CHAPTER 7

Experimental Methods in Research in Psychopathology

L. ROWELL HUESMANN

The legitimate approaches to empirical research practiced by clinical psychologists fall into three major groupings: true experiments, quasi-experiments, and observational studies. Each type has advantages and disadvantages for the clinical researcher, and each presents different technical problems. The failure of researchers to master these technical considerations can result in deficient studies from which valid conclusions cannot be drawn.

Investigators conducting empirical research on psychopathology aim to infer population characteristics from sample characteristics. Sometimes an investigator may wish to estimate a population parameter. More commonly, though, the researcher wants to test an hypothesis about the relations among variables. In some cases he or she may be satisfied with determining whether or not two variables are *stochastically independent,* but more frequently he or she would like to determine much more precisely the functional relation between the variables. In particular the researcher may want to test the hypothesis that a certain *causal model* of the relationship between the variables is true. Most of the inferences psychologists made about populations fall into these three classes: (1) simple parameter estimation, for example, what proportion of the population has schizophrenia? (2) testing dependency hypotheses about the relation between two variables, for example, do schizophrenics have slower reaction times than normal persons? (3) testing causal hypotheses about the relation between variables, for example, is a defect in attention a cause of slower reaction times in schizophrenics? Similarly, the statistical procedures used in such studies can be divided into two broad classes defined by the primary sample statistic used: tests on means or tests on correlations. Tests on means are generally most appropriate for true experiments, somewhat less appropriate for quasi-experiments, and least appropriate for observational studies. Tests on correlations are used most frequently in observational studies, less frequently in quasi-experiments, and least frequently in true experiments. Yet either class of analyses may be appropriate in many situations. A common misconception is that correlational analyses cannot be used to test causal models. Most scientists are taught very early in their careers that a correlation (i.e., stochastic dependency) between variables does not imply causation. However, this does not mean that correlational analyses can never be used to test causal models. In fact, with certain types of data, that is, longitudinal observational data, special types of correlational analyses provide good tests of causal hypotheses.

It is common to hear methodologists talk about the *validity* of a particular study. Validity, as used in such a context, means somthing slightly different than when one talks about the validity of a variable. Methodol-

ogists are talking about the truth of the conclusions of a study when they refer to the validity of a study. Campbell and Stanley (1963) have distinguished between two types of validity for the conclusions of studies: internal validity and external validity. *Internal validity* is the degree to which the conclusions of the study are justified within the narrow context in which the study was performed. *External validity* is the degree to which the conclusions are generalizable to other contexts. Internal validity is necessary but not sufficient for external validity. Consider the case of a scientist who measures the concept learning ability of a sample of schizophrenics, administers a drug, remeasures their concept learning ability, and finds it greatly enhanced. Such a study would have low internal validity. Even the narrow conclusion that the drug facilitated concept learning in schizophrenics would be unjustified, since the improvement could have been due to many factors including retesting, maturation, or an unknown factor intervening in time. Suppose, however, the experimenter had tested a comparable *control group,* administered a placebo to them, retested the control group, and found that their concept learning ability was enhanced significantly less. Now the experimenter has a much more internally valid study, since he or she has eliminated many alternative explanations for the result by adding a control group. The conclusion that the drug enhances concept learning is more justifiable. However, the external validity of the study is limited. One cannot really say whether the drug would enhance learning in subjects suffering from other disorders or enhance learning on other tasks.

This chapter focuses on the methodology for conducting "experimental" studies of psychopathology. By an "experimental" study we mean a study in which one or more variables are *manipulated* (the independent variables), and the effects on other variables (the dependent variables) are observed. Such manipulative studies can be subdivided along a dimension of increasing validity for causal inferences into preexperimental, quasi-experimental, and true experimental designs. Preexperimental designs have virtually no potential for testing causal hypotheses. Probably the most widely used paradigm of this type has been the pretest-manipulation-posttest design without any control group. A great deal of psychotherapy research has been of this type, with a client's improvement from pretest to posttest being accredited to the therapy manipulation. The threats to the validity of such a causal conclusion (alternative causal explanations) include the uncontrolled between-test history of the subject, improvement due to retesting, motivation, subject mortality, statistical regression upward (spontaneous recovery), the interaction of subject selection with the treatment, and other factors. Fortunately, journal editors no longer smile on such research, and it is being seen less often.

True experiments, on the other hand, possess the greatest validity for causal inferences. The sine qua non of a true experiment is a control group that is functionally equivalent to the treatment groups on all relevant dimensions except the one being manipulated. While perfect equivalence is an impossible ideal, some procedures for achieving equivalence are much better than others. Without exception the most acceptable procedures all involve a *random* assignment of observations to treatments. There is *no* completely adequate substitute for randomization. Unfortunately, randomization is not always possible because of ethical considerations, institutional policies, subject limitations, or other reasons. Furthermore, even perfect randomization cannot counteract the effect of confounding factors introduced with the manipulation.

A quasi-experiment is defined here as a manipulative study that does not meet the rigid standards required of true experiments, but whose design permits some causal inferences to be drawn, though with less validity than in true experiments. In the canonical

case the experimental and control group are nonequivalent, perhaps because of limitations on who is acceptable for the experimental group, for example, requirements of skill, volunteering, pathology, or other characteristics. As a result, the manipulation is confounded with preexisting subject attributes. However, also included in the quasi-experimental category are the differential studies in which groups differing on one or more attributes such as psychopathology are exposed to a common experimental manipulation. While the group differences are orthogonal to the manipulation, the problems encountered are similar to those of other quasi-experiments. Probably the most frequently employed experimental design in clinical research is the quasi-experiment in which one performs the same manipulation on a pathological sample and a control sample presumed to be equivalent except for the lack of pathology. But there is no guarantee that the two samples are equivalent except for the pathology. Even when there are no obvious confounding factors such as medication, age, hospitalization, or socioeconomic status, there is no guarantee of equivalence. Furthermore, the manipulation may interact with subject selection within each population to cause confounding within as well as across the samples. For example, certain segments of a pathological population may be systematically eliminated from an experimental group because they cannot function sufficiently well, while they are accepted in a control group (Kisz & Parsons, 1979). In such differential studies one must remember that the distinction between the pathological and normal group is an *observed* not manipulated distinction. As a result, the validity of any conclusion about the pathology's effect on the manipulation can never be strong and depends critically on the elimination of alternative correlated group differences.

One of the best ways to understand the relative merits of a particular experimental or quasi-experimental design is by examining three components of each design: how the subjects were selected, what manipulation was performed, and how the data were analyzed. While the components are obviously interrelated, the advantages and disadvantages of the possible alternatives within each design component can be discussed relatively independently.

SUBJECT SELECTION AND ASSIGNMENT

The procedure used in selecting the samples is one of the most critical determinants of both the external and internal validity of a study. The extent to which experimental and control groups can be considered equivalent depends upon how they are selected, and the generality of any conclusion depends upon how representative the samples are of the population. Yet surprisingly many researchers neglect to specify exactly *how* the sample was chosen, especially in quasi-experimental studies. It is not sufficient to state every characteristic of the sample with great precision; one needs to designate exactly *how* the sample was selected from the population of interest. It may be interesting to know that the college-student subjects in a drug experiment were all seniors with no history of mental illness and a mean IQ of 120; but it is more important to know whether they were chosen randomly from the senior class, randomly from volunteers who responded to an advertisement, were the first students to reply to an advertisement, or were chosen in some other way.

Random Assignment from a Single Population

The most powerful approach for testing causal hypotheses requires the researcher to randomly select subjects from a single population and then randomly assign them to the different experimental conditions. Unfortunately, such a technique is often impractical

for the researcher of psychopathology. Nevertheless, when possible, random selection and assignment allow the strongest causal conclusions to be drawn. Experiments of this type using only abnormal subjects are rare, since one generally wishes to compare abnormals with normals. Nonetheless, a few studies of schizophrenic cognition (Chapman & Knowles, 1964), many experiments on hypnosis (e.g., Spanos & Bodorik, 1977; Spanos, Ansari, & Stam, 1979), and a few other experiments have employed randomly assigned abnormal subjects. More commonly, random assignment has been used with normal populations to select subjects for treatments in which abnormal states would be induced. The abnormal states might be induced physically with alcohol (Abrams & Wilson, 1979; Polivy, Schueneman, & Carson, 1976; Wilson & Lawson, 1978), drugs (Pihl & Sigal, 1978), or food (Ruderman & Wilson, 1979); or they might be induced through behavioral manipulations as in studies of mood states (Velten, 1968; Polivy & Doyle, 1980), hypnosis (Banyai & Hilgard, 1976; Coe, Basden, Basden, & Graham, 1976), learned helplessness (Hiroto & Seligman, 1975), anxiety (Glad & Adesso, 1976) and schizophrenic associations (Meiselman, 1978).

Random assignment of subjects to conditions can be accomplished in at least three different ways. First, each subject selected from the population may be assigned at random to one experimental condition independently of the assignment of any other subject (*independent groups design*). Second, subjects may be paired (or grouped into triplets, quadruplets, etc.) on the basis of possessing similar attributes (e.g., IQ and SES). Then one member of each matched *k*-tuple would be assigned at random to one of the experimental conditions. Such *randomized block designs* have the advantage that much of the between-subject variance is removed from the error variance. One particular type of randomized block design deserves special mention. The *yoked control* design increases the equivalence of the paired subjects by assuring that they receive exactly the same exposure to irrelevant characteristics of the stimuli. For example, in many learned helplessness studies each experimental subject is paired with a control subject and is given the same number of reinforcements as the control subject earns (Cole & Coyne, 1977; Seligman & Maier, 1967). Yoking each experimental subject's reinforcements to a control subject's usually eliminates reinforcements as a confounding variable (for some potential problems with yoking, see Church, 1964, and Costello 1978).

The third approach to random assignment involves letting each subject appear in each experimental condition with the order of presentation randomized and counterbalanced (Chapman & Knowles, 1964; Teasdale & Fogarty, 1979). These *repeated measures designs* allows between-subject variance to be removed from the error variance. Unfortunately, other error variance may be introduced because of "testing" effects. Repeated measures designs are probably overused in research on psychopathology.

While random assignment is the one technique that assures equivalence of subject groups, it has not been utilized as frequently in clinical research as have other less demanding procedures. Random assignment is not possible in many settings, and where possible, often is not used because the experimental conditions of interest are existing subject attributes such as existing pathologies. Thus quasi-experiments appear more frequently than true experiments in clinical research.

Selection from Different Populations

By far the most common procedure for subject selection in experimental studies of psychopathology has involved drawing some subjects from a pathological population and others from a control population. Over 75% of the experimental articles published in the *Journal of Abnormal Psychology* between

1976 and 1979 were of this type. In most studies of schizophrenic cognition, the effect of a variety of manipulations on schizophrenics' memory processes, reaction times, and perceptual processes has been compared with the manipulation's effect on control groups consisting of normal subjects or subjects with another pathology (Koh & Peterson, 1978; Knight, Sherer, Putchat, & Carter, 1978; Bellissimo & Steffy, 1972). Similarly, the differences between the effects of various manipulations on pathological subjects and their effects on control subjects have been measured to investigate social and cognitive processes in depression (Smolen 1978; Sacco & Hokanson, 1978; Miller & Lewis, 1977) psychopaths (Hare, 1978; Steinberg & Schwartz, 1976), alcoholics (Brown & Cutler, 1977; Tucker, Vuchinich, & Sobel, 1979), and those suffering appetitive disorders (Elman, Schroeder, & Schwartz, 1977; Spencer & Fremouw, 1979). These studies are quasi-experimental because the major "independent" variable, pathology, is a subject attribute, and other subject differences may be confounded with it (see Chapman & Chapman, 1977, for good examples). The manipulation is sometimes also confounded with populations, but more often is performed within both subject groups. For example, in studies comparing reaction time in schizophrenics with reaction time in normals, the preparatory interval is manipulated over a range of values (Bellissimo & Steffy, 1972); so the interaction of preparatory interval with pathology can be examined.

Often researchers comparing a pathological and normal population attempt to increase the equivalence of their samples by *forming matched pairs* across the different populations (e.g., schizophrenic and normal) and eliminating unmatched subjects from the experiment. This procedure should not be confused with the matching of subjects in randomized blocks experiments. There, subjects are assigned randomly to conditions after matching. Here, their population determines their treatment condition. As a result, the matching may introduce more problems than it remedies. Matching subjects' observed scores may seem to increase the equivalence of the groups, but the subject's underlying "true" scores may be quite different if the matching variable is less than perfectly reliable. If the two populations have substantially different means on the matching variable (as is often the case), the matching procedure selects subjects unrepresentative of the populations, because most of those who match will fall between the means. Further, regression effects are likely to operate in opposite directions and make the groups seem to move apart on the matching variable and on any correlated variable (e.g., the dependent variable). Generally the statistical corrections discussed later in the chapter are preferable to such matching.

Selection from Population Extremes

Another mode of subject assignment frequently employed in clinical research requires the selection of the extreme scorers on a particular test. For example, in much of the learned helplessness research on humans, two groups of subjects are selected and contrasted: those who score high and those who score low on a depression inventory (Miller & Seligman, 1975; Rizley, 1978). While such an approach may seem very similar to the comparison of the analogous clinically differentiated groups, the similarity is based on an assumption that psychopathology falls on a continuum. Many question this presumption (Depue & Monroe, 1978). Are college students who score above the median on a depression inventory similar in essential characteristics to clinically depressed patients? Certainly the external validity of the conclusions drawn from such studies is weaker than for those drawn from studies employing clinical populations. One must also be careful in this kind of quasi-experiment to avoid the matching errors discussed earlier with regard to clinically differentiated populations. While one may be

tempted to match the subjects selected from the two extremes on such variables as IQ and SES, it is likely to harm more than it helps for the reasons discussed above.

One type of selection from the "extremes" that is particularly prevalent is the *volunteer* experiment. It is common for the subjects in one experimental condition to represent a population that volunteered for a therapy, treatment, or procedure for which the subjects in the other condition did not volunteer. For example, one might compare women who come to a clinic, ask for, and receive assertiveness training with control women selected from a class in introductory psychology. Why are such volunteers considered to be from an extreme? It is assumed that the volunteers selected themselves for the treatment on the basis of their true assertiveness scores, which were probably below average for the population. While it would be possible to measure, estimate, and statistically correct for the volunteers' initial assertiveness level, it would be preferable to divide the volunteers into an early treatment and late treatment group randomly. The late treatment group could then serve as a control (waiting list control) for the early treatment group. In experimental studies of psychopathology, one should be suspicious of any design in which one group consists of volunteers and the other does not.

MANIPULATIONS

The manipulations commonly performed in experimental studies of psychopathology are diverse enough to defy easy classification. Nevertheless, a few general categories stand out.

Induction of Abnormal States

One broad category of manipulations are those clearly designed to induce special psychological states or modes of behavior. These states may be of interest because they occur often among the population (e.g., inebriation), because they are analogous to certain pathological states (e.g., learned helplessness), or because they provide insight into certain psychological processes (e.g., hypnotic trance). The states may be induced with physiological or behavioral procedures.

Among the most common of such experiments are those employing drugs, food, or alcohol. Alochol–no-alcohol manipulations have been used to investigate alcohol's effects on social behavior such as aggression (Lang, Goeckner, Adisso, & Marlatt, 1975; Zeichner & Pihl, 1979) and sexual behavior (Wilson & Lawson, 1978), to investigate its influence on certain cognitive processes such as memory (Rosen & Lee, 1976), and pain (Brown & Cutler, 1977) and to investigate the etiology of alcoholism (Brown & Williams, 1975; Marlatt, Demming, & Reed, 1973; Tucker, Vuchinich, & Sobell, 1979). Similarly, drug manipulations have been used to simulate abnormal behavior such as hallucinations (Heaton & Victor, 1976; Klee, Bertino, Weintraub, & Calaway, 1961; Metzner, Litwin, & Weil, 1965). In experiments using such physiological manipulations, it is particularly critical to control subject expectations. The best manipulations appear to be those that vary actual substance dose and expected substance dose orthogonally. The control group receives a placebo that is indistinguishable from the "real thing," and instructions are carefully designed to manipulate what the subject expects to receive. For example, in most alcohol studies the drinks are made bitter enough that alcohol cannot be detected. Half the subjects in each group are thus told they are getting alcohol and half told they are not. The need for such an expectancy manipulation has been amply demonstrated by the numerous experiments finding significant expectancy effects (e.g., Abrams & Wilson, 1979; Polivy, Schueneman, & Carlson, 1976). Of course the physiological effects of a substance administered in a large dose may be detected by the subject; so posttest measure-

ments of subjects' attributions are a necessity.

A more difficult problem associated with using substances to induce psychological states concerns the equivalency of the states across subjects. The same dose may have quite different effects on different subjects. While varying dose as a function of body weight may reduce the differences, it may not eliminate them. Probably the best solution is to employ more than two dose levels so the effect of dose on behavior can be estimated. Still another problem is presented in these studies by the ethical necessities of substance research. All subjects must be informed ahead of time of what substances they may receive, and must be given a chance to withdraw. This introduces two potential threats to the validity of an experiment: (1) certain classes of subjects may be less likely to agree to participate; and (2) subject expectancies may predispose them to certain behavior. For example, in Ruderman and Wilson's (1979) experiment on eating behavior, it appears that the ethical warning that subjects would be asked to consume a high caloric drink predisposed subjects to behave as if they already had received the drink.

Abnormal psychological states may be induced by other than physiological means. The subject may be exposed to certain environmental situations or contingencies designed to elicit behavior analogous to that observed in certain pathologies. The learned helplessness induction procedure is perhaps the best recent example of such a manipulation. Be exposing a subject to a situation in which his or her reinforcements are noncontingent on behavior, a state of helplessness can be induced that is putatively analogous to depression (Blaney, 1977; Abramson, Seligman, & Teasdale, 1978; Hiroto & Seligman, 1975). Mild mood staes may be induced by even simpler procedures such as reading lists of affectively toned words (Teasdale & Fogarty, 1979), while states of heightened aggressiveness may be induced by frustration paradigms (Buss, 1966). Sensory and sleep deprivation have been used to engender psychological states in which hallucinations, anxiety, or disconnected thoughts are displayed (Cohen, 1979; Dement, 1960; Heaton & Victor, 1976).

For anyone using either physiological or nonphysiological induction techniques to investigate a pathology, a major concern should be how similar the induced state is to the real state. Too often analog states are treated as if they are real states when general conclusions are enumerated. A nonalcoholic who has been given alcohol may behave quite differently than an alcoholic; a helpless subject may behave quite differently than a depressed subject; and the hallucinations experienced in states of sensory deprivation may be quite different from those experienced by schizophrenics. But what if they do appear to be the same? That does not prove the analog either. The analog and real subject may behave similarly for quite different reasons. This is not to say that analog research is not valuable. It is valuable because it is the only experimental research on psychopathology allowing the manipulation of psychological states and hence random assignment of subjects. Still the conclusions of analog experiments must be supported by the results of other kinds of research on pathological subjects before they gain substantial credibility.

One other widely used induction technique deserves special mention: hypnosis. Experimental studies utilizing hypnotic induction have focused most often on characteristics of the hypnotic state itself rather than on its similarity to any psychopathology. The exact technique used to hypnotize a subject has varied greatly, but most frequently subjects selected for their suggestibility are exposed to tape recorded instructions for entering a trancelike state and engaging in certain behavior. The susceptibility of such subjects to obeying instructions has provided unusual opportunities to gather data on a number of kinds of behavior and processes including

memory (Coe, Basdin, Basdin, & Graham, 1976; Kihlstrom, 1980), tolerance for pain (Spanos, Radike-Bodorek, Ferguson, & Jones, 1979), perception (Miller & Leibowitz, 1976) and logical thought (Sheehan, Obstoj, & McConkey, 1976). However, a major difficulty in interpreting the results of hypnosis experiments arises because it is unclear whether the behavior observed is more a function of the induction or simply of the subjects selected. The best design would seem to be one in which subjects are randomly assigned to the hypnosis and no hypnosis group either from the general population or the "hypnotically susceptible" population. However, this procedure is not commonly used, because many researchers believe that "susceptible" subjects will be hypnotized no matter what the procedure. Unfortunately, these researchers do not seem to realize that a valid experiment is impossible under such an assumption.

While most experiments in which special psychological states are induced are aimed at understanding specific states or psychopathologies, one type is aimed more at understanding variations in cognitive processes across states. These are the state-dependent learning studies. In such experiments, the subject learns certain responses while in the special state (e.g., inebriated). Then the subject's performance is measured in the same state and the normal state (i.e., sober). The evidence from these experiments suggests that the reimposition of the state active during learning enhances performance (Eich, Weingartner, Stillman, & Gellin, 1975; Weingartner, Miller, & Murphy, 1977). This general approach of comparing the cognitive processes of the same subject in different states appears to have substantial potential for the understanding of cognitive processes in psychopathology.

Mitigation of Pathological Behavior

The methodology for research on therapy techniques is discussed elsewhere in this book. Therefore only a few general comments about the experimental paradigms most often employed will be made here. First, one must be careful about generalizing the power of a manipulation. A technique that inoculates college students against the effect of a learned helplessness induction (Jones, Nation, & Massad, 1977; Teasdale, 1978) may be ineffective in preventing clinical depression. Second, natural improvement over time and subject attrition must always be carefully considered as alternative explanations for any improvements. Usually a random assignment to treatment and control group is impossible, but control must be accomplished in some way. A waiting group control is often the best solution. In any case, control groups in such studies should always receive placebo treatments with demand characteristics equivalent to those found in the experimental group. With small numbers of subjects the best approach may be a time series design with repeated introductions of the manipulation and measurements of the behavior. If enough observations are made, even the data from a single subject can be statistically significant. What suffers is the external validity (generalizability) of the results. Time series designs are discussed in more detail in the section on data analysis.

Interaction of Pathology with Cognitive or Social Processes

As mentioned earlier, it is very common for researchers of psychopathology to compare the responses of normal and abnormal subjects to a particular manipulation. Most often the manipulation is directed at a particular cognitive or social process believed to be of relevance to a psychopathology. The objective of such an experiment is usually twofold: to learn more about the pathology by examining how the abnormal subjects perform compared to normals, and to learn more about the cognitive or social process by examining how it is affected by the pathology. In these experiments the manipulation is performed on both the normal and abnormal groups, so the outcome of interest is the in-

teraction between group and manipulation. With schizophrenics, for example, one popular paradigm has been to study reaction times as a function of the interval between a warning signal and the stimulus signal. The duration and regularity of the preparatory interval are manipulated factorally within the schizophrenic and normal groups, and the interaction of these manipulations with the group (schizophrenic-normal) are examined. These experiments led to the discovery of the "reaction time corss-over effect" for regular and irregular intervals in schizophrenics (Bellissimo & Steffy, 1972). Numerous other paradigms from experimental psychology have been employed to investigate cognitive processes of schizophrenics and depressives in much the same way. Among these have been experiments on recall, (Koh & Peterson, 1978; Raulin & Chapman, 1976), recognition (Miller & Lewis, 1977), memory scanning (Koh, Szoc, & Peterson, 1977), organization of memory (Russel & Beekhuis, 1976), visual search (Russel & Knight, 1977), iconic memory (Knight, Sherer, & Shapiro, 1977), sentence verification (Neufeld, 1977), interference in short-term memory (Bauman & Kolisnyk, 1976), attention (Chapman, 1956; Straube & Germer, 1979; Oltmans, 1978), and concept formation (Payne, Mattusek, & George, 1959). Many of these manipulations were originally motivated more by a general interest in schizophrenic cognition than by particular theoretical expectations.

The manipulations employed in investigation of depression have had a firmer theoretical grounding, probably because of the existence of theories suggesting effects of certain social and cognitive processes. Many manipulations have been stimulated by the learned helplessness paradigm described earlier. One popular approach has been to induce learned helplessness in groups differing on level of depression (Miller & Seligman, 1975; Price, Tryon, & Raps, 1978). Unfortunately, the importance of these experiments has been overemphasized. A comparison of learned helplessness effects in depressed and normal populations is not a strong test of the learned helplessness model of depression. A causal variable may have a quite different effect on a pathological population after the pathology emerges than before it emerges.

Recently some researchers studying depression have attempted to manipulate and measure attributions about success and failure among depressed and normal subjects (Dweck, 1975; Rizley, 1978; Krantz & Hammen, 1979; Seligman, Abramson, Semmel, & von Baeyer, 1979). Attributions are generally manipulated by giving a subject false information about himself or herself and the task he or she is performing, for example, "this task clearly relates to intelligence." Because one cannot directly observe attributions, most of the evidence regarding the success of such manipulations has been inferential. However, a few psychologists have also either asked subjects to report what attributions they made (retrospective verbalizations) or asked subjects to think aloud while they are making attributions (concurrent verbalizations). Nisbett & Wilson (1977) have pointed out the weakness of retrospective verbalization as data, but concurrent verbalizations can be valid and useful (Ericsson & Simon, 1980). Unfortunately, the distinction between retrospective and concurrent is often missed.

While the above classification of manipulations is by no means exaustive, it does provide a framework in which most experiments and quasi-experiments on psychopathology can be placed. Let us now turn to the types of analyses most commonly employed in psychopathology research and to a more formal classification of experimental designs. Table 7.1 summarizes this classification.

ANALYSIS OF DATA FROM EXPERIMENTAL AND QUASIEXPERIMENTAL DESIGNS

The analysis of data from experimental studies of psychopathology differs from the analysis of other experimental data only in that

Table 7.1. Summary of Commonly Used Experimental Designs for Research on Psychopathology

Design	*Subject selection and assignment*	*Manipulation*	*Recommended analyses*	*Some recent examples*
True Experiments				
Independent groups	Randomly select subjects from a single population. Assign subjects randomly to treatments.	Treatments differ only in value of independent variable.	Posttest only: ANOVA or MANOVA	Polivy et al., 1976 Meiselman, 1978
			Pre–post test: ANOVA or MANOVA on difference scores or with test of time by treatment interaction.	Abrams & Wilson, 1979 Sacco & Hokanson, 1978
			Intergroup comparison of regressions.	Bachrach et al., 1977
Randomized blocks	Randomly select subjects from a single population. (a) Match subjects into blocks of k subjects and randomly assign one from each block to each of the k treatment groups. (b) Or randomly pair subjects and randomly assign one of each pair to the experimental and "yoked" control group.	Treatments differ only in value of independent variable. In "yoked" control design, each subject's manipulation is matched to his or her controls except for value of independent variable.	Same as for independent groups except that *randomized blocks* ANOVA or MANOVA is used.	Knox & Shum, 1977 Cole & Coyne, 1977 Hiroto & Seligman, 1975 Seligman & Maier, 1967
Counterbalanced repeated measures	Randomly select subjects from a single population.	Each subject receives each treatment but in a different order; so order can be removed as a potential explanation for treatment effects	MANOVA with the subjects score under each treatment as a separate dependent variable (or, if assumptions are not violated, a repeated measures ANOVA)	Chapman & Knowles, 1964 Teasdale & Fogarty, 1979

Quasi-Experiments				
Nonequivalent independent groups	Subjects are selected randomly from different populations, or from the extremes of one population, or in some other way so groups differ in essential characteristics.	Confounded: The manipulation is confounded with the group differences.	Post test only: None Pre–posttest: ANCOVA, or ANCOVA with reliability correction in which pretest is covariate, or ANCOVA on standardized change scores without pretest as covariate.	Kenny, 1965
		Factorial: The manipulation is orthogonal to the group differences. Within groups, treatments differ only in value of independent variable.	Same as independent groups under true experiments except measured group differences should be used as covariates in ANCOVA, if assumptions are not violated.	Bellissimo & Steffy, 1972 Koh & Peterson, 1978 Chapman, 1976
Matched groups	Subjects are selected from two different populations or from the extremes of one population so they *match* on one or more variables believed to correlate with the dependent variable.	Confounded: Same as nonequivalent independent groups. Factorial: Same as nonequivalent independent groups.	It is recommended that these designs be avoided because the effect of matching on observed scores is uncertain. If used, analyses would be similar to randomized blocks design.	Miller et al., 1979 Straube & Germer, 1979
Time series	Randomly select subjects from a single population.	Each subject receives the treatment one or more times and is measured on the dependent variable many times before and after each treatment	Autoregressive time-series analysis or, as a first approximation, a comparison of before and after regression of the dependent variable on time.	Kazdin & Wilson, 1978 Ayllon & Axrin, 1967

the preponderance of studies are quasi-experiments rather than true experiments. Nevertheless, the starting point for any discussion of analysis issues must be the true experiment.

Independent Randomized Groups Design

Posttest Only. While not the most commonly occurring design in clinical research, the independent groups posttest-only design serves as a good starting point for a discussion of analysis issues. Subjects from a single population are assigned at random to treatment cells and measured after the treatments are performed. If p independent variables are manipulated and each has r treatment levels, such a design will require $p \times r$ groups of subjects (cells). For maximizing the power of the analysis, one would like to have an equal number of subjects in each cell, but this may not always be possible. Recent examples of this approach include Polivy, Schueneman, & Carson (1976), and Meiselman (1978). Discussion of the numerous variations of this basic design is beyond the scope of this chapter; rather I will deal with only a few specific issues that have sometimes arisen when the independent groups design is used in experimental studies of psychopathology.

Usually the principal hypothesis tested in an independent groups design is that the population means for the different levels of the independent variables are equal, and the principal technique applied is analysis of variance. Frequently the experimenter has measured a number of dependent variables in each cell. Unfortunately, a common approach for this situation has been to compute separate analyses of variance for each of the dependent variables. Such a method is deficient on two grounds: (1) it raises the probability of at least one Type I error (rejecting the null hypothesis when it's really true) from α to about about $1 - (1 - \alpha)^v$ where v is the number of dependent variables;[1] and (2) the probability of a Type II error (not rejecting the null hypothesis when it's really false) is higher than for alternative techniques if the dependent variables are correlated. The appropriate procedure with multiple dependent variables is a multivariate analysis of variance (MANOVA) followed by tests of the individual variables.

A second common error concerns the appropriate model for the ANOVA or MANOVA. A fixed-effects model is legitimate for an independent variable whose levels represent exaustively the range of levels of the variable (e.g., hospitalized or not hospitalized), while a random-effects model is appropriate for an independent variable whose levels represent a few arbitrary or randomly selected values from a range of values (e.g., Hospital A or Hospital B). The choice of a fixed-effects model, which most computer programs use as a default, when random effects are appropriate, can lead to an inflated Type I error rate (Clark, 1973; Martindale, 1978).

A third type of error may be observed when the experimenter utilizes a design requiring tests of the interactive effect of the independent variables on the dependent variables. Too often only the F-statistics for the interaction and main effects are examined in these experiments. But *if an interaction is significant,* tests of the main effects for the variables involved in the interaction are misleading. Instead one should test the *simple effects* of each variable within each level of the other variables in the interaction. Care must be taken with such procedures so that the probability of a Type I error is not inflated unacceptably because too large a number of

[1]Alpha (α) is the significance level chosen for each test by the experimenter. If, for example, the experimenter decides to reject a hypothesis whenever the probability of his or her being wrong is less than .05, then Alpha is .05. But the probability of *at least* one error in 10 such tests (i.e., on 10 dependent variables) will be much higher than .05. In fact, according to the formula, it will be about .40.

F-statisitcs have been tested. A similar problem occurs when one attempts to determine which levels of a manipulated variable produced any significant main effect. While a discussion of the various techniques for overcoming the multiple comparison problem is beyond the scope of this chapter (see Morrison, 1976; Miller, 1966), even a cursory survey of clinical experimentation reveals that researchers need to be more cognizant of the problem.

Pretest–Posttest Designs. In deciding whether to measure the dependent variable before and after the treatment or only after, two factors must be balanced. The main advantage of the pre–post design is that the within-subject error variance can be estimated separately from the between-subject variance. As a result, the statistical tests generally are more powerful (i.e., have lower probability of a Type II error) when both pretreatment and posttreatment measurements are made. The effect of the treatment can be measured by comparing the change from pretest to posttest in the experimental groups with the change in the control groups relative to the within-subject variance (Abrams & Wilson, 1979; Sacco & Hokanson, 1978). One can perform a one-way analysis of variance directly on the difference scores of the groups, or one can test the treatment by time interaction with a two-factor repeated measures analysis of variance model. The procedures generate identical results.

One should remember that the advantage of the difference score analysis stems from the fact that posttest and pretest scores are substantially correlated. If they are not highly correlated (e.g., $r^2 > .25$) for some reason, then the difference score analysis may be less powerful than a posttest-only analysis. Usually, however, the scores are highly correlated.

While the greater statistical power of pre–post designs is desirable, it must be balanced against a potential deficit—the pretest may substantially affect the posttest. Such a testing artifact is particularly serious if it raises the dependent variable's mean close to a ceiling, or if the magnitude of the testing effect depends upon the treatment the subject receives (treatment by testing interaction). For example, in experimental studies of memory, a pretreatment recall test may both enhance posttreatment recall for all subjects and sensitize experimental subjects to an objective of the study (recall changes) before the subjects receive the treatment. The effect of such a treatment might be quite different if a posttest-only design were used.

One other quite different type of analysis is also appropriate for pre–post independent groups designs in which the researcher is primarially interested in the effect of a manipulation on the relation between two psychological variables, for example, the effect of setting on the relation between mood and performance. If both variables of interest are measured before and after the manipulation, one can examine how the manipulation affected their relation by comparing the multiple regression equations predicting postmanipulation scores from premanipulation scores in the experimental group with the corresponding equations in the control group. One of the variables may be the manipulated variable, or they may both be unmanipulated. The advantage of this approach is that it can reveal both how the manipulation affected the relation between the variables and how the variables influenced the power of the manipulation. While this technique has been used widely by economists (see Johnston, 1972, for details), it has only been employed infrequently by psychologists (Bachrach, Huesmann, & Peterson, 1977).

Randomized Blocks Designs

In a randomized blocks design with k treatment levels, subjects are formed into sets of k people (blocks) on the basis of similarities (matching). Then the k subjects in each block are randomly distributed among the k treat-

ment groups. Thus every subject in treatment condition *i* can be paired with a subject in any other condition *j* during the data analysis (Cohen, 1979; Cole & Coyne, 1977; Hiroto & Seligman, 1975; Knox & Shum, 1977). The value of this procedure depends upon the extent to which blocking reduces error variance in the analysis of the results. Error variance is partitioned into "between block" variance and "within block" variance. If the blocking accounts for a substantial portion of variance, the error term in an analysis of variance will be reduced, and the test of the null hypothesis will be more powerful than for the corresponding independent groups design. While the "yoked control" version of the randomized blocks design is often used in clinical research (Seligman & Maier, 1967; Cole & Coyne, 1977), other types of randomized blocks designs are less frequently seen (although they are recommended in therapy outcome research, Kendall & Norton-Ford, this volume). It is worth reemphasing that a randomized blocks design is a valid true experimental design and is quite different from the experiments that use matched subjects from two different populations. When the subject's characteristics determine the subject's treatment, matching is not very valuable; but when matching can be followed by random assignment (as with a randomized blocks design), it can be a powerful tool.

With a randomized blocks design, one can use a posttest-only procedure or a pretest-posttest procedure. The same arguments apply for the advantages and disadvantages of pretesting as with independent groups designs. Similarly, the analysis of variance procedures are not greatly different, and the cautionary statements outlined above apply here as well.

Counterbalanced Repeated Measures Designs

In these designs every subject receives every treatment but in a different order (Chapman & Knowles, 1964; Teasdale & Fogarty, 1979). The ordering of the treatments must be carefully counterbalanced to eliminate order as a potential explanation. The repeated measures design looks similar to a pre–post design, but differs in that every subject receives every treatment; so there are no separate experimental and control groups. With pre–post designs the important analysis of variance effect is the time by treatment interaction (i.e., differences in change scores between groups). In repeated measures designs it is the main effect for treatment level. The traditional linear model for conducting an analysis of variance of a repeated measures design requires one to assume that the correlations between the repeated measurements on the dependent variable are the same for every pair of measurements, for example, placebo versus dose 1 is the same placebo versus dose 2 is the same as dose 1 versus dose 2, and so on. Since this assumption is often seriously violated, it is preferable to treat each measurement as a separate dependent variable in a multivariate analysis of variance (MANOVA). With such an analysis one examines the hypothesis of equal means for all pairs of treatment levels simultaneously (Morrison, 1976).

Quasi-Experimental Nonequivalent Independent Groups Design

By far the most common type of design in experimental studies of psychopathology is the nonequivalent independent groups design. In this design samples are selected from different populations for the same or different treatments. For example, clustering in free recall might be compared between hypnotically susceptible subjects and normal control subjects. If the only question were whether the populations differed in their scores on the dependent variable, the statistical test could be a simple one-factor analysis of variance. But such a design is not really even a quasi-experiment, since no variables were manipulated. The designs of more interest occur when another independent variable is either manipulated identically within each group

(*factorial manipulation*), or manipulated across groups so its levels are confounded with population (*confounded manipulation*). An example of the first kind would be experiments on schizophrenic cognition in which a sample of schizophrenics and normals receive the same manipulation (Bellissimo & Steffy, 1972; Chapman, 1956; Koh & Peterson, 1978). The hypothesis to be tested is whether the pathology (schizophrenia vs. normal) caused the manipulation to have a different effect, that is, is there a group by treatment interaction? Examples of the second kind of experiment would be studies of hypnosis in which the hypnotized and control group differed on susceptibility (Stevenson, 1976), studies of therapist effectiveness in which therapist-patient pairs are not randomly assigned (Gomes-Schwartz, 1978), studies of cognitive processes in which subject pathology is confounded with drugs received (Russel & Beekhuis, 1976), or studies in which the experimenter does not control who gets the treatment (Dennis, 1960; Feldman-Summers et al., 1979). The question of interest in these experiments would be whether the manipulations caused a difference between the groups, that is, is there a main effect for therapy? Of course, both factorial and confounded designs can be generalized to have several grouping variables (e.g., psychopathology, sex, hospitalized status) and to have several manipulated variables. In all cases, however, these designs cannot be called true experiments because the difference between groups is due to a preexisting condition that might be correlated with many other variables. As a result, one can never answer the causal questions posed with as much confidence as when differences are produced by an experimental manipulation across equivalent groups (i.e., randomly selected).

As with randomized independent group experiments, an important decision for nonequivalent group quasi-experiments is whether one uses both a *pretest and a posttest* or *only a posttest*. The arguments are similar and clearly favor the pretest-posttest approach for quasi-experiments in which population is *confounded* with treatment. If one does not measure preexisting differences in these designs, it is almost impossible to know whether to attribute a postmanipulation difference to preexisting conditions or the manipulation.

Factorial Manipulations. The analysis of data from nonequivalent independent group designs with factorial manipulations proceeds very similarly to analysis of data from randomized independent group designs. The analysis of variance model appropriate for the exact design is applied and the group by manipulation interaction is tested. One cannot, however, state a strong causal conclusion even if a significant group by manipulation interaction is found. For example, a difference in the effect of preparatory intervals on the reaction times of schizophrenics and normals might be "caused" by the pathology, but it might also be "caused" by some other factors correlated with the pathology in the same studied. One technique that may aid in clarifying such relations is *analysis of covariance*. With it the effect of any variable that correlates similarly with the dependent variable within all groups can be removed as a potential explanation for the group by manipulation interaction. For example, if the ages of schizophrenics and normal persons were significantly different, covarying out age could eliminate it as a potential confounding factor. Analysis of covariance is discussed in more detail in the next section.

Confounded Manipulations. When the treatment condition is confounded with preexisting differences in the subjects, analysis of the data is more difficult. Yet often researchers are forced to use designs in which pathological subjects are assigned to experimental and control groups by other than random means, for example, on the basis of volunteering, hours of availability, severity of psychopathology, or ethical considerations (who needs a potentially valuable treatment the most?). Abnormal subjects almost

inevitably differ on other factors than diagnosis, for example, drug maintenance. In some cases the manipulation itself may be a naturally occurring event out of the experimenter's control, for example, changes in hospitalized patients' diets, changes in psychotherapy programs, changes in drug administration. Pretests are almost always given in such cases, but a simple analysis of variance of the pre–post change scores to test for a difference between groups (or equivalently a test for a group by time interaction in a repeated measures analysis) would not be adequate because of the confounding. One common alternative is to use the pretest as a covariate along with any other measured variables that discriminate between the groups. In the resulting analysis of covariance the main effect for groups (instead of a group by time interaction) would be of greatest interest. Of course, if a number of dependent variables need to be examined, a multivariate analysis of covariance would be used.

The appropriateness of any analysis of covariance depends upon a number of assumptions. First, the covariates must correlate about the same with the dependent variable within each treatment group (homogeneity of regression). For example, if age related to performance quite differently within a sample of schizophrenics and a sample of normal persons, then analysis of covariance with age as a covariate should not be used. Second, an analysis of covariance assumes that everyone is regressing toward the same mean. If subjects were selected by pretest for the manipulation they would receive (e.g., as they might be for a remedial or therapy treatment), then the samples might represent different ends of the distribution. If the dependent variable were only moderately reliable, there would be substantial regression effects on the posttest. An analysis of covariance would correct for the assumed pre-post regression appropriately. But if the groups represent two qualitatively different populations with different mean true scores, this assumption would be incorrect. The analysis would be based on an overestimate of regression toward the grand mean, and would have an increased chance of Type I or Type II errors. For groups that appear to be from distinctly different populations with substantially different mean true scores on the dependent variable, Kenny (1975) advocates deleting the pretest as a covariate because of this difficulty. Instead one can use the change in standardized score for each subject as the dependent variable in the analysis of covariance. The scores are standardized at each time using the entire sample of subjects; so the change in standardized score for a subject represents his or her change corrected for changes in the overall mean or standard deviation. On the other hand, if the groups come from one population and have different pretest means because the subjects *volunteered* for the treatments they received, the pretest can be used as a covariate as long as one corrects for its unreliability. The reasoning is that volunteering is based on a true score; but the covariance analysis is based on the less reliable observed pretest score. Therefore, the covariance analysis predicts too much regression, and is inaccurate. The solution is to divide the regression coefficient for the pretest by its reliability. In other cases in which the subjects can be considered to come from one population but differ in pretest scores, an uncorrected analysis of covariance can be used.

Quasi-Experimental Matched Groups Design

Sometimes researchers attempt to remedy the problem of nonequivalent groups by discarding the subjects in each group who cannot be matched with any subject in the other groups (Miller, Saccuzzo, & Braf, 1979; Straube & Germer, 1979). Such a technique is quite different from the randomized blocks technique because the matched subjects come from distinctly different populations and are not randomly assigned to a treatment con-

dition after being matched. The population determines the treatment condition. On the whole, this design probably should be avoided. The problem is that one can only match subjects on observed scores not on true scores. If the populations from which the samples are drawn have different mean true scores, matching biases both samples away from their own means and exacerbates the likelihood of regression effects masking or being mistaken for real effects. For example, if a schizophrenic and normal sample were matched for pretest performance on a cognitive task, but the population mean were really lower for schizophrenics, one would have selected above-average schizophrenics and below-average normal persons. The best prediction for posttest performance would be an increase for normals (toward their mean) and a decrease for schizophrenics (toward their mean). Of course, if the matching variable were uncorrelated with the dependent variable, matching could not hurt; but then why match? The only situation in which a case might be made for matching subjects from two different populations would be when the *population* means on the matching variable were equal and the matching variable was correlated with the dependent variable. If the available subjects were unrepresentative of their populations, then simple random sampling would yield biased samples. In this case matching might produce a more representative sample and would allow a paired observation statistical test with reduced error variance and greater power. Unfortunately, this is not the situation in which matching is most often used.

Not everyone agrees with this viewpoint on matching. For example, Chapman and Chapman (1977) have recently argued in favor of matching in studies of schizophrenic cognition. They argue that there are many differences between the normal and schizophrenic populations that are correlated with cognitive processes, and these differences somehow must be controlled. There is merit to this argument. The problem is that when control is accomplished by matching, one ends up with samples unrepresentative of the populations and perhaps unmatched on true scores. Still this can be an acceptable situation if one is more interested in differences between certain subgroups of the populations than in drawing inferences about the populations, for example, one may be interested in comparing young schizophrenics with young normals. The danger is that the generalizations made from such research may go beyond the subgroups studied. For example, if one found a difference in the cognitive processes of schizophrenics and normal persons matched for IQ, one would certainly be tempted to conclude it was a difference between schizophrenics and normal persons in general. Yet it may only be a difference in high IQ schizophrenics. The generality of the conclusion must depend upon the matching variable, its distribution in the two populations, and its theoretical relevance to the pathology. Certainly, in some cases, if used with care, matching can be beneficial as long as judicious conclusions are drawn. However, more often an appropriate use of covariates with an independent groups design can accomplish the desired result with greater safety.

Quasi-Experimental Time Series Design

In time series designs control is achieved by repeating the measurement of the dependent variable a large number of times before and after each of a series of interventions. Then one tests whether the intervention produced a change in the relation between time and the dependent variable. A time series design is a special kind of repeated measures experiment in which a trend over time (perhaps due to maturation, testing, or spontaneous remission) is expected and corrected for. In a typical repeated measures experiment a subject is observed only once in each treatment condition, and the order of treatments is counterbalanced across subjects. In a time series quasi-experiment a subject may be

observed hundreds of times in each condition, and no attempt is made to counterbalance order of treatments. The internal validity of the quasi-experiment does not depend so much on the number of subjects studied as on the number of observations made. The sample size for the statistical tests is the number of observations not the number of subjects. Even a single-subject time series design can be internally valid, though the external validity (generality) of such single-subject experiments may be very limited. While time series designs have been widely used in therapy outcome research (Kazdin & Wilson, 1978) and in behavior modification experiments with abnormal subjects (Ayllon & Azrin, 1965), most researchers have not employed an appropriate statistical analysis.

For a complete description of the methods for analyzing time series data, one must consult other texts (Cook & Campbell, 1979; Box & Jenkins, 1976). The simplistic approach is to derive three linear equations for predicting the dependent variable as a function of time. One equation is computed for all observed data; one is computed for the data collected before each intervention; and one is computed for the data collected after each intervention. The researcher then tests the null hypothesis that the slopes and intercepts of the three lines are identical.

There are several problems with this approach. First, the particular theoretical model being tested may predict a different pattern of results than a simple change in slope or intercept after the intervention. For example some intervention procedures might have delayed, temporary, nonlinear, or cyclical effects. As a result, the exact prediction equation to be used and the range of times from which data are used must depend upon the theoretical model being tested. Second, the use of an ordinary least-squares regression analysis to estimate and compare the prediction equations is of questionable validity. The slopes and intercepts of the three lines would be valid unbiased estimates, but the statistical tests for comparing such regression lines would be faulty. These tests assume independence of observations, while in fact each observation in a time series is almost inevitably correlated with the previous observations (autocorrelated). In other words, a score observed at time t is correlated not only with time itself but with the score observed at time $t - 1$, time $t - 2$, and so on. The procedures developed to cure this difficulty allow the specification of a variety of models and remove the autoregressive effect before testing the intervention effect (Box & Jenkins, 1976). Yet like many multivariate techniques, the analysis is only feasible with specialized computer programs and requires statistical sophistication. As a result, many researchers use ordinary regression analysis as a first approximation (see Johnston, 1972, for details).

THREE EXPERIMENTAL STUDIES OF PSYCHOPATHOLOGY

Having examined systematically the gamut of designs and analyses frequently used in experimental studies of psychopathology, let us look at three specific experiments in greater detail in order to see how well theory and practice coincide. These three are specimens of some of the better published experiments, and if flawed in places, they are near the top overall. Each also represents a popular area of experimental research on psychopathology.

Cognitive Processes in Schizophrenics

Bellissimo and Steffy's (1972) experiment on attention in schizophrenics is representative of the large body of experiments performed to understand cognitive processes in schizophrenics. The authors selected their schizophrenic and nonschizophrenic patients in a manner typical of many similar experiments. Essentially all the female patients in two wards who met the diagnostic criteria and agreed to participate were used. However, as usual a few patients refused to participate, and a few had some characteristics

the experimenters considered inappropriate (e.g., excessive hospitalization) and were excluded. Another normal control group consisted of volunteers from the hospital staff. Obviously this study must be considered a *quasi-experiment with nonequivalent groups*.

The dependent variable was reaction time to a 300 Hz, 70db tone, and the manipulated independent variables were the duration and regularity of the time between when the experimenter said "ready" and when the tone was sounded (preparatory interval). Within each group a Latin square repeated measures design was used for manipulating the preparatory interval along two dimensions: duration, and regular versus irregular. A particular preparatory interval would be called regular if the preceding three intervals had been the same length. Since the manipulations were combined factorially with groups, the major statistics of interest would be the groups by manipulation (duration and regularity) interactions.

After some preliminary tests, Bellissimo and Steffy appropriately examined the group by duration by regularity interaction, which was found to be highly significant. They then proceeded with tests of the two-way duration by regularity interations within each group to determine the cause of the three-way interaction. This is equivalent to testing for simple effects after a significant two-way interaction is found. The only group for which the two-way interaction was significant was the process schizophrenics, indicating the reason for the three-way interaction. The process schizophrenics respond more rapidly to regularly spaced warning signals when the interval is short and to irregularly spaced warning signals when the interval is long (reaction time crossover). All other subjects always respond more rapidly to regularly spaced warnings. Other analyses were performed, but these are the critical ones.

Overall, the Bellissimo and Steffy experiment provides a reasonably good model for experiments comparing cognitive processes in schizophrenics and normals. It has flaws, for example, pathology is confounded with other factors, but not many flaws compared to most experiments of this type.

Learned Helplessness

Cole and Coyne's (1977) experiment on learned helplessness is a good example of the use of *yoked* controls in experiments on psychopathology. It was an analog study in that the manipulation was intended to *induce* a state analogous to depression in human subjects. The dependent variables were the subject's performance (latency, failures, and trials to criterion) on an anagram task and the subject's change in affective mood. Each subject was assigned *randomly* to either "escapable noise" pretraining or "inescapable noise" pretraining and to posttesting in either the "same" room or a "different" room. These two independent variables were combined factorially in a posttest only design for the anagram task and a pre–post design for the mood measure. Although the groups were independently selected and randomly assigned, each subject in the "inescapable" group was yoked to a subject in the "escapable" group for the pretraining. During pretraining, subjects in the escapable noise group would learn to turn off an aversive tone by pressing the appropriate pattern of buttons. Since the tone might terminate anyway, a green light would appear when the subject caused it to turn off. The yoked control subject's tone would be terminated at the same instant, but a red light would always be shown to him or her denoting failure. Thus the escapable subject and yoked control inescapable subject received the same amount and pattern of aversive stimulation, but the escapable subject could learn to control the aversive stimulation, and the inescapable subject could not. This procedure was intended to induce learned helplessness in the inescapable noise subjects, which would be revealed by lowered performance on the anagram task and depressed mood.

Cole and Coyne believed this effect might not generalize to a different setting (room); so their major statistic of interest should have

been the interaction of escapability and room on the dependent variables. Since there were three performance measures and one mood measure, a multivariate analysis of variance would seem to be appropriate. Cole and Coyne recognized this fact, unlike most learned helplessness researchers, and executed a MANOVA. Unfortunately, the application of the MANOVA and ordering of tests were not the most appropriate. First, Cole and Coyne treated the escapable variable as an independent groups variable when it could have been analyzed with a randomized blocks model (because yoking was used). Second, instead of including all the dependent variables in the MANOVA, they only included the three performance measures. Third, univariate tests were reported before the MANOVA was discussed for hypotheses made irrelevant by the MANOVA's results. The most appropriate procedure would have been to examine the escapability by room interaction first with the MANOVA. Since it was insignificant, the authors could then look at the two main effects, of which only the room effect was significant. An inspection of which dependent variables contributed to the effect was then proper and revealed high canonical loadings for anagram failures and latency but a low loading for trials to criterion. However, since the three dependent variables are very highly intercorrelated, not much should be made of this difference. When multicolinearity is high among dependent variables, regression and canonical coefficients will fluctuate wildly over replications.

While Cole and Coyne's analysis of their data did not proceed in the most efficient manner, they reached the correct conclusions, and the experiment is a well-executed example of the type of study widely seen in learned helplessness research.

The Effect of Alcohol on Normal Persons

The study of alcohol's effects on normal persons is one of the topics in clinical psychology that is most amenable to investigation with true experiments. Abrams and Wilson's (1979) experiment on alcohol and social anxiety in women serves as a good specimen of this genre. The subjects were paid female undergraduates selected from volunteers who reported only moderate drinking. Two independent variables, alcohol dose and expectancy, were manipulated by randomly assigning subjects to one of four groups—expect alcohol and receive alcohol, expect alcohol and receive placebo, expect placebo and receive alcohol, or expect placebo and receive placebo. This design is typical of a large number of studies, and procedures have been devised that seem to assure its integrity. The manipulation was double blind—neither the subject nor the person running the subject knew the subject's actual treatment condition, that is, what the subject was drinking. Each subject was told that both alcohol and nonalcohol groups must be tested in the study and that he or she had been randomly assigned to the alcohol (or tonic only) group. The subjects gargled with a mouthwash that reduced their taste sensitivity, under the guise of preparing for a breathalyzer test. The drinks were mixed from labeled bottles in full view of the subject, and the glasses in the "receive placebo" group were surreptitiously smeared with vodka. Finally, the breathalyzer was altered to give false feedback. Manipulation checks based on self-reports of alcohol consumed clearly indicated the manipulations were highly successful. In experiments involving alcohol consumption expectancy checks are just as necessary as breathalyzer tests and should always be included.

After the drink manipulation and breathalyzer test, the subjects were placed in a controlled social interaction with a male confederate. The dependent variables included physiological measures of anxiety (heart rate and skin conductance) taken before, during, and after the interaction, self-report measures of anxiety taken before and after the interaction, and observer ratings of the subject's

anxiety during the interaction. Thus the design was an independent randomized groups design with pre–post data on the physiological and self-report measures.

Abrams and Wilson's major analysis was a neatly done multivariate analysis of variance of change scores including all the dependent variables. After finding no significant interaction, the authors reported a significant main effect for "expectancy". Those subjects who expected alcohol were significantly more anxious than those who did not expect alcohol regardless of what they actually received. Examination of the canonical coefficients and appropriate post-hoc univariate analyses of variance of each dependent variable showed that expectancy had its greatest effect on the physiological measures and observer ratings. The effect on self-reported anxiety appears minimal, but when change in self-reported anxiety was used as a dependent variable, a significant effect was found. This is a good illustration of a situation in which a pre–post design produced greater statistical power. All in all, Abrams and Wilson's analyses serve as one of the best published examples of the proper way to deal with data from experiments of this type.

SOME GENERAL THREATS TO THE VALIDITY OF CAUSAL CONCLUSIONS

Experimental studies of psychopathology are no different from any other experiments in their susceptibility to sampling errors, manipulation mistakes, and analysis failures. Each of the specific designs discussed above has its own pitfalls for the unwary researcher. Some of the common errors discussed earlier have been *biased sampling, matching across populations, testing main effects before interactions,* and *using a large number of univariate tests* when one multivariate test is appropriate. Nevertheless, a few threats to the validity of causal conclusions have not been discussed sufficiently because they are not specific to any particular design.

Differential Mortality Among Subjects

When one selects pathological subjects for an experiment, it is almost inevitable that certain subjects will be dropped from the sample before the experiment is completed or even before it is begun. The seriously pathological subjects may be too incapacitated to perform the experimental task, biasing the sample in the direction of the less pathological. Worse, subject mortality may interact with the treatment conditions, so that the subjects drop out at different rates from the experimental and control groups. Klisz and Parsons (1979) have demonstrated the importance of this factor for investigations of cognition in alcoholics, but it is no less serious a problem for other pathologies. For example, when one discards subjects from a hypnotic treatment because they could not be hypnotized, one has committed the same error. One cannot know whether differences between the hypnotic group and other subjects are due to the treatment or subject selection. The most satisfactory solution for the mortality problem is to impose strong enough motivational conditions that mortality is minimized. Unfortunately, this is sometimes not possible, for example, in hypnosis experiments. In such cases the experimenter can resort to using additional control groups to measure the effect of mortality, that is, compare the dropouts with an unhypnotized group of hypnotizable subjects.

Attention and Demand Effects

In many experiments the procedures are designed so that the researcher must spend a great deal of time with the experimental subjects and only a little with the control subjects. Among hospitalized subjects such attention may be highly rewarding, and the subjects may be particularly complaint in order to maintain the attention. Thus differ-

ential attention is always an alternative explanation for a difference between an experimental and control group of pathological subjects. The solution is to be sure that the control group receives equal attention through a placebo treatment with similar demand characteristics.

Stimulus Specificity

Cognitive and social processes in abnormal subjects may be particularly sensitive to stimulus characteristics to which the experimenter is unattentive, for example, affective tone, remote associations, experimenter attractiveness, sex of experimenter. Too often there has been a tendency to conclude that one particular characteristic of a stimulus caused an effect when other characteristics could just as easily have produced it (Maher, 1978). If a sex of therapist by treatment interaction is found, was it really due to sex of therapist or to another characteristic correlated with sex in the particular therapists used? If a noncontingent aversive noise produces a helplessness effect, does one conclude that "noncontingency" is necessary? In order to generalize one's conclusions beyond the narrow context of the stimuli used, one should employ a sample of stimuli in each experimental treatment. If sex of experimenter is an independent variable, then use several different male and female experimenters. If the critical stimulus has several attributes besides the ones manipulated, then use a random sample of stimuli of each type (for example, see Elman, Schroeder, & Schwartz, 1977, and Hammen & Peters, 1978).

Fallacious Causal Reasoning in Normal–Abnormal Comparisons

Often in experimental studies comparing normal and abnormal populations, one discovers that a particular cognitive or social process operates differently in pathological than in normal populations. This difference may coincide perfectly with the researcher's model for the etiology of the pathology. Nevertheless, a conclusion that the difference causes the pathology is unjustifiable. These studies are only quasi-experiments. Any difference between the normal and abnormal subjects may be a cause of the pathology, may be caused by the pathology, may be caused by another characteristic of the pathological sample, or may be caused by some variable correlated with the pathology. Furthermore, the lack of a difference between the groups on a cognitive or social process does not imply that the process is irrelevant to the etiology of the pathology. A causal process critical to the development of a disorder may be masked by the emergence of the full symtomatology of the disorder. In general the results of any quasi-experiment comparing normal and pathological subjects cannot provide definitive evidence about the etiology of the pathology.

SUMMARY

In reading over this chapter one might be tempted to become very pessimistic about the utility of experimental research for answering important questions about the development of psychopathologies. Clearly researchers are handicapped by the fact that psychopathological states cannot be manipulated very well, and analog states differ from true pathological states in important respects. Nevertheless, if one evaluates the contributions of experimental research relative to other types of empirical research, one can be optimistic. Recent experimental and quasi-experimental studies have yielded major contributions to our understanding of schizophrenic cognition (Chapman, 1956; Bellissimo & Steffy, 1972; Knight et al., 1977; Koh et al., 1977; Payne et al. 1959), depression (Cole & Coyne, 1977; Hammen & Peters, 1978; Hiroto & Seligman, 1975; Price et al. 1978; Rizley, 1978; Seligman & Maier, 1967; Smolen, 1978), appetitive disorders (Brown & Williams, 1975; Glad & Adesso, 1976; Marlatt

et al., 1973; Polivy et al., 1976; Ruderman & Wilson, 1979), state-dependent learning (Eich et al., 1975; Weingartner et al., 1977), and hypnotism (Banyai & Hilgard, 1976; Coe et al., 1976; Kihlstrom, 1980; Spanos & Brodorik, 1977), as well as other disorders. The goal for future researchers should be to take advantage of recent advances in design and analysis techniques to expand the experimental study of these and other psychopathologies.

REFERENCES

Abrams, D. B., & Wilson, G. T. (1979). Effects of alcohol on social anxiety in women: cognitive versus physiological processes. *Journal of Abnormal Psychology,* **88,** 161–173.

Abramson, L. Y., Seligam, M. E. P., & Teasdale, J. (1978). Learned helplessness in humans: critique and reformulation. *Journal of Abnormal Psychology,* **87,** 49–74.

Aylon, T., & Azrin, N. H. (1967). The measurement and reinforcement of behavior of psychotics. *Journal of Experimental Analysis of Behavior,* **8,** 357–383.

Bachrach, R. S., Huesmann, L. R., & Peterson, R. (1977). The relation between locus of control and the development of moral judgment. *Child Development,* **48,** December, 1340–1352.

Banyai, E., & Hilgard, E. R. (1976). A comparison of active-alert hypnotic induction with traditional relaxation induction. *Journal of Abnormal Psychology,* **85,** 218–224.

Bauman, E., & Kolisnyk, E. (1976). Interference effects in schizophrenic short-term memory. *Journal of Abnormal Psychology,* **85,** 303–308.

Belissimo, A., & Steffy, R. A. (1972). Redundancy associated deficit in schizophrenics reaction time performance. *Journal of Abnormal Psychology,* **80,** 229–307.

Blaney, P. H. (1977). Contemporary theories of depression: critique and comparison. *Journal of Abnormal Psychology,* **86,** 203–223.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forcasting and control.* San Francisco: Holden-Day.

Brown, R. A., & Williams, R. J. (1975). Internal and external cues relating food intake in obese and alcoholic persons. *Journal of Abnormal Psychology,* **84,** 660–665.

Brown, R. A., & Cutler, H. S. (1977). Alcohol, customary drinking behavior and pain. *Journal of Abnormal Psychology,* **86,** 179–188.

Buss, A. H. (1966). Instrumentality of aggression, feedback, and frustration as determinants of physical aggression. *Journal of Personality and Social Psychology,* **3,** 153–162.

Chapman, L. J. (1956). Distractibility in the conceptual performance of schizophrenics. *Journal of Abnormal and Social Psychology,* **53,** 286–291.

Chapman, L. J., & Knowles, R. R. (1964). The effects of phenothiazine on disordered thought in schizophrenia. *Journal of Consulting Psychology,* **28,** 165–169.

Chapman, L. J., & Chapman, J. P. (1977). Selection of subjects in studies of schizophrenic cognition. *Journal of Abnormal Psychology,* **86,** 10–15.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior,* **12,** 335–359.

Coe, W. C., Basden, G., Basden, D., & Graham, C. (1976). Posthypnotic amnesia: Suggestions of an active process in dissociative phenomena. *Journal of Abnormal Psychology,* **85,** 455–458.

Cohen, D. B. (1979). Dysphoric affect and REM sleep. *Journal of Abnormal Psychology,* **88,** 73–77.

Cole, C. S., & Coyne, J. C. (1977). Situational specificity of laboratory induced learned helplessness. *Journal of Abnormal Psychology,* **86,** 615–623.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Costello, C. G. (1978). Learned helplessness and depression in humans. *Journal of Abnormal Psychology,* **87,** 21–31.

Dement, W. (1960). The effect of dream deprivation. *Science,* **131,** 1705–1707.

Dennis, W. (1960). Causes of retardation among institutional children. *Journal of Genetic Psychology,* **96,** 47–59.

Dweck, C. S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology,* **31,** 674–685.

Eich, J. E., Weingartner, H., Stillman, R., & Gillin, J. C. (1975). State-dependent accessibility of retrieval cues in the retention of a categorized list. *Journal of Verbal Learning and Verbal Behavior,* **14,** 408–417.

Elman, D., Schroeder, H. E., & Schwartz, M. J. (1977). Reciprocal social influence of obese and normal weight persons. *Journal of Abnormal Psychology,* **86,** 408–413.

Ericson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review,* **87,** 215–251.

Feldman-Summers, S., Gordon, P. E., & Meagher, J. R. (1979). The impact of rape on sexual satisfaction. *Journal of Abnormal Psychology,* **88,** 101–105.

Gald, W., & Adesso, V. J. (1976). The relative importance of socially induced tension and behavioral contagion for smoking behavior. *Journal of Abnormal Psychology,* **85,** 119–121.

Gomes-Schwartz, B. (1978). Effective ingredients in psychotherapy: Predictors of outcome from process variables. *Journal of Consulting and Clinical Psychology,* **46,** 1023–1035.

Hammen, C. L., & Peters, S. D. (1978). Interpersonal consequences of depression: Responses to men and women enacting a depressed role. *Journal of Abnormal Psychology,* **87,** 322–332.

Hare, R. D. (1979). Psychopathy and literality of cerebral function. *Journal of Abnormal Psychology,* **88,** 605–610.

Heaton, R. K., & Victor, R. G. (1976). Personality characteristics associated with psychedelic flashbacks in natural and experimental settings. *Journal of Abnormal Psychology,* **85,** 83–90.

Hiroto, D. S., & Seligman, M. E. P. (1975). Generality of learned helplessness in man. *Journal of Personality and Social Psychology,* **31,** 311–327.

Johnston, J. (1972). *Econometric methods.* New York: McGraw-Hill.

Jones, S. L., Nation, J. R., & Massad, P. (1977). Immunization against learned helplessness in man. *Journal of Abnormal Psychology,* **86,** 75–83.

Kazdin, A. L., & Wilson, G. T. (1978). *Evaluation of behavior therapy: Issues, evidence, and research strategies.* Cambridge, Mass.: Ballinger.

Kihlstrom, J. F. (1980). Posthypnotic amnesia for recently learned material: Interactions with episodic and semantic memory. *Cognitive Psychology.* **12,** 227–251.

Klee, G. D., Bertino, J., Weintraub, W., & Calaway, E. (1961). The influence of varying dosage on the effects of lysergic acid diethylamide. *Journal of Nervous and Mental Disease,* **132,** 404–409.

Kenny, D. A. (1975). A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. *Psychological Bulletin.* **82,** 345–362.

Klisz, D. K., & Parsons, O. A. (1979). Cognitive functioning in alcoholics: The role of subject attrition. *Journal of Abnormal Psychology,* **88,** 268–276.

Knight, R., Shirer, M., & Shapiro, J. (1977). Iconic imagery in overinclusive and nonoverinclusive schizophrenics. *Journal of Abnormal Psychology,* **86,** 242–255.

Knight, R., Shirer, M., Putchat, C., & Carter, G. (1978). A picture integration task for measuring iconic memory in schizophrenics. *Journal of Abnormal Psychology,* **87,** 314–321.

Knox, V. J., & Shum, K. (1977). Reductions of cold-pressor pain with acupuncture analgesia in high and low hypnotic subjects. *Journal of Abnormal Psychology,* **86,** 639–643.

Koh, S. D., Szoc, R., & Peterson, R. A. (1977). Short-term memory scanning in schizophrenic young adults. *Journal of Abnormal Psychology,* **86,** 451–460.

Koh, S. D., & Peterson, R. A. (1978). Encoding orientation and the remembering of schizophrenic young adults. *Journal of Abnormal Psychology,* **87** 303–313.

Krantz, S., & Hammen, C. (1979). Assessment of cognitive bias in depression. *Journal of Abnormal Psychology,* **88,** 611–619.

Lang, A. R., Goecker, D. J., Adesso, V. J., & Marlatt, G. A. (1975). Effects of alcohol on aggression in male social drinkers. *Journal of Abnormal Psychology,* **84,** 508–518.

Maher, B. A. Stimulus sampling in clinical research: Representative design reviewed. *Journal of Consulting and Clinical Psychology,* 1978, **46,** 643–647.

Marlatt, G. A., Demming, B., & Reid, J. B. (1973). Loss of control drinking in alcoholics: An experimental analogue. *Journal of Abnormal Psychology,* **81,** 233–241.

Martindale, C. (1978). The therapist-as-fixed effect fallacy in psychotherapy research. *Journal of Consulting and Clinical Psychology,* **46,** 1526–1530.

Meiselman, K. C. (1978). Inducing "Schizophrenic" associations in normal subjects. *Journal of Abnormal Psychology,* **87,** 291–293.

Metzner, R., Litwin, G., & Weil, G. W. (1965). The relation of expectations and mood to psilocybin reactions. *Psychedelic Review,* **5,** 3–39.

Miller, E. & Lewis, P. (1977). Recognition memory in elderly patients with depression and dementia: A signal detection analysis. *Journal of Abnormal Psychology,* **86,** 84–86.

Miller, R. G. (1966). *Simultaneous statistical inference.* New York: McGraw-Hill.

Miller, R. J., & Leibowitz, H. W. (1976). A signal detection analysis of hypnotically induced narrowing of the peripheral visual field. *Journal of Abnormal Psychology,* **85,** 446–454.

Miller, S., Saccuzzo, D., & Braf, D. (1979). Information processing deficits in remitted schizophrenics. *Journal of Abnormal Psychology,* **88,** 446–449.

Miller, W. R., and Seligman, M. E. P. (1975). Depression and learned helplessness in man. *Journal of Abnormal Psychology,* **84,** 228–238.

Maher, B. A. (1978). Stimulus sampling in clinical research: Representative design reviewed. *Journal of Consulting and Clinical Psychology,* **46,** 643–647.

Morrison, D. F. (1976). *Multivariate statistical methods.* New York: McGraw-Hill.

Neufeld, R. W. J. (1977). Components of processing deficit among paranoid and nonparanoid schizophrenics. *Journal of Abnormal Psychology,* **86,** 60–64.

Nisbett, R. F., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review,* **84,** 231–259.

Oltmans, T. (1978). Selective attention in schizophrenia and manic psychoses: The effect of distraction on information processing. *Journal of Abnormal Psychology,* **87,** 212–225.

Payne, R. W., Mattusek, P., & George, E. I. (1959). An experimental study of schizophrenic thought disorder. *Journal of Medical Sciences,* **105,** 627–652.

Pihl, R. O., & Sigal, H. (1978). Motivation levels and the marihuana high. *Journal of Abnormal Psychology,* **87,** 280–295.

Polivy, J., Schueneman, A. L., & Carson, K. (1976). Alcohol and tension reduction: Cognitive and Physiological effects. *Journal of Abnormal Psychology,* **85,** 595–600.

Polivy, J., & Doyle, C. (1980). Laboratory induction of mood states through the reading of self-referent mood statements: Affective change or demand characteristics? *Journal of Abnormal Psychology,* **89,** 286–290.

Price, K. P., Tryon, W. W., & Raps, C. S. (1978). Learned helplessness and depression in a clinical population: A test of two behavioral hypotheses. *Journal of Abnormal Psychology,* **87,** 113–121.

Rizley, R. (1978). Depression and distortion in the attribution of causality. *Journal of Abnormal Psychology,* **87,** 32–48.

Rosen, L. J., & Lee, C. L. (1976). Acute and chronic effects of alcohol on organizational processes in memory. *Journal of Abnormal Psychology,* **85,** 309–317.

Ruderman, A. J., & Wilson, G. T. (1979). Weight, restraint, cognitions, and counterregulation. *Behavior Research and Therapy,* **17,** 581–590.

Russel, P. N., & Beekhuis, M. E. (1976). Organization in memory: A comparison of psychotics and normals. *Journal of Abnormal Psychology,* **85,** 527–534.

Russel, P. N., & Knight, R. G. (1977). Performance of process schizophrenics on tasks involving visual search. *Journal of Abnormal Psychology,* **86,** 16–26.

Raulin, M. L., & Chapman, L. J. (1976). Schizophrenic recall and contextual constraint. *Journal of Abnormal Psychology,* **85,** 151–154.

Sacco, W. P., & Hokanson, J. E. (1978). Expectations of success and anagram performance of depressives in a public and private setting. *Journal of Abnormal Psychology,* **87,** 122–130.

Seligman, M. E. P., & Maier, S. F. (1967). Failure to escape traumatic shock. *Journal of Experimental Psychology,* **74,** 1–9.

Seligman, M. E. P., Abramson, L., Semmel, A., & von Baeyer, C. (1979). Depressive attributional style. *Journal of Abnormal Psychology,* **88,** 242–247.

Shakow, D. (1963). Psychological deficit in schizophrenia. *Behavioral Science,* **8,** 275–305.

Sheehan, P. W., Obstoj, I., & McConkey, K. (1976). Trans logic and cue structure as supplied by the hypnotist. *Journal of Abnormal Psychology,* **85,** 459–472.

Smolin, R. C. (1978). Expectancies, mood, and performance of depressed and non-depressed psychiatric inpatients on chance and skill tasks. *Journal of Abnormal Psychology,* **87,** 91–101.

Spanos, N. P., & Bodorik, H. L. (1977). Suggested amnesia and disorganized recall in hypnotic and task-motivated subjects. *Journal of Abnormal Psychology,* **86,** 295–305.

Spanos, N. P., Ansari, F., & Stain, H. (1979). Hypnotic age regression: A failure to replicate. *Journal of Abnormal Psychology,* **88,** 88–91.

Spanos, N. P., Radtke-Bodorik, H. L., Ferguson, J. D., & Jones, B. (1979). The effects of hypnotic susceptibility, suggestions for analysis, and the utilization of cognitive strategies on the reduction of pain. *Journal of Abnormal Psychology,* **88,** 282–292.

Spencer, J. A., & Fremouw, W. J. (1979). Binge eating as a function of restraint and weight classification. *Journal of Abnormal Psychology,* **88,** 262–267.

Steinberg, E. P., & Schwartz, G. E. (1976). Biofeedback and electrodermal self-regulation in psychopathy. *Journal of Abnormal Psychology,* **85,** 408–413.

Stevenson, J. H. (1976). Effect of post-hypnotic dissociation on the performance of interfering tasks. *Journal of Abnormal Psychology,* **85,** 398–407.

Straube, E. R., & Germer, C. K. (1979). Dichotic shadowing and selective attention to word meaning in schizophrenia. *Journal of Abnormal Psychology,* **88,** 346–353.

Teasdale, J. D. (1978). Effects of real and recalled success on learned helplessness and depression. *Journal of Abnormal Psychology,* **87,** 155–164.

Teasdale, J. D., & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology,* **88,** 248–257.

Tucker, J. A., Vuchinich, R. E., & Sobeil, M. B. (1979). Differential discriminative stimulus control of nonalcoholic beverage consumption in alcoholics and in normal drinkers. *Journal of Abnormal Psychology,* **88,** 145–160.

Velten, E. (1968). A laboratory task for induction of mood states. *Behavior Research and Therapy,* **6,** 473–482.

Weingartner, H., Miller, H., & Murphey, D. L. (1977). Mood-state-dependent retrieval of verbal associations. *Journal of Abnormal Psychology,* **86,** 276–284.

Wilson, G. T., & Lawson, D. M. (1978). Expectancies, alcohol, and sexual arousal in women. *Journal of Abnormal Psychology,* **87,** 358–367.

Zeichner, A., & Pihl, R. O. (1979). Effects of alcohol and behavior contingencies on human aggression. *Journal of Abnormal Psychology,* **88,** 153–160.