. The subject's task is to identify at the end of each trial whether the stimulus occurred in the first, second, or third interval of the trial. While the detection procedure gives a measure of the sensory factor which is independent of criterion, the temporal forced choice does not permit varying degrees of caution. This arises from the fact that the subject must in effect say "present" once for every trial, and there is nothing about the first, second, or third intervals which make choosing them a more or less cautious response. (Of course, if position biases occur, they do weaken this argument. But they can be easily tested for.)

There is another sense in which the role of criterion is minimized in the forced choice technique. In the detection method, the subject must maintain some memory of the stimulus across the testing session and when the perception matches this memory he can say present, when it does not he can say absent. In the temporal forced choice technique, each trial is independent in this sense since the subject is simply identifying which of the three intervals is most different. He knows that the positive stimulus occurs in only one interval of every trial and any perceptual difference may be used to identify the correct interval. Consequently, he is always making a comparison. There is nothing in the structure of the task that requires him to decide how much of a difference is necessary before he can say "present." Finally, the temporal forced choice technique shares with the detection technique the presence of an accuracy indicator. This permits the experimenter to detect and to correct for chance or bizarre performance.

One of the first applications of the power of the forced choice technique to research in psychopathology was made by Clark et al., (1967). Using a method of limits, they were able to replicate the often-reported finding of poorer flicker fusion thresholds in schizophrenic patients. However, by use of the forced choice technique on the same subjects, these differences evaporated. Schizophrenic patients performed as well as normals on CFF. This strongly suggests that the finding of poorer flicker fusion on the part of the schizophrenic patients with the method of limits is due to their conservative bias -- a desire for greater certainty. It is such results that bring into serious question so many of the reported
differences between schizophrenics and normals.

Multiple Values of a Variable

An approach which has generality for a wider range of studies than those involving thresholds is to compare patients and normals at several values on some variable rather than at some arbitrary isolated point. This approach, which is becoming more widely used, is very desirable because it provides additional information and reduces the likelihood of misinterpreting the findings. Thus differences between patients and normals on a two choice reaction time procedure may lead to interpretations that the patients' retardation is due to the choice, to deficient information processing, or to difficulty in discrimination. However, Venables (1968) has used a range from no choice (simple reaction time) to eight choices. He found that the curve for the schizophrenic patients exactly parallels the curve for the normals; it is simply displaced by a constant. The interpretation, therefore, is that an information processing deficit is not supported by these data; rather, the deficit, whatever its source, is constant at all information levels.

The measuring of more than one point or a complete function permits one to detect differences in slope of some part of the curve between groups. This is a more interesting finding but the interpretation is not straightforward. For example, reduced motivation or attention might disproportionately hamper a difficult task more than an easy task. This would give a difference in slope whose meaning is trivial. This question of level of difficulty of tasks to be compared is a critical issue and findings which show patients doing proportionately worse on the more difficult part of the task are open to interpretation in terms of general state variables such as attention or motivation.

Optimally, it is desirable to construct experiments so that those aspects which are to be compared are not salient or even evident to the subject. Otherwise positive findings may arise from differential attitudes between patients and normals to some part of the task. Consequently, when we studied the effect of shift of sensory modality on reaction time in consecutive trials, the stimuli were presented at random and the subject made the identical finger lift response to both light and sound stimuli. There was nothing in the task or the instructions to alert the subject that our interest was in reaction time for sequential trials or in modality shift.

One aspect of the problem which is highlighted in reaction time research is what may be called the level problem. Since Kraepelin we have known that reaction time in schizophrenic patients is slower than in normals. When we use reaction time to test some new hypothesis and obtain positive results, has the new hypothesis actually been supported? Or are the positive findings trivial in that they arise secondarily out of the fact of slower reaction time in the patients?

Another aspect of the level problem is known under the law of initial value (Lacey, 1956). This has been of particular concern in physiological measures. If one group's initial level is closer to the top of the range for that variable, then the application of a treatment or experimental variable which moves the response toward the ceiling may result in smaller effects for the group which is
closer to the ceiling to begin with. For example, is the smaller contraction of
the pupil of schizophrenics to light (Hakerem & Lidsky, 1970) due to the fact that
their pupils are more contracted to begin with?

There are several approaches to the solution of the level problem. In our
modality shift research, schizophrenics respond more slowly than normals both when
the sequence is in the same sensory modality and when the sequence involves shift
in sensory modality (Krieger et al., 1971). Our first impulse is to use difference
scores -- by subtracting ipsimodal reaction time from crossmodal reaction time we
can see whether the residual is greater for the schizophrenics. However, there
are many statistical arguments that have been marshalled against such difference
scores. We have generally used covariance analysis which can be viewed as a method
of statistically equating the groups on the initial variable (ipsimodal reaction
time) in order to assess whether the increment in the experimental variable (cross-
modal reaction time) significantly discriminates between groups.

Drs. Fleiss and Tanur (1971) of our Biometrics Unit have published a system-
atic critique of the use of covariance analysis in these situations. Among their
arguments is that the covariance technique involves extrapolation to values which
are typically beyond the ranges of either group. Such extrapolations would require
that linearity holds over values not observed in the actual samples. Instead of
covariance analysis, they propose that several groups (including schizophrenics and
normals) be tested. The relationship between ipsimodal reaction time and cross-
modal reaction time can be plotted for all groups. The schizophrenic point on the
curve can then be tested to see whether it deviates significantly from the general
trend.

An excellent, but rarely feasible alternative approach to the level problem
is to select and match subjects on the basis of the initial variable (e.g., ipsi-
modal reaction time), and for only these subjects make the comparison between
groups in crossmodal reaction time. This approach has been used occasionally
(Zahn et al., 1961), but we have not found it too practical. The attrition in
sample size is generally too severe to permit the use of this method. Addition-
ally, such selection may introduce a bias. The fastest schizophrenics and the
slowest normals may be atypical for their groups. As a complementary approach it
may be very useful but it would be quite risky to use by itself.

Improving Patient Performance

In contrast to the foregoing, one need not always accept the differences in
level between groups. One may use the strategy of attempting to bring patient
performance up to the level of the normals. This may be considered a negative
strategy, as its main purpose is not so much the finding of differences between
patients and normals, but rather the avoidance of artifactual differences. It has
been shown that by manipulating the motivational situation reaction time differ-
ences between schizophrenics and normals disappear (Rosenbaum et al., 1957), sug-
gestng that the source of the difference between the groups is not the motor or
information processing system. Rather, the difference appears to be motivational
-- under ordinary conditions normals perform optimally, while patients do not.
This approach becomes less of a negative strategy when the experimenter then turns to the questions of examining exactly what has to be altered in order to improve patient performance (e.g., positive versus negative reinforcement) or investigating the causes of the poor motivation in the patients (e.g., the nature of the institutional experience). By learning what variables have been effective in improving patient performance, it may become a powerful source of hypotheses as to the origin and treatment of the illness. This is the approach that has been favored by researchers oriented toward learning theory.

**Superior Patient Performance**

Since the problems with which we are trying to cope are ordinarily associated with poorer performance by patients on all measures, it would add powerfully to the degree of conviction we have in the group differences if measures could be devised in which patients perform better than normals. Such a result could hardly be a trivial byproduct of poor attention or motivation in the patients. We have been working for the last several years to develop such measures and on one of these we have now obtained sufficient data. This measure is based on the well-known Bunsen-Roscoe or Bloch's Law of time-intensity reciprocity for stimuli shorter than some duration known as the critical duration. The law states that below critical duration the determinant of the response is the total energy; it does not matter how the energy is distributed in time. Thus a 4 msec pulse of light of 5 intensity units is equivalent to a 2 msec pulse of light of 10 intensity units. The product, 20, is the same for both, and both of these pulses will yield the same reaction time, or they cannot be discriminated from each other, etc. If, however, either or both pulses exceed critical duration, then this equivalence of response for equal energy stimuli breaks down. Several years ago, I proposed the possibility that schizophrenic patients would have shorter critical durations than normals. We tested and found that at fairly dim intensity levels, normals gave the same reaction time to a 4 and a 2 msec light pulse which were equal in energy, while our schizophrenic patients gave a slightly longer reaction time to the 4 msec pulse than to the 2 msec pulse. This would be consistent with the notion that the patients had a critical duration that was shorter than 4 msec and were not completely utilizing the full 4 msec of the light stimulus in generating the speed of the reaction time response. However, the group differences were quite small, and we were not completely satisfied with the precision and reliability of our control over the intensity of the light flashes. We temporarily abandoned this venture and awaited the completion of several years of systematic work in normals on time-intensity reciprocity by Dr. Kietzman of our laboratory. Last year we felt ready to resume the patient work, and Pat Collins of our laboratory did so using the design shown in Figure 1. The square waves at the bottom of the figure

Insert Figure 1 about here

show the stimuli. We used a 2 msec light pulse at an intensity about one log unit above detection threshold. The second pulse is at the same intensity but is a duration of 4 msec. It might be useful to think of this as two 2 msec pulses with
Figure 1. Hypothetical relationship between reaction time and duration of light pulse package for schizophrenics and normals. The 2 and 4 msec packages are single pulse while the 6 msec package is a double pulse. The dotted line shows the projected duration of a single pulse which would yield the same reaction time as the 6 msec double pulse package. The projection assumes full integration of the 6 msec package for the normals and less than full integration for the schizophrenics.
a zero interval. The third stimulus is two 2 msec pulses of light separated by a 2 msec interval of darkness. These three stimuli are presented in random order and the subject is instructed to lift his finger as rapidly as possible at the occurrence of each stimulus. (Though there are three stimuli it is not a choice reaction.) On the left is a hypothetical curve for normals based on our knowledge that at this intensity critical duration for a reaction time response is about 10 msec. The reaction time response to the 4 msec pulse is more rapid than to the 2 msec pulse since it contains more energy. However, the 2 on - 2 off - 2 on stimulus which has the same energy as the 4 msec pulse yields the same reaction time. They contain the same energy. In passing, I should note that in psychophysical studies with normals these equal energy stimuli cannot be discriminated from each other. The right side of Figure 1 shows the predicted performance for schizophrenic patients if they have critical durations shorter than 6 msec. First, as is well known, reaction time is slower for the schizophrenic patients. However, note that since we predict that critical duration is less than 6 msec for the patients, the reaction time to the 2-2-2 package is slower than to the 4 msec package. This would arise from the fact that not all the energy of the second pulse of the 2-2-2 package is integrated and therefore the patient is reacting as if this package contained less energy than the 4 msec package. This is shown by projecting the 2-2-2 reaction time on to the time axis to obtain an estimate of the duration of a single pulse whose energy is equal to the amount of energy of the 2-2-2 package which has been utilized.

Note that in the predicted result for the patient, we have, despite the patient's slower reaction time, better discrimination for the patient. With his motor response, he is making a discrimination which the normal cannot make, either in reaction time or as a psychophysical discrimination. We suspect, and we are now in the process of checking this, that the patient also cannot discriminate these stimuli psychophysically.

Some of you may be reasoning as follows. The patient's reaction time is slower than the normal's reaction time. Furthermore, the patient reacts more slowly to the 2-2-2 package than to the 4 msec package. How can you call this better performance? My answer is that we know of no way, either by instruction, or by any experimental manipulation to get the normal to produce different reaction times to the two packages. In fact, we know of no response that the normal can use to distinguish these packages. Yet the schizophrenic can distinguish them with his reaction time response.

The other nice feature of this design is that it requires a positive difference for the schizophrenic patient and no difference in the normal. The implication of this difference, under the conditions of this experiment where 1) the response is the same for all stimuli -- finger-lift, 2) the stimuli are presented in random sequence, and 3) the differences between stimuli are probably not available to awareness, is that poorly motivated performance would tend to lead to greater variability and to the reduction of the likelihood of obtaining significant differences between the 4 msec package and the 2-2-2 package. In other words, a patient not performing the task optimally would make him appear more like a normal -- not responding differentially to the stimuli. This feature, therefore, increases our confidence in the validity of positive findings.
In Figure 2 the actual data for three groups: schizophrenic patients, non-schizophrenic hospitalized psychotic patients, and normals are presented. Each subject is shown as a vertical bar. On the ordinate is the projected duration of a single pulse which would yield the same reaction time as the 2-2-2 stimulus package. The bars which are hatched represent subjects in whom the reaction times for the 4 msec package and for the 2-2-2 package are significantly different (p < .05) by the Mann-Whitney U Test. Note that only among the schizophrenic patients are these two packages ever significantly different and that this is always in the direction of a slower reaction time for the 2-2-2 package than the 4 msec package, i.e., shorter critical duration. This is true for the majority (6 out of 10) of the schizophrenic sample.

I wish to note that all of the patients were under phenothiazine treatment and we are currently looking at which phenothiazine and the various dosages that were administered. Several other controls with respect to drugs are also in process.

The Iterative Method

The iterative method is not new in science nor new to research in psychopathology. The purpose of giving it a name and developing it explicitly is to enter a plea for a more extensive use of iterative approaches. The problem to which we are addressing the iterative method arises from the diagnostic imprecision and the overlap among diagnoses that now exists in psychopathology. The investigator attempting to use precise and objective research measures must deal with the problem of specifying the diagnostic group in which he expects his measures to detect differences. If a measure results in average differences between schizophrenics and normals, but there is a high degree of overlap between the groups -- is this because the measure is not very good, or is it because even with the most careful diagnosis, schizophrenia consists of several diagnostic entities and normality conceals many undetected pathologies? Further, the researcher may often ask himself whether his measures may be of use for improving diagnostic precision. These questions may be approached via the iterative method. In essence the method involves successive alternating in the specification of which is the dependent and which is the independent variable.

Figure 2 shows that 40 percent of the schizophrenic group did not show a significant difference between the equal energy packages. We first examined whether this was because these patients had more variable reaction times, i.e., were they the patients who performed least optimally? As it turns out, this is not so. Alternatively, are these, perhaps, the patients who are less ill, or more importantly, is it possible that these patients are a different subgroup of schizophrenics. In this study all the subjects had been given a structured interv
Figure 2. Distribution of equivalent single pulse packages which would yield the same reaction time as the 6 msec double pulse package for schizophrenic, non-schizophrenic psychotic, and normal subjects. The dashed line is the 4 msec single pulse, equal in energy to the 6 msec double pulse package. A cross-hatched bar indicates a statistically significant difference (p < .05) on the Mann-Whitney U Test between the 6 msec double pulse package and the 4 msec single pulse of equal energy.
(the CAPPS scale developed by Dr. Spitzer of our Biometrics Unit) which produces ratings on 30 symptom scales (Spitzer & Endicott, 1969). We compared the two groups of schizophrenic patients, those with significant differences on the equal energy stimuli with those who do not show significant differences. On 28 of the 30 scales the ratings of the two groups were essentially the same, varying only at random. However, on a scale entitled Speech Disorganization, patients who had a high rating were the ones who showed significant differences on the equal energy packages, while those who had a low rating were the ones who did not show significant differences. The other scale, Auditory Hallucinations, showed a similar but somewhat weaker difference between the two schizophrenic groups. Now two out of 30 comparisons may turn up by chance, but here is the strength of the iterative method. We can now undertake a new study in which subjects are selected in terms of their ratings on these scales and see if the critical duration differences hold up. Our hypothesis now is that only those who have high ratings on these symptoms will show the reduced critical duration.

It should be noted that the Speech Disorganization scale of the CAPPS is probably misnamed. Our probing of the rating criteria used suggest that the scale taps what is normally called thought disorder or the degree to which sequential utterances are relevant to each other or appear to flow logically.

If the first iterative step is successful, other iterative steps can be followed. What I mean by the first iterative step is that, having found (without a prior prediction based in theory) that only the schizophrenic patients with the low critical duration were rated as high on thought disorder, we should then select a sample based on the best measures of thought disorder currently available, and test for critical duration in this sample as compared with other subjects. If our initial results are now essentially replicated, we can subsequently use critical duration as an independent criterion for cases of thought disorder and begin to refine the first measures of thought disorder in order to form a clearer diagnostic cluster with respect to which we may then begin to discover other physiological and psychological characteristics. For example are these the patients who show familial incidence of psychopathology; do they show differences on physiological and biochemical measures; do they respond differently to drugs; will their time-intensity reciprocity change in other sensory modalities; what are other psychological measures which characterize this group, etc.? By proceeding along such paths, it may become possible for the objective investigator to contribute to the problem of diagnosis in psychiatry as well as to move in the direction of developing good etiological hypotheses.

We have not yet taken the first iterative step in our critical duration study. Our first venture using the iterative procedure was done in the context of our studies of the ease or difficulty of shifting sensory modality in sequential reaction time trials (Kriegl et al., 1971). In that study, Dr. Kriegl of our laboratory began with patients specified as schizophrenic by hospital diagnosis and normals who were obtained from the usual sources -- hospital attendants, students, etc. The two samples chosen in this way showed a significant difference. The schizophrenic patients were disproportionately retarded in their reaction time by shift of sensory modality. However, Dr. Kriegl also used an objective structured diagnostic interview. According to the interview results, only half of the schizophrenic sample received a diagnosis of schizophrenia and only half of the normal sample received a diagnosis of normality.
When the two purified groups were compared -- the "interview" schizophrenics with the "interview" normals -- the modality shift differences between groups were heightened. Furthermore, despite the reduction of each original sample by half, the inter-individual reaction time variability in the "purified" groups was significantly reduced. In this instance the first step in the iterative procedure has added conviction both to the structured interview diagnosis and to the modality shift measure as being associated with schizophrenia.

Construct Validation

The iterative method is a special case of the general approach of construct validation. The iterative method is most useful in the special situation where one variable is difficult to specify precisely (diagnosis) while the variable which can be defined precisely (objective measure) depends for its relevance on its association with the imprecisely defined variable. The construct validation approach, however, is more general. It involves the use of converging operations -- the process of making and testing additional inferences in order to interpret a finding properly. For our critical duration experiment, the use of a two-pulse package of light may lead some to attempt to interpret our finding as having to do with temporal resolution and not with temporal integration. Therefore, our next step will be to test whether a single pulse package of the necessary duration predicted from our current experiment will yield the same reaction time. If for some individual, we extrapolate that the 2-2-2 package is giving a reaction time level which is equivalent to a 3 msec single pulse, for this individual we can make a direct reaction time comparison between the two stimuli. If the level is different, then the temporal integration interpretation of the finding is weakened.

There are a number of other properties of temporal integration which can be used to further validate our inference. For example, beyond critical duration there is a region of partial integration in which duration is still relevant but is no longer equivalent to intensity. Finally, there is a duration usually referred to as utilization time beyond which only the intensity of the stimulus is relevant. These further characteristics of the temporal integration curve provide other means of testing that schizophrenics are indeed different from normals in the way they process energy over time.

In the modality shift experiments referred to above, we have been involved in a 10 year process of construct validation; even so, we do not as yet have a single clear interpretation of the findings. But the review of the sequence of experiments makes a good illustration of the construct validation approach.

In the first of these experiments, the subject was instructed to make a choice reaction time response to each of a sequence of light and sound stimuli presented in quasi-random order. It was found that his reaction time is influenced by the relationship between the stimulus to which he is responding and the stimulus of the previous trial. When the sensory modality of the stimulus to which the subject is responding is different from the sensory modality of the stimulus in the previous trial (crossmodal sequences), the reaction time is longer than when the two consecutive stimuli are in the same sensory modality (ipsimodal sequences). The basic finding is that while this lengthening or retardation of reaction time
when modality is shifted is found for both normal and schizophrenic subjects, it
is greater for the schizophrenic subjects. (Sutton et al., 1961)

Dr. Hakerem of our laboratory who designed the initial experiment in this
area (Sutton et al., 1961) was interested in testing a neurological model
proposed by Mettler (1955) which suggested that some of the perceptual findings
in schizophrenia might be explained by dysfunction of neural structures involved
in environmental scanning and in shifting the organism's attention from stimuli
in one sensory modality to stimuli in another sensory modality. To get closer
to a test of this hypothesis, we eliminated the choice aspect of the reaction
time and had the subject simply perform the same finger lift to all stimuli. The
finding remains the same: schizophrenics are more retarded in shifting sensory
modality. To confirm that shift in sensory modality, and not just difference
between stimuli produced the retardation, we used at this stage two different
stimuli in each modality. We found that although both schizophrenics and normals
have a slight retardation due to intramodality difference, this ipsemodal
retardation is not significantly different between schizophrenics and normals.

However, our next several controls experiments did not add much conviction
to the Mettler hypothesis (Sutton & Zubin, 1964). We reasoned that the amount of
retardation caused by what amounts to a "sticky switch" should decrease with
increased interval between stimuli. While findings were in this direction,
lengthening the interval to 10 seconds did not reduce the retardation strongly
or systematically. We also found that the retardation is reduced with practice,
an effect not predicted by the Mettler model.

Our next step was to see whether we could increase the differences between
schizophrenics and normals by making the switch "stickier." By increasing the
number of repetitions in the same modality before shifting to the other modality,
we thought we might increase the crosemodal retardation and reduce the overlap
between schizophrenics and normals. Again, we found that while differences
between patients and normals remain, the repetition before shifting did not
increase the clarity of the results.

At this point we began to lean toward a more global concept of "set" or
"expectancy." However, even from the beginning we were uncomfortable with this
as an explanatory concept. In Hakerem's initial experiment, a choice reaction
time design was used -- the subject had to move his finger to one key for light
and to another key for sound. In this situation, one might propose differences
in motor set as a possible explanation for increased reaction time in the
patient group when modality is shifted. If the patient makes some assumption as
to which key he is to move and the stimulus signals that he should move to the
other key, one might conceive that the occurrence of the unexpected stimulus
interrupts and lengthens the time of the prepared reflex. Therefore, in sub-
sequent experiments we asked subjects to make an identical finger-lift response
regardless of whether the stimulus was a sound or a light. Hence, there is no
stimulus-specific motor set to be disrupted by an unexpected stimulus. To retain
the expectancy explanation, one must resort to the notion of a sensory set --
even though it is not intuitively apparent by what mechanisms one can prepare to
sense with the ear as opposed to the eye. Perhaps descending inhibitory pathways
in the afferent systems could be involved in such selective readiness.
These questions relating to differences in set between schizophrenic patients and normals have been tested in our more recent work (Kriegl et al., 1971). We reasoned that since the schizophrenic seemed to be more retarded by a shift of sensory modality in consecutive trials, it would seem that the patient, more than the normal, must be making the assumption that the next stimulus was likely to be in the same sensory modality. As we used a random stimulus sequence and the next stimulus was just as likely to be in the other modality, the patient would in effect be receiving more unexpected stimuli than the normal -- resulting in greater average reaction time retardation.

While there is no objective way of knowing exactly what a subject expects, we used a guessing response prior to each stimulus presentation to provide some index of expectancy. We asked the subject in each trial, prior to stimulus presentation, to place his finger on one of two keys. One key was clearly labeled sound and the other key was clearly labeled light. The key on which the subject placed his finger indicated his guess as to whether sound or light would be the next stimulus. At the occurrence of the stimulus (which was determined by the experimenter on the basis of a prearranged random sequence independently of the subject's guess) the subject lifted his finger as rapidly as possible. For the reaction time response, no choice was required of the subject. Regardless of whether the stimulus was a sound or a light, and regardless of whether the stimulus confirmed or disconfirmed his guess, the subject's task was simply to lift his finger as rapidly as possible.

The results militated strongly against the expectancy or set explanation. First, while the patients had a mild ipsimodal bias in their guessing, the normals had a much greater ipsimodal bias -- exactly opposite to what would be required for set to be the explanation of the greater crossmodal retardation for schizophrenic patients. Second, the patients did not make more wrong guesses. Third, the reaction time retardation was not due to increased reaction time for wrong guesses, because the modality shift difference between patients and normals occurs only in the reaction times associated with right guesses.

Finally, we did one control that should have been done 10 years earlier. We tested the modality shift in a situation in which there is no uncertainty. Prior to each stimulus the subject was told by the experimenter what the next stimulus would be. It seemed to us that if incorrect expectancies had anything to do with the explanation of the difference, this condition would eliminate it altogether. The patients showed a significantly greater crossmodal retardation than the normals even in this condition. So we are now back with some version of the "sticky switch" hypothesis. Zubin (1970) has recently attempted to construct a new model along these lines. Our current experiments involve looking at scalp recorded, evoked potentials in schizophrenics and normals in relation to shift of sensory modality.

The above line of work illustrates the long and still incomplete process of construct validation which is necessary for interpretation of a finding even when the finding itself is completely reliable.
Summary

I have tried to review the available strategies for designing experiments in psychopathology which permit meaningful and valid interpretations. I have been particularly concerned with the kinds of psychological experiments which may contribute to organic etiological formulations. The topics I have considered are:

1. The use of threshold procedures which eliminate or control criterion variables and permit an estimate of true sensory sensitivity.

2. The measurement of complete parametric functions and the initial level problem.

3. Bringing patient performance up to the level of the normals where motivational and attitudinal factors are at fault.

4. Finding measures on which patients perform better than normals.

5. The iterative approach to the problem of diagnostic imprecision.

6. The role of construct validation in the interpretation of findings.

Most of these are not alternative so much as complementary approaches.
References


Zubin, J. Problem of attention in schizophrenia. Presented at the meetings of the American College of Neuropsychopharmacology, 1968, San Juan, Puerto Rico.