CHAPTER 10

Fact and Artifact in the Psychology of Schizophrenia

SAMUEL SUTTON

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

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 rather 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. Although it is conceivable that patients are different from normals in all of these functions, one's suspicions are aroused when the general finding is that the patient performs more poorly than the normal, whatever the measure used. The patient's reaction time is slower; his sensory threshold is less sensitive; his perceptual and conceptual performance is less adequate. Are these all, 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? Certainly such factors would tend to reduce the efficiency of patient performance on any function tested. Additionally, a
number of attitudinal states are 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 (Raush, 1952). 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, 1969), and such objective measures may prove valuable 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, that is, when we wish to use our psychological data to develop inferences about an 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 secondary by-product of the overall motivational state of the patient. Similarly, it would be incorrect to interpret as a sensory defect behavior that arises from the patient’s preoccupation with his hallucinations.

The issues raised here may have implications beyond the problem 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, or reduced neural inhibitory control. 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 intraindividual variability. The problem, however, 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 shows a lower percentage of alpha in their electroencephalograms (EEGs)—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 pre-
dominance of low-voltage fact activity may simply be reflecting this state, since low-voltage fast EEG activity is known to be correlated with focused 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 we could show that the low-voltage fast activity were primary—arising perhaps from an excess or deficiency of some brain metabolite—then we would have the kind of information that 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. Although many of the symptoms of the illness may be dependent on psychological and cultural conditions, the disease does not develop without prior invasion of the brain by the spirochete. In this example, if some individuals were infected and did not develop general paresis, the spirochete would not be eliminated as an etiological agent; it would then be a necessary but not sufficient condition, and such a finding would necessitate the specification of 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 effects of drugs. To the extent that we wish to use objective psychological measures to evaluate the impact of a drug on the nervous system, it serves no purpose to interpret the 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 that are worked out for more valid comparisons between patients and normals in regard to basic psychological functions should have implications and perhaps direct application to the evaluation of the effects of drugs.

Unfortunately the present status of our knowledge with respect to comparisons of patients and normals on objective measures can 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. The purpose of this chapter is to examine what strategies may be adopted which would begin to meet test of the criticisms that I have indicated thus far. I intend, not to undertake a critical review of available findings, but rather to try to summarize the available strategies in psychological research for obtaining findings that 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, for example, the use of reaction time as opposed to a verbal response, does not get us much nearer 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 galvanic skin response, the electromyogram, and the heart rate. All of these measurements are strongly 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, to which, the evidence shows, the evoked potential is enormously sensitive.

Actually, in threshold situations, verbal or choice responses are often much better than a measure such as reaction time. The former at least permit accuracy indicators, as one can not only ask a subject whether a light is flickering but also inquire 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, reliability and variability measures can be used to add some conviction to the obtained value, but one is in a much better position with the addition of an accuracy indicator. This makes it possible 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
attempt to specify what strategies are necessary to make such responses yield valid data for our purposes.

DETECTION AND FORCED-CHOICE TECHNIQUES IN THRESHOLD WORK

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 that 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 that 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 that permit the obtaining of a separate measure for the sensory factor ($d'$) and for the criterion factor ($\beta$). They have shown that under certain conditions $d'$ can be constant despite experimental manipulation of $\beta$ (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 must say whether it is present or absent. The temporal forced-choice procedure may be even more useful for work in psychopathology. Here every trial consists of three (or more) observation intervals, and the stimulus is present in only one of them. Whether the stimulus is in the first, second, or third interval is random. 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 can give a measure of the sensory factor that is independent of the subject’s criterion, the temporal forced-choice procedure 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 interval that makes choosing it a more or less cautious response. (Of course, if the subject shows a consistent preference for one of the intervals, the argument is weakened. But such position biases 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; when the perception matches this memory he can say “present,” and 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 intervals is most different. He knows that the positive stimulus occurs only one interval of every trial and that any perceptual difference may be used to identify the correct interval. Consequently, he is always making a comparison. Nothing in the structure of the task requires him to exercise some criterion in order 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, Brown, and Rutschmann (1967). Using a method of limits, they were able to replicate the often-reported finding of poorer flicker fusion thresholds in schizophrenic patients. However, when the forced-choice technique was used on the same subjects, these differences evaporated—schizophrenic patients performed as well as normals. This strongly suggests that the findings of poorer flicker fusion on the part of schizophrenic patients with the method of limits is due to their conservative bias, that is, their 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 that has generality for a wider range of studies than those involving thresholds is to compare patients and normals at several values of 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 normal persons on a two-choice reaction time procedure may lead to interpretations that the patients' slower reaction time is due to the choice, to deficient information processing, or to difficulty in discrimination. However, Venables (1958) used a range of conditions from no choice (simple reaction time) to eight choices. He found that the curve for the schizophrenic patients exactly parallels that 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 of a complete function permits one to detect differences between groups in the slope of some part of the
curve. This would be an interesting finding, but the interpretation is not straightforward. For example, reduced motivation or attention might hamper the difficult portion of the task disproportionately more than the easy portion. The result would be a difference in slope whose meaning is trivial.

Optimally, it is desirable to construct experiments so that the 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 toward 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 as a function of sequence or in modality shift.

One aspect of the problem of comparing patients and normals which is highlighted in reaction time research is what may be called the level problem. Since Kraepelin we have known that reaction time is slower in schizophrenic patients than in normals. When we use reaction time to test some new hypothesis and obtain positive results, can we therefore assume that the new hypothesis has 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 ceiling of the range for that variable, then the application of a treatment or experimental variable that moves the response toward the ceiling may result in smaller effects for the group that is closer to the ceiling to begin with. For example, is the small contraction of the pupil of a schizophrenic in response to light (1...sky, Hakerem, & Sutton, 1971) due to the fact that the pupils of these patients 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 a shift in sensory modality (Kriegel, Sutton, & Kerr, in press). Our first impulse is to use difference scores—by subtracting ipsi-modal reaction time from crossmodal reaction time, we can see whether the residual is greater for the schizophrenics. However, many statistical arguments 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 (ipsi-modal reaction time) in order to assess whether the increment in the experimental variable (crossmodal reaction time) significantly discriminates between groups.

Fleiss and Tanur (Chapter 28) of our Biometrics Unit have made a
systematic critique of the use of covariance analysis in these situations. Among their arguments is that the covariance technique involves extrapolation to values that are typically beyond the ranges of either group. In extrapolations would require that linearity hold 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 crossmodal 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., ipsimodal reaction time), and for only these matched subjects make the comparison between groups in crossmodal reaction time. This approach has been used occasionally (Zahn, Rosenthal, & Shakow, 1961), but we have not found it very practical. The attrition in sample size is generally too severe to permit the use of this method. Additionally, such selection may introduce a bias. The fastest schizophrenics and the slowest normals may be atypical for their groups. As a complementary approach this may be very valuable, but using it by itself would be quite risky.

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 normal individuals. This may be considered a negative strategy, as its main purpose is not so much to find differences between patients and normals, but rather to avoid artificial differences. It has been shown that by manipulating the motivational variation reaction time differences between schizophrenics and normals disappear (Rosenbaum, MacKavey, & Grissel, 1957), suggesting 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, whereas patients do not.

This approach becomes less of a negative strategy when the experimenter then turns to the question of examining exactly what has to be altered in order to improve patient performance (e.g., positive versus negative reinforcement) or of investigating the causes of the poor motivation in the patients (e.g., the nature of the institutional experience). By indicating what variables have been effective in improving patient performance, this approach may become a powerful source of hypotheses as to the origin
and treatment of the illness. It 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 one's degree of conviction in regard to group differences if measures could be devised in which patients perform better than normals. Such a result could hardly be a trivial by-product 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 a constant response: there will be the same reaction time, or they cannot be discriminated from each other, and so on. If, however, either pulse or both of them 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

and a 2-msec light pulse which were equal in energy, whereas our schizophrenic patients gave a slightly longer reaction time to the 4-msec than to the 2-msec pulse. This would be consistent with an interpretation that the patients had a critical duration shorter than 4 msec and therefore were not completely utilizing the full 4 msec of the light stimulus in generating the speed of the reaction time response.

The group differences were quite small, however, and we were not completely satisfied with the precision and reliability of our control over the intensity of the light flashes. Therefore we temporarily abandoned this venture and awaited the completion of several years of systematic work with normals on time-intensity reciprocity by Mitchell Kietzman of our laboratory. Last year we felt ready to resume the patient work, and Patrick Collins did so, using the design shown in Figure 1. The square waves at the bottom of the figure show the stimuli. We used a 2-msec light pulse at an intensity about one log unit above detection threshold. The second
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 double pulse. The dotted line shows the projected duration of a single pulse that 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.

pulse is at the same intensity but has a duration of 4 msec. It might be useful to think of this as two 2-msec pulses with 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 the critical duration for a reaction time response is about 10 msec. The reaction time response is more rapid to
the 4-msec pulse 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 1-sec pulse, yields the same reaction time; they contain the same energy.

In passing, I should note that in psychophysical studies with normals using verbal reports 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 the critical duration is less than 6 msec for the patients, the reaction time is slower to the 2–2–2 package 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 onto 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 that has been utilized.

Note that in the predicted result for the patient we have, despite his slower reaction time, better discrimination. With his motor response, he would be making a discrimination that the normal person cannot make, either with 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.

One other issue should be clarified. The reaction time of the patient is slower than that of the normal. Furthermore, the patient reacts more slowly to the 2–2–2 than to the 4-msec package. How could one call such a finding better performance? The 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.

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 requirement, 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 reduction of the likelihood of obtaining significant differences between the 4-msec package and the 2–2–2 package. In other words, a patient's not performing the task optimally would make him appear more like a normal—not responding differentially to the stimuli. This feature, therefore, would increase our confidence in the validity of positive findings.
The actual findings (Collins, 1972) were in accord with the predictions. All of the 10 normals tested gave essentially equal reaction times to the two equal-energy packages. Of the 10 schizophrenics tested, 7 were significantly outside the limits of normal variability for equal reaction time for the two equal-energy packages. They gave longer reaction time to the 2-2-2 than to the 4-msec package—the finding that would be predicted by the hypothesis of shorter critical duration. By contrast, none of a third group, who consisted of hospitalized nonschizophrenic patients, gave significantly longer reaction times to the 2-2-2 package.

It should be noted that all of the patients were under phenothiazine treatment, thus suggesting that the finding in the schizophrenic group is not a drug effect. We are currently looking at which phenothiazine was administered, as well as the various dosages used. Several other controls with respect to drugs are also in process.

THE ITERATIVE METHOD

The iterative method is not new in science or 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 so because the measure is not very good, or because, even with the most careful diagnosis, schizophrenia consists of several diagnostic entities and normality conceals many undetected pathologies? Furthermore, the researcher may often ask himself whether his measures may be of use for improving diagnostic precision. These questions can 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.

In our study of critical duration 3 of the 10 schizophrenic patients did not show a significant difference between the equal-energy packages. We first examined whether the reason was that these 3 patients had more variable reaction times, that is, were they the patients who performed least optimally? As it turned out, this was not so. Alternatively, were these, perhaps, the patients who were less ill, or, more importantly, was it pos-
sible that these patients constituted a different subgroup of schizophrenics? In this study all the subjects had been given a structured interview (the Present and Past Psychopathology Scale—CAPPs—developed in our Biometrics Unit) which produces ratings on 30 symptom scales (Spitzer & Endicott, 1969). We compared the two groups of schizophrenic patients, the 7 with significant differences on the equal-energy stimuli and the 3 who did 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 did not show significant differences. The other scale, Auditory Hallucinations, showed a similar but somewhat weaker difference between the two schizophrenic groups. Now 2 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 whether 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 suggests 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 shorter critical duration were rated as high on thought disorder, we would 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 results are replicated, we can subsequently use critical duration as an independent criterion for cases of thought disorder and begin to refine the initial 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 the patients with shorter critical duration and thought disorder also those 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 in other sensory modalities also be different; what other psychological measures characterize this group? 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 dur study. Our first venture using the iterative procedure was carried out in the context of our studies of the ease or difficulty of shifting sensory modality in sequential reaction time trials (Kriigel et al., in press). In that study, Kriigel began with patients specified as schizophrenic by hospital diagnosis and with normals obtained from the usual sources—hospital attendants, students, and so on. 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, Kriigel et al. 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 interindividual reaction time variability in the “purified” groups was significantly reduced. In this instance, the first step in the iterative procedure added conviction both to the structured interview diagnosis and to the modality shift measure as being associated with schizophrenia.

Still other applications of an iterative approach are cited by Gurland in Chapter 22 of this book.

**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 that 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. However, such an interpretation is in conflict with our earlier finding of longer reaction time for the schizophrenic patients to a 4-msec light flash than to a 2-msec light flash which was equal in energy (by being more intense).
The study in the same patients of temporal processing in other sensory modalities, as well as the use of physiological measures such as the evoked potential and the pupillary response, will permit still further delineation of the exact nature of the differences. Furthermore, such converging operations may permit the specification of the physiological basis of the findings.

A number of other properties of temporal integration 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. We can therefore investigate whether partial integration and utilization time are also altered in patients who display a shorter critical duration. Furthermore, we are repeating these studies with a verbal report discrimination rather than reaction time. In normals, we know that critical duration and utilization time are longer for verbal report than for reaction time. We can thus check whether for the schizophrenic patients shorter critical duration and utilization time will be obtained for verbal report as well. All of these characteristics of temporal integration could provide additional means of testing whether schizophrenics are indeed different from normals in the way in which they process energy over time.

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 that may contribute to organic etiological formulations. Six topics have been considered:

1. The use of threshold procedures that 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 in cases 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 so much alternative as complementary approaches.
ACKNOWLEDGMENTS

I am indebted to Patrick Collins, Jean Endicott, Barry Gurland, Mit Grossberg, Gad Hakerem, Muriel Hammer, Mitchell Kietzman, Jewell Kriegl, Joseph Fleiss, Karen Olson, and Robert Spitzer for assistance with various aspects of this work. The research was supported under Grants MH-11688 and MH-07776 from the National Institute of Mental Health, United States Public Health Service.

REFERENCES

Kety, S. S. Biochemical theories of schizophrenia. Part II. Science, 1959, 129, 1590–1596. (b)
Sutton, S. The specification of psychological variables in an average evoked potential experiment. In E. Donchin, & D. B. Lindsley (Eds.), Average


