returning change, they randomly assigned customers to one of three experimental conditions: a no-touch control condition; a brief hand-touch condition; or a longer shoulder-touch condition. Because customers were randomly assigned to conditions at the *close* of their interactions with the waitress-confederates—blind to subjects' experimental condition until that point—only the touch manipulation differentiated customers from one another. Thus, the independent variable in this experiment was the nature of physical contact initiated by the waitress-confederates. Levels of the independent variable defined the conditions of the experiment. In this case, there were three: not touching the customer; touching only fleetingly on the hand; or touching on the shoulder.

We often begin by thinking of the independent variable at a more abstract conceptual level. There are several different types of relationships that can exist between the experiment's conceptual independent variable and the events that actually happen to a subject in the experiment. At the simplest level, as in the experiment on touching, the experimenter may be interested in a particular concrete variable for its own sake. The independent variable, the touch, was not a representation of some more abstract conceptual variable; instead, the concept and the event were so closely linked as to be almost identical. The study on touching may be viewed as the investigation of a simple stimulus-response relationship. The purpose of the experiment was to determine the effects of an external stimulus (touching) on subjects' responses (questionnaire service ratings and tipping). The studies of bystander intervention discussed earlier (Latané & Darley, 1968; Latané & Rodin, 1969) also exemplify this type of research. Here, the independent variable was the mere presence of other people in a situation in which the subject was faced with an emergency. There are relatively few difficulties involved in translating the idea of "presence of others" into an experiment treatment.

In the majority of social psychological experiments, the situation is more complicated. Consider the Asch (1951) experiment, in which the subject was faced with a unanimous majority who disagreed with him. Asch was not *specifically* interested in unanimous majorities of seven or eight people. Rather, he was interested in the broader, more abstract conceptual variable "group pressure," and the unanimous majority was the particular concrete event he chose as his empirical realization (Chapter 1) of the concept of group pressure. The particular event is taken as *representative* of the broad class of events that are included in the concept of group pressure. Other experimenters may use different, equally valid empirical realizations of the same conceptual variable. In Chapter 2 we discussed the relationship between conceptual variables and their empirical realizations in detail and noted that the experimenter's concerns are not solely with the specific realization of the independent variable being used in any particular experiment. Instead, the experimenter's interest may extend beyond the specific events contained in the particular empirical realization to include the whole class of events subsumed within the abstract category. Again, the experimenter is interested in a stimulus-response relationship, but the real "stimulus" of interest is much broader than the specific situation set up for the relationship, as shown in Figure 7-1.

The situation becomes still more complex when the investigator's hypothesis involves an internal stimulus variable that cannot be manipulated directly. For example, several researchers have been interested in studying the effects of guilt. Recall the field experiment of Regan, Williams, and Sparling (1972) described in Chapter 6, in which subjects were asked to take a confederate's picture at a shopping mall. Some subjects were induced to believe that they had broken his very expensive camera, and others were informed that the camera malfunctioned all the time and that they were not responsible. The subjects who believed they had broken the camera—the guilty subjects—subsequently helped a second confederate more than did the nonguilty control subjects.

The conceptual independent variable in this study was a feeling of guilt. That is to say, the experimenters were interested in comparing the behavior of subjects who felt guilty with the behavior of subjects who did not. Since it is impossible to vary the subject's internal state directly, the experimenter must use the indirect method of creating an external stimulus designed to bring about that state. Regan, Williams, and Sparling created a situation in which subjects were induced to take a photograph, and thus to break an expensive camera. They expected that subjects would react to this transgression with a feeling of guilt. The feeling of guilt was then expected to lead to differences in the subject's willingness to help another person.

As in the case of all studies involving a broad conceptual variable, the conceptual independent variable and its realization are not the same, and investigators are not particularly interested in the specific events involved in the stimulus situation, for example, as in the case above, destruction of a camera. In studies in which the conceptual variable is an abstract category of external stimulus events, such as "group pressure," the experimenter is interested in the particular events used only insofar as they are representative of the broad category. In studies in which the conceptual variable is an internal state

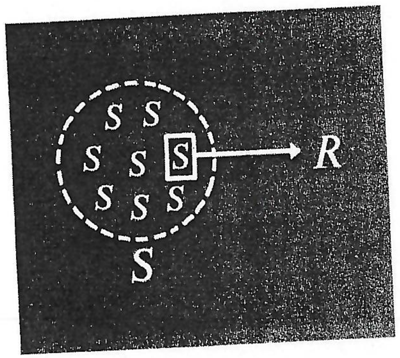

FIGURE 7-1
Independent variable as representative of a class of variables.

$$S \longrightarrow (R = S) \longrightarrow R$$

FIGURE 7-2
Conceptual independent variable as an internal state.

such as guilt, the experimenter is not interested in the particular events except in terms of their success in arousing the internal state whose effects are to be studied. Theoretically, the external event produces the independent variable (internal event), and the independent variable produces the behavior.[1] Thus, there are two stimulus-response relationships involved: (1) The internal state (conceptual variable) is a *response* to the particular experimental situation, and (2) at the same time it is a *stimulus* for the dependent variable. These relationships are shown in Figure 7-2.

When internal states are of concern, the investigator is primarily interested in the *second* of these stimulus-response relationships; the first is merely a means of producing the stimulus for the second. Thus, in other studies designed to test the effects of guilt, a wide variety of methods for producing guilt has been used. Subjects have been induced to break an expensive piece of laboratory equipment (Wallace & Sadalla, 1966); to upset a graduate student's carefully ordered pile of index cards (Freedman, Wallington, & Bless, 1967); to lie to give electric shocks to another subject (Carlsmith & Gross, 1969); to lie (Freedman, Wallington, & Bless, 1967); and to succeed on an experimental task because they had been tipped off to a trick for solving it (Carlsmith, Ellsworth, & Whiteside, 1968). None of these situations was particularly interesting to the experimenter for its own sake, but only as a means of inducing guilt. The conceptual variable guilt is the unifying factor tying these events to the experimental results.

In experiments involving the effects of an abstract conceptual variable, the investigator must face the difficult and complicated task of constructing an empirical realization that will represent the general concept or that will produce the desired state of mind in the subject. How can this be done? Unfortunately, there are few standard techniques. In social psychology, few experimental manipulations of the same conceptual variable are identical. The researcher must usually construct an experimental situation appropriate for the particular question or hypothesis, borrowing only bits and pieces from previous work. A situation that is appropriate for studying the effects of negative effort on liking for a group—as in the Aronson-Mills (1959) experiment—might not work if the experiment later wants to investigate the effects on liking for a color (Aronson, 1961) or liking for the whole experiment. As the experiment is modified to create a plausible situation in which to study these other questions, the experimental treatments will usually have to be modified to fit the new situations, as in the variety of guilt studies described above. Likewise, social psychological experiments depend so heavily on the special nature of the subculture under study that even those relatively standard procedures that *do* exist must often be altered drastically so that they make sense to the particular sample with which the experimenter is working.

[1] Technically, the term "independent variable" refers only to the concrete manipulations or empirical realization involved. We use the term loosely, however, to include the conceptual independent variable and hope that the meaning will be clear in context.

The fact that subject populations differ across places and that populations change over time has been a source of criticism of experimental work in social psychology. For example, Kenneth Gergen (1973) contends that cultural groups differ from one another in terms of values, norms, and patterns of social behavior, making it impossible to make valid generalizations based on one or a few studies. He adds that the capriciousness of social phenomena—that is, temporal changes in the values, norms, and ways of behaving which characterize any given culture—also make the kind of broad, sweeping regularities sought by psychologists impossible to discover. Naturally, social phenomena vary across populations. For example, subjects far removed from academia might not appreciate the importance of knocking over an ordered pile of index cards. Similarly, a low score on an IQ test might not be a blow to their self-esteem. Subject populations can change over time, as well. If Aronson and Mills tried to replicate their severity-of-initiation study today using a new population of sophisticated, sexually liberated students, they might find that reading a passage from *Lady Chatterley's Lover* would not be embarrassing for their subjects and thus would no longer constitute a "severe" initiation; what is deemed a severe initiation or unpleasant at one point in time—say, by college women in the 1950s—might hardly raise an eyebrow in the 1990s.

However, the abstract conceptual variables (e.g. "severe" initiation or "unpleasant effort") change only in the sense that experimenters might need to consider alternative *specific empirical realizations* of a construct as social phenomena differ across populations, or as they change with the passage of time. Although concrete *manifestations* of abstract constructs may vary as a function of time or locale, relationships between or among the constructs *themselves* may be far less variable. From a broad perspective, it is trivial that concrete manifestations of such concepts as "cohesiveness," "guilt," "aggression," "altruism," "self-esteem," "conformity pressure," or whatever might vary as a function of different subject populations or of temporal changes. But because these manifestations frequently *do* change, it is of *practical* importance to the experimenter and makes it of little use to outline *specific* techniques for experimentally varying such abstract concepts: There is no "Top 10" list for producing a good mood, for example. Of course, in the short run treatments that have been successful can often be reused, and when circumstances permit the use of a situation and subject population similar to those used in other experiments, it is a good idea to try using the same procedure, rather than thinking up a whole new idiosyncratic treatment. When an established technique can be used, it is much easier to integrate the experiıental findings with those of other experiments in the same area. When a new ıethod is introduced, it is sometimes difficult to assess the extent to which the ːsults are a function of procedural changes rather than of the real conceptual ifferences the experimenter has in mind. In designing an experiment, the in- estigator should read other people's research on the same topic and use the ame or similar techniques when appropriate. Often, however, the investigator vill not find a good method that fits the problem. What we hope to be able to lo in this chapter is to provide some general guidelines, some rules of thumb, ome intuitions, and some recommendations which may serve to direct an ex- erimenter toward a sensible and effective empirical realization of the inde- endent variable.

Our discussion of the ways in which independent variables may be realized ests heavily on the use of deception. In Chapters 3 and 6 we discussed many f the ethical and methodological issues involved in the use of deception. The ocus on such techniques in this chapter should not be construed as an indica- ion of a preference for deceptive techniques. Rather, it will become apparent hat procedural difficulties are often compounded when deception is used, and hus it becomes necessary to devote considerable discussion to those difficul- ies and some ways around them. Conversely, a lengthy discussion is hardly in order to describe how to experimentally vary the size of a group between two and eight, for example.

THE BASIC REQUIREMENT: RANDOMIZED ASSIGNMENT

The first and most important guideline has already been suggested in our dis- cussion of the nature of an experiment in Chapter 1. If we are to carry out an experiment that allows for causal inferences, we must determine which treat- ment each subject gets by the principle of randomization. If chance determines who gets which treatment, we know that any differences we observe are due to differences in the treatments, not to preexisting differences between or among subjects who were assigned to the various conditions.

In most laboratory experiments, random assignment of subjects to condi- tions is relatively easy to achieve by flipping a coin, rolling dice, or more com- monly, consulting a table of random numbers. And in some field experiments, it is likewise relatively easy to assign subjects to conditions at random, to de- termine, for example, whether to touch a specific customer in a restaurant or whether to stare at a specific driver.

In some rare situations field researchers may safely forgo random assign- ment and rely instead on the haphazard distribution of naturally occurring events to serve as real-life analogs of controlled experimental designs. For ex- ample, Stanley Parker, Marilynn Brewer, and Janie Spencer (1980) conducted a study on the effects of a natural disaster, a devastating brush fire that swept through a Southern California community, on the premise that the hit-or-miss pattern of destruction caused by the fire was essentially the product of a

"natural randomization" process. Among the homes in close proximity at the height of the fire, only chance factors—changes in wind direction and velocity, location of the fire-fighting equipment, traffic congestion, and so on—deter- mined which homes were burned to the ground and which were intact when the fire was finally brought under control. Thus, homeowners who were vic- tims of the fire were essentially equivalent to those who were more fortunate, and any group differences in attitudes and perceptions following the fire could safely be attributed to the differential experience of the "natural" administra- tion of the independent variable. Of course, in such "experiments," the bur- den of proof is on the investigator to make a convincing case that subjects were unlikely to differ systematically on relevant variables *prior* to the event. For example, if stone houses survived the fire and wooden houses burned down, we could not be so sure that differences between the victims and the nonvictims were due to the fire; as with the Three Little Pigs, people who choose to live in stone houses may be more practical, safety-conscious, or per- haps neurotic than their happy-go-lucky neighbors in wooden houses.

Whatever the setting of an experiment, the randomization requirement may seem obvious, but in practice any one of a number of considerations can lure the experimenter into forgetting about it, thereby defeating the purpose of do- ing an *experimental* study in the first place. It is to these considerations to which we now turn.

Individual Differences as "Independent Variables"

The most common pitfall occurs when the independent variable is something that we feel is already a characteristic of the person (a "subject variable"), and so we do not try to vary it experimentally. Instead, we measure it and then look at differences between people with widely separated scores on our mea- sure. For example, if we are concerned with differences in performance be- tween subjects with high self-confidence and those with low self-confidence, we might measure self-confidence in a sample of people and then compare the performance of the high-scoring subjects with that of the low-scoring subjects. This of course is a correlational study, even though it may be conducted in the laboratory with a high degree of control over the subsequent measures. As a correlational study, it suffers from the indigenous weakness already men- tioned: One cannot know whether differences in the subjects' behavior are *caused* by differences in self-confidence. Self-confident subjects differ from self-doubting subjects in many ways, any one of which might cause the ob- served differences in their behavior (see Chapter 1).

We are not suggesting that individual differences do not exist; anyone who has ever tried to persuade an "introverted" friend to attend a large, raucous party is confronted with the difference between someone who is introverted and someone who is extraverted. Neither are we suggesting that such differ- ences are unimportant in a scientific sense. Many investigators are interested in exactly these kinds of differences.[2] Our purpose here is just to make clear once again the distinction between experiments in which the table of random numbers decides which subject gets which treatment and studies in which something or somebody else makes that decision: studies in which individual differences are manipulated versus studies in which they are taken "as is."

Often, variables such as test scores, sex, age, and class are included in a study in which the main focus is an experimentally controlled independent variable. When used in conjunction with random assignment, evidence about individual differences can add to the precision of the experiment. Here, the purpose is not to look at the *effects* of these subject variables but simply to measure them to determine the *generality* of the effects of the true independent variable *or* to decrease the unexplained (error) variability in the experiment. For example, recall from Chapter 4 that Helmreich, Aronson, and LeFan (1970) conducted an experiment designed to explore the generality of the finding that very competent people are liked better when they commit some kind of "humanizing" blunder or pratfall (Aronson, Willerman, & Floyd, 1966). As in the original experiment, the investigators *manipulated* the competence of the stimulus person by showing a videotape recording of an interview in which the stimulus person's answers to questions showed him to be either a top student with many honors and a wide variety of successes, or a dull student with no honors and a wide variety of failures. They also manipulated the stimulus person's experience of a pratfall, by showing half the subjects in each condition a sequence in which the person accidentally spilled a cup of hot coffee all over himself. In addition to the two manipulated variables, the experimenters *measured* the subjects' self-esteem. They found that the pratfall increased liking for the competent stimulus person only when the subjects had average levels of self-esteem. Subjects with very high or very low self-esteem liked the competent person better when his perfection was unmarred by a blunder. Thus, the authors learned that the generality of the original finding was limited to subjects of average self-confidence, and this refinement in their knowledge about the applicability of the "pratfall effect" led to new and interesting hypotheses about the phenomenon. In this experiment, of course, we would not claim that low (or high) self-esteem *caused* the subjects to reject the blundering hero; we can make no inferences about self-esteem as a causal variable. Statements about causality can be made only when the independent variable consists of treatments that are randomly assigned to subjects.

But what if we are interested in the effects of some subject variable? Do we have to resort to a nonexperimental method and give up the chance to make causal statements? In many cases it is possible to effect a temporary change,

[2] For example, the extensive literature on sex differences and research in this area is important and interesting. However, the problems of drawing causal inferences from such research have been reemphasized by the women's movement: Observed "sex differences" may be due to differences in upbringing, education, social pressures, the relationship between the experimenter and the subject, or any other number of experiential factors.

even in relatively important personality characteristics, through the judicious use of an experimental treatment. Ordinarily, one can influence such habitual modes of responding for only a short period of time and in only one small area, but that may be enough for the purposes of testing the hypothesis involved.

For example, Aronson and David Mettee (1968) varied subjects' self-concept experimentally, so that they could assign subjects at random to "high," "medium," and "low," self-esteem conditions. They did this by giving the subjects a set of personality tests at the first session of an experiment and giving them false feedback about their personality at the beginning of the second session. When they arrived at the second session, the subjects read evaluations of their total personality, supposedly based on the tests they had taken. Some received very positive assessments, some very negative assessments, and some were told that their tests had not yet been evaluated. Thus, subjects entered the real experiment (the second session) with temporarily induced high, "unaltered," or low self-esteem, based on personality-test feedback. They were then subjected to a situation where it was worth their while to cheat in a game of cards. Since the levels of self-esteem were under the control of the experimenters and were randomly assigned, the experimenters were able to conclude that the differences in self-esteem *caused* differences in the subjects' cheating behavior. Thus, when it turned out that subjects in the low self-esteem condition were more likely to cheat at a card game than were subjects in the high or medium self-esteem conditions, the experimenters could conclude that the low self-esteem treatment caused cheating. Had they relied on a personality measure, it could easily have been the case that unknown factors that *produce* low self-esteem *also produce* cheating.

Performing an Internal Analysis

Nonrandom assignment of subjects to experimental conditions is not confined to the use of personality measures in lieu of experimental treatments. It often takes place in more subtle ways, one of the most common occurring when the experimenter is forced to perform an **internal analysis** in order to make sense out of the data. Such a situation arises when the experiment apparently did not "work." The experimenter has assigned subjects randomly to treatments, but finds that when the dependent variable is measured, subjects in the two conditions do not behave differently as predicted. But suppose that besides measuring the dependent variable, the experimenter has had the foresight to provide a separate measure designed to find out whether the experimental treatment succeeded in producing the internal state that constituted the conceptual variable. This measure, often called a **manipulation check**, provides the experimenter with important information.

For example, suppose that the experimenter wants to study the effects of anxiety on ability to concentrate. In the design, the experimenter might want the treatment in one condition to make the subjects anxious and that in the other condition to keep them calm. In addition to measuring the subjects' ability to concentrate (the dependent variable), the experimenter includes a measure of anxiety to determine the success of the procedure. Now, suppose that in conducting the experiment, the experimenter finds that subjects in the "anxious" condition and those in the "nonanxious" condition exhibit no differences in ability to concentrate. The experimenter can return to the measure of anxiety and find out whether the anxiety-producing procedure was successful, that is, whether the subjects in the anxious condition show more anxiety than do those in the nonanxious condition on this measure. If there are no differences in anxiety (as measured by the manipulation check), the experimenter concludes that the anxiety-producing procedure was unsuccessful. Because no differences in anxiety were produced, the experimenter can hardly expect the dependent variable to show differences. That is, no differences in ability to concentrate can be expected between the subjects who were supposedly made anxious and those who were not.

In such a case, experimenters often reanalyze their data, using the subjects' responses on the manipulation check as a substitute for the independent variable. That is, subjects who appear to have been made anxious in the experiment (regardless of whether the experimenter tried to *make* them so) are considered to belong to one condition (high anxiety); those who did not become anxious, to the other (low anxiety). The experimenter *now* predicts that the dependent variable measure will differ significantly between these two reconstituted groups of subjects. This is an internal analysis.

Let us pursue this example. The experimenter attempts to vary the extent of anxiety experienced by the subjects and then measures their ability to concentrate on some test. The measure of concentration is the number of problems the subject finishes on a concentration test composed of 10 problems. The data are shown in Table 7-1.

Clearly, there are no differences between the two conditions in ability to concentrate as measured by the number of problems subjects successfully completed. But suppose that the experimenter has conducted a manipulation check, asking the subjects to rate how anxious they felt on a scale from 0 (no

TABLE 7-1
NUMBER OF PROBLEMS COMPLETED BY SUBJECTS RANDOMLY ASSIGNED TO HIGH-ANXIOUS AND LOW-ANXIOUS CONDITIONS

Subject	High-anxiety condition	Subject	Low-anxiety condition
1	5	6	2
2	10	7	2
3	2	8	8
4	3	9	9
5	10	10	9
Total completed	30		30
Average	6		6

TABLE 7-2
RESULTS OF MANIPULATION CHECK, BASED ON SUBJECTS' SELF-REPORTED ANXIETY RATING

Subject	High-anxiety condition	Subject	Low-anxiety condition
1	3	6	3
2	1	7	2
3	2	8	0
4	3	9	0
5	0	10	1

anxiety) to 3 (very high anxiety). The results of the manipulation check are given in Table 7-2.

It is apparent that the treatment designed to produce anxiety was not very effective. Although the overall level of anxiety in the high-anxiety condition is higher than it is in the low-anxiety condition, some subjects who were given the treatment designed to keep them calm (low-anxiety condition) are actually more anxious than some who were given the high-anxiety treatment. The experimenter reasons, in effect, "Well, my treatments didn't work very well, but my *hypothesis* about anxiety and concentration may still be true." So the experimenter now divides the subjects into two new groups, placing those who scored high (anxiety rating of 2 or 3) on the manipulation check in one group and those who scored low (0 or 1) in another group, and performs the internal analysis shown in Table 7-3.

When one looks at the results of the internal analysis of these data, anxiety *does* seem to be related to ability to concentrate. Such data can be useful and provocative, but statements about causality should be approached cautiously. Indeed, since the effect was not due to the experimental variable, no causal statement can be made. Some of the "highly anxious" subjects may have been made so by the anxiety manipulation, but clearly others were anxious for other

TABLE 7-3
INTERNAL ANALYSIS TO TEST HYPOTHESIS ABOUT ANXIETY AND CONCENTRATION

Subjects with high scores on anxiety test			Subjects with low scores on anxiety test		
Subject	Problems completed	Anxiety rating	Subject	Problems completed	Anxiety rating
1	5	3	2	10	1
3	2	2	5	10	0
4	3	3	8	8	0
6	2	3	9	9	1
7	2	2	10	9	0
Total	14			46	
Average	2.8			9.2	

reasons. Since people who become anxious easily may be different from people who do not, we are dealing with a set of unknown variables. We don't know what made the subject check a 0 or 1 instead of a 2 or a 3 on the anxiety scale (it certainly wasn't the anxiety manipulation), and thus we don't know what was responsible for the subject's performance on the test. If the manipulation check was given at the end of the experiment, it is possible that subjects who realized that they weren't getting many of the problems done *became* anxious about being so slow. In this case, the direction of causality would be the reverse of that proposed by the experimenter; instead of anxiety causing poor performance, poor performance is causing anxiety. We have no way of knowing. The fact that the experimenter *once* assigned subjects to treatments at random is irrelevant: The current analysis is not based on that assignment, but rather on a *measured* subject variable, and thus it is a *correlational study*. The experimenter may find that the results of this internal analysis are encouraging enough to warrant a new experiment on the relationship between anxiety and concentration, this time using a more forceful manipulation of anxiety. The internal analysis is useful, since it provides such suggestive information. Until the second experiment has actually been conducted and the predicted results obtained, however, the experimenter is not justified in making causal claims about the effect of anxiety on concentration.

It should be noted that the useful function performed by an internal analysis—suggesting new hypotheses and interpretations of the data—is not limited to experiments in which the manipulation is a failure. In a successful experiment the results of an internal analysis may lead to qualifications of the main results, suggestions about the principles of operation of the variables involved, hypotheses for further experiments, or even assurance that some potentially confounding variable is irrelevant. A provocative discussion of these phenomena is given by Lee Cronbach (1975). In some cases an internal analysis can indicate how likely it is that a manipulated personality variable duplicates the effects of the corresponding "real" personality variable. For example, in the Aronson-Mettee (1968) study, the experimenters manipulated self-esteem through false feedback from personality tests and found that low self-esteem led to more cheating. In addition to manipulating self-esteem, however, the authors had a *measure* of self-esteem from one of the personality tests that the subjects had taken during the first session. They found that subjects whose personality test scores actually *did* indicate low self-esteem tended to behave like subjects who were placed in the "low" condition on the basis of false feedback. Similarly, subjects with high "chronic" self-esteem tended to behave like subjects with high "manipulated" self-esteem. Thus, by performing an internal analysis, the experimenters gained additional information which increased their confidence that it was really self-esteem that was being affected by the personality-feedback treatments. This illustrates another function for internal correlations using a measured subject variable: providing support for the experimenter's conceptual definition of a manipulated variable.

Allowing Subjects to Choose Their Own Conditions

Another situation in which treatments are assigned nonrandomly occurs when subjects assign themselves to the experimental condition: **self-selection**. In certain experimental situations the subject is given a choice about what to do. The experimenter then compares the subsequent behavior of subjects who choose one alternative with that of subjects who choose the other. For example, in an experiment designed to study how moral attitudes change after people have been exposed to a temptation to cheat, Judson Mills (1958) gave sixth-grade subjects a test as part of a "contest" in a situation in which it was either very easy or very difficult to cheat. It turned out that there were honest subjects and cheaters in all the conditions and that their attitudes about cheating differed after the contest was over. Since the subjects had not been assigned to conditions by the experimenter but rather decided for themselves whether or not to cheat, it is impossible to say that cheating *caused* differences in their attitude change. If this had been all there was to the experiment, the results would not have been very useful; there may in fact be important differences between those who choose to cheat and those who do not. In relinquishing control of the situation to the subjects, the experimenter is left with a nonexperimental study.

Curiously, the problem of choice is a particularly sticky one, since several interesting questions in social psychology involve hypotheses about the effects of making a decision. To study these questions, it is obviously important to give the subjects a perception of free choice, for if they think that their decisions are made for them by the experimenter, their behavior will not be relevant to the basic question. Yet the perception of choice must remain nothing more than a perception, for as soon as the subjects take advantage of it, we are beset by the problem of nonrandom assignment.

In some experimental situations this problem can be solved by using *instructions* that create a perception of choice, although little choice is actually present. For example, one might wait to present some of the drawbacks of a choice until after the subject has made an irrevocable decision. In presenting the negative aspects of the chosen alternative, the experimenter can often create a strong impression that any reasonably intelligent person would have been aware of these drawbacks and that, in fact, hardly any other subjects chose that alternative.

A different solution to this problem is to explore the situation and to pilot-test until finding a level of a variable just sufficient to inhibit subjects from actually choosing a "wrong" behavior. For example, in the experiment by Aronson and Carlsmith (1963) in which children were given either a mild threat or a severe threat to prevent them from playing with a desirable toy, it was important that the threat be strong enough, even in the mild condition, so that no children actually played with the forbidden toy. However, the threat could not be too strong, since the experimental hypothesis hinged on the child's not having a terribly good reason for leaving the toy alone. The situation had to be

ıch that although making a choice about whether to play, the child would al-
ays end up choosing not to play and would be bothered by the lack of a good
ːason for this choice. It is sometimes possible to find such a level by careful
retesting.

Such pretesting can also be used to create situations that will reliably elicit
ıny kind of behavior the experimenter wishes to study; it is not limited to stud-
es in which choice is the independent variable in the experiment. Carlsmith,
Ellsworth, and Jane Whiteside (1968) conducted an experiment in which sub-
jects who were about to take an experimental test were "tipped off" by a con-
federate; that is, the confederate told them how to do well on the test. The
independent variable was confession; the experimenters were interested in
whether subjects who confessed to having heard about the experiment would
still show the compliance effect reported in other studies of guilt. Thus, the
experimenters wanted subjects in one condition to confess that they knew the
trick and subjects in the other condition not to confess. After extensive pre-
testing, the experimenters found a situation which invariably induced confes-
sions. Since they were interested in the effects of confession itself rather than
in the *decision* to confess, it was easy to arrange a no-confession condition:
The experimenters simply left the room before the subject had a chance to
confess. Coarse tactics (like leaving the room) are often useful when the inde-
pendent variable involves some interpersonal behavior which the experi-
menter wants to elicit in one condition and to prevent in another condition.
The experimenter simply removes the opportunity.

CREATING AN EMPIRICAL REALIZATION OF THE INDEPENDENT VARIABLE

An experimenter has three major questions when trying to construct an em-
pirical realization of the independent variable. First, what specific event
should be used for the treatment in the experiment? Second, how should this
event be presented so as to have maximum impact on the subject, and how
does the experimenter know whether the intended impact has been achieved?
Third, if the predicted results are such that they could be spuriously achieved
by "cooperating" subjects (or distorted by uncooperative subjects), how does
the experimenter keep the subjects from guessing what effect the independent
variable is supposed to have on their behavior?

There are two general classes of experimental treatments; the empirical re-
alization can consist of (1) a set of instructions to the subject or (2) an "event"
to which the subject is exposed. In practice, these two techniques are not al-
ways separable; they usually blend into each other. Most "event" manipula-
tions contain verbal instructions to the subject, at least as a means of setting
the stage. In experiments in which the independent variable is manipulated by
means of instructions, the instruction often consists of a description of events
that might happen to the subject. Nevertheless, it is useful to separate the two
techniques conceptually. A good illustration of the manipulation of an inde-

pendent variable primarily through the use of instructions is to be found in
well-known "group cohesiveness" experiments conducted by Leon Festinger
and his colleagues (for example, Back, 1951; Festinger & Thibaut, 1951;
Schachter, 1951). In these experiments the cohesiveness of a group was varied by
informing the subject that the group members were especially selected so that
they would like one another (high cohesiveness) or that, despite repeated at-
tempts, the experimenter was unable to accomplish this feat (low cohesiveness).

In contrast, were such an experiment to be carried out by an experimenter
who wished to avoid instructions and instead present the subject with a vivid
"event," we would expect minimal verbal description. Instead, a group of
confederates would agree with any opinion voiced by the subject, approve any
actions, and express effusive positive regard at every turn. Another subject, of
course, would be disagreed with, disapproved of, and disliked by a set of con-
federates well trained in this unpleasant task.

Theoretically, the difference between the two techniques is intimately tied
up with the issue of control versus impact. Typically, when things happen to a
subject, we have much less control over them and much less confidence in
how to interpret the subject's reactions than we do when we read instructions.
However, it is almost always the case that events that happen during the
course of an experiment will have a far greater impact on a subject than will a
mere set of instructions. The crux of the problem is illustrated by the following
example: Stating that a person and the other members of a group will like each
other is almost certain to have less impact than presenting the person with
real people doing likeable things. But can we be sure that our subjects will
interpret these behaviors as likeable? Will they assume that the others like
them? Once again, we find that the choice between two approaches is neither
absolute nor easy. Both techniques have advantages, and the experimenter
must choose the one that better fits the particular problem. Finally, the exper-
imenter's choice is likely to be a combination of both. Below we try to accen-
tuate the distinction and point to important considerations for each technique.

Instructions

The investigator who chooses to use instructions must attempt to make them
interesting and forceful, since instructions are usually weak in impact. *First*,
the experimenter must make sure that the instructions command the subject's
attention; the subject whose mind is on other things may never even attend to
the manipulation. The experimenter should speak clearly and emphatically,
maintain eye contact with the subject, and pause to let the important points
sink in.

Second, the experimenter should ask questions to make sure that the sub-
ject has understood the essential features of the manipulation. If deception is
involved, this may be a delicate task, since the experimenter must ask the
questions and get the relevant information without revealing the purpose of the
experiment.

Third, it is a good idea, when possible, to set up a situation in which the ubject *must* understand the instructions in order to know what to do in the est of the experiment. That way, the subject will be motivated to understand he instructions, and the behavior of one who doesn't comprehend instructions will reveal the misunderstanding to the experimenter in the rest of the exper-ment. The importance of motivating subjects to attend to and understand task instructions cannot be overestimated, especially in judgment studies in which instructions are often designed to direct subjects' attention to the relevant di-mensions of the stimuli presented to them. Unless subjects are alert and atten-tive, and motivated to stay that way, all the care and effort devoted to system-atic control over stimulus conditions will have been wasted.

Fourth, instructions should be kept simple. Complicated instructions are more likely to lead to variability in the subjects' interpretations and may also make it more difficult for the experimenter to know what aspects of the in-structions constitute the effective stimulus. The experimenter who is manipu-lating several independent variables in the same experiment should try to avoid putting more than one manipulation into the instructions; instructions containing several manipulations are almost certain to be more complicated and difficult to absorb than those containing only a single manipulation. In-structions, like other types of treatment, can be pretested and revised in order to increase their clarity and impact and reduce the variability of the subjects' interpretations.

Sometimes the impact of an instructional manipulation can be increased by preparing the subject for it beforehand. For example, in the group-cohesiveness experiments mentioned above, the subjects were told upon re-cruitment about the experimenters' effort to create highly cohesive groups and were asked to answer some questions which would enable the experimenters to place them in a group whose other members had similar interests. The an-swers to the questions were not used, but presumably this advance informa-tion served to pique a subject's interest in the composition of his or her par-ticular group and to render more plausible the instructional treatment that took place during the actual experiment.

Another way to make sure that the subject is attending to a verbal treatment is to put the manipulation in the form of false feedback about an event that he or she has already experienced in the course of the experiment. Individual sub-jects can be told that they have scored high or low on a test they have taken, that opinions they have already expressed agree with those of another subject, and so on. Groups can be given false results of secret balloting or can be told that their interaction has been high or low on some dimension (cooperation, quality of performance on a group task, inhibition, etc.). Here, an event is used to get the subjects involved, so that a verbal treatment (presumably based on the event) will have greater impact.

An experimental treatment of this sort falls somewhere on the continuum between simple verbal instructions and the construction of an "event." The main drawback of this technique is that the report may not be plausible to the subject, who may have had different experiences with the event. A subject who cannot solve more than one or two of the questions on a test is unlikely to believe she received a high score; similarly, a subject who sits frowning and taciturn on the fringe of a group is not likely to believe a vote indicating that she is the most popular member of the group. The experimenter should try to make the event ambiguous enough so that the subject is deprived of informa-tion that could contradict the false feedback.

For example, in one experiment Lee Ross, Mark Lepper, and Michael Hubbard (1975) gave subjects the task of discriminating between fake suicide notes and notes actually written by people who had attempted suicide. Be-cause the subjects had little experience with similar tasks and no clear defini-tion of the requisite skills, and because there was no way of telling whether one's guesses were correct or incorrect, it was easy to convince the subjects that they succeeded or failed at the task, much easier than it would have been if the experimenter had tried to give false feedback based on a more objective or familiar test, such as an IQ test.

Events

We used the term "events" in connection with the independent variable as if events are all more or less alike. This is not the case. Sometimes an event is something that *happens to* subjects: A confederate attacks subjects' phi-losophy of life; subjects inadvertently break a camera or blow up an expensive piece of equipment; they are stared at or touched; their perceptions are called into question by a unanimous majority; they are threatened with an electric shock; and so on. Other times, events are far less personally involving and far more benign: Subjects read newspaper headlines; they watch a videotape in which someone spills coffee on himself; they see another person's answers to an attitude survey; they look at cartoons and rate how funny they are; and so on.

One dimension on which these events can be ordered is that of impact on one end and control on the other. Of course, one characteristic of a "di-mension" is that events can fall anywhere along a continuum. For example, at what point does group pressure cease to matter much to subjects? When mem-bers of a unanimous majority disagree with subjects' judgments and berate them for their stupidity face-to-face? When subjects hear the voices of the ma-jority over an intercom? When a series of flashing lights communicates the majority's disagreement? When subjects are handed a note summarizing dis-crepant judgments of the unanimous majority? At what point will these stimuli fail to capture and hold subjects' attention?

In current social psychological research, particularly the "judgment" stud-ies carried out by psychologists interested in attribution and social cognition, subjects frequently make judgments about written descriptions of hypothetical people or events, rather than events they experience directly. Control is high, impact is low, and researchers are faced with a troublesome problem: The way people reason about hypothetical events may differ from the way they reason about real ones.

Not all social cognition research uses simple verbal stimuli and paper-and-pencil measures. Throughout the history of social psychology, research styles have ranged from highly controlled, relatively unrealistic "pure" techniques to noisy, involving techniques high in experimental realism. The Hovland attitude change group relied heavily on written communication and paper-and-pencil measures; La Piere studied attitudes by traveling around the country with a Chinese couple and attempting to register at hotels. Likewise, although many judgment studies rely on vignettes, it is possible and often preferable to study social cognition using high-impact events.

Lee Ross, David Greene, and Pamela House (1977), for example, asked subjects to read descriptions of hypothetical situations and to indicate which of two responses they would make. They found that people see whichever choice they make as relatively common. The authors termed this the "false consensus" effect. People also believe that making the other choice reveals quite a bit about a person's personality, while their own choice (the "normal" choice) is not particularly revealing of personal disposition.

So far, the study was like many social cognition studies, a paper-and-pencil study of responses to hypothetical situations. The subjects never really had to choose between two alternatives. In a follow-up study, the experimenter told subjects that he was interested in attitude change and "communication techniques," and asked them to exit the building wearing a large sandwich-board sign with the painted message "Eat at Joe's" and to observe people's reactions. Some subjects agreed to do it; others did not. The results replicated those of the first study: Subjects who agreed to wear the sign thought that 62 percent of those who were asked would agree; subjects who refused thought that 67 percent would refuse. Those who agreed made inferences about the kind of person (up-tight, humorless) who would refuse; those who refused made inferences about the kind of person (attention-grabbing, frivolous) who would agree. The realistic second study convincingly demonstrates that the false consensus effect is not unique to paper-and-pencil games, but occurs when people really commit themselves to decisions. As Lee Ross argues, "[The] cost of procedures that simultaneously eliminate both realism and noise are . . . formidable. One loses the capacity to produce theories and empirical generalizations that are not merely 'correct,' but also powerful relative to competing processes, i.e., theories and generalizations likely to be useful in the goals of understanding, predicting, and controlling behavior in real social contexts"(Ross, 1987, p. 143).

In recent years, experimenters have made increasing use of interactive microcomputers. Such microcomputers are frequently used to present stimuli to subjects and to record their responses. Not only does such technology improve the precision with which the stimuli can be varied and with which sub-

jects' responses can be recorded, but it also makes for a more interesting, vivid and engaging task for subjects.

Throughout the remainder of this section, we devote out attention to experimental events which fall toward the impact end of the continuum. This is *not* to say that experiments in which "events" are verbal stimuli are not valuable and important tools for understanding social behavior. It *is* to say that because impactful events or "strong" manipulations that happen to subjects are often complex and more difficult to "pull off" successfully, there are more problems and methodological concerns associated with their use. That being the case, it seems prudent to provide a more extensive discussion of ways in which these events can be implemented, always with an eye toward enabling experimenters to infer valid causal relationships between empirical realizations of the independent variable and some effect of those realizations.

Often, we favor the use of an event rather than a set of instructions, despite the problems posed by the difficulty of interpretation. The meaning of an event can often be ascertained through the process of systematic replication, whereby different events with overlapping meanings can be used to test the same hypothesis in different experiments. In addition, it is often possible to increase the likelihood that a subject will arrive at the intended interpretation of the event. Sometimes this can be accomplished through the skillful combination of events and instructions. An example is an experiment by David Landy and Aronson (1968) in which the investigators wanted to find out if people react more strongly to personal evaluations if they regard the evaluator as a "discerning" person. The authors predicted that subjects who are evaluated positively by a discerning confederate should like the confederate better than if he were not discerning. However, subjects who are evaluated negatively by a discerning confederate should dislike the confederate more than if he were not discerning. In other words, they were predicting an interaction between type of feedback (positive versus negative) and the confederate's power of discernment (discerning versus not discerning). How does one vary the subjects' perception of the confederate's discernment? One could do it by instructions; that is, one could simply say to the subject, "Say, by the way, this fellow is really a discerning person; I thought you might be interested." But for reasons of credibility, the investigators felt that it would be better to allow the subject to arrive at this judgment independently. They therefore started out by having the confederate perform a task in the presence of the subject; the task was such that by varying the confederate's behavior, the subject might easily be induced to regard the confederate as either discerning or not discerning.

The word "might" epitomizes the problem of manipulating events. The subject *might* interpret this behavior in a multitude of ways. In order to maximize the likelihood that the subject would consider this behavior relevant to discernment and nothing else, the investigators: (1) asked the subject to observe the confederate's behavior on a task (in the context of an experiment on social judgment); (2) told the subject that "degree of discernment" was an aspect of the confederate's behavior that was of particular interest to them; (3)

ıked the subject to rate the confederate's discernment; (4) informed the sub- ct exactly how the confederate's behavior might reflect either high or low iscernment; (5) had the confederate behave either one way or the other; (6) ad a handy and meaningful check on the manipulation, in the form of the sub- ect's actual rating. It can readily be seen that this technique is a compromise. t may lack the impact of obscene words or electric shock, but it has more mpact than does a set of instructions. At the same time, it capitalizes on the easy interpretability of verbal instructions by focusing the subject's attention on the variable (discernment) that the experimenter wants to manipulate. In effect, Landy and Aronson told the subjects in advance what the event would mean and then created the event, giving it the clarity of a set of verbal instructions, but without sacrificing the impact characteristic of event manipulations.

One advantage of events over instructions is apparent when we consider the problem of subjects' becoming aware of our hypothesis and allowing this awareness to influence their behavior. For many experiments in social psychology, the ideal empirical realization of an independent variable is an event that the subject does not connect with the experiment at all. This is the best way to guarantee that the subject has no hypothesis of concern to the experimenter. Frequently, it is also the best way to guarantee that the manipulation has an impact on the subject. For example, a subject told that a particular communication was written by T. S. Eliot (Aronson, Turner, & Carlsmith, 1963) may yawn and ignore this information or, even more important, may have the detachment to sit back, look at the ceiling, and begin to hypothesize that the experimenter is concerned with the effect of a high-prestige communicator. For this reason, the manipulation is a relatively weak one. But consider a subject faced with a person to whom he has been delivering electric shocks who is now screaming, beating on the walls, and begging to be let out of the room (Milgram, 1973); consider a subject who has a dozen wires attaching him to an apparatus that has suddenly short-circuited, so that it looks as though he may be in danger of being electrocuted (Ax, 1953); consider a subject who has just broken someone else's expensive camera (Regan, Williams, & Sparling, 1972); consider a subject who, to his dismay, discovers that a group of normal-looking people all judge the length of a line differently than he does (Asch, 1951). These subjects are very unlikely to yawn or to start playing intellectual games; they have a problem of their own—for example, what to do about this poor fellow who is screaming in the next room.

Several classes of techniques have been used successfully to present the independent variable as an event unrelated to the experiment, so as to have a maximum impact on the subject. Not perceiving them as part of the experiment, the subject will not speculate about their purpose or their relationship to the experimenter's hypothesis. Many experiments have actually used a combination of several of these techniques.

The Accident Perhaps the most effective of these techniques, but one of the most difficult to set up, is the "accident." An experiment by Albert Ax (1953), designed to compare the physiological reactions characteristic of fear and anger, provides one of the best examples of this procedure. The subjects believed that the experiment was designed to study metabolic differences between hypertensives and normals and that all they had to do was lie on a mattress and listen to music while their physiological responses were being measured. In the fear condition, while the subject was lying there with leads from the recording apparatus attached to his ear, leg, chest, abdomen, face, and fingers, one of the finger electrodes "accidentally" began to emit shocks that gradually became stronger. When the subject spoke to the experimenter about the shocks the experimenter expressed alarm and acted as though the situation was out of control, pushing a lever which caused sparks to fly from the apparatus near the subject and exclaiming in confusion that there appeared to be a dangerous short-circuit in the system. This procedure enabled the experimenter to produce fear without its being blunted by the subject's knowledge of being in a protected environment. From the subject's point of view, the "accident" occurred outside of the context of the preplanned experiment and thus retained the impact and unpredictability of a real-world event.

Another ingenious example of the accident technique comes from the experiment by John Wallace and Edward Sadalla (1966), in which the subjects worked on a task that involved a large, expensive looking machine. During the course of the experiment, a confederate induced the subject to push a forbidden button labeled "Do Not Touch" on the experimental apparatus, whereupon the machine exploded and appeared to be totally destroyed. Festinger and Carlsmith (1959) used a similar technique when they told the subject that due to an accident, the regular confederate had not shown up and then asked the subject to play the role of the confederate. Indeed, the "accident" procedure has been used so frequently and so successfully that it might be said that part of being a good experimental social psychologist involves learning how to say "whoops" convincingly.

The Confederate A variation on the accident procedure is to use a confederate who introduces the manipulation of the independent variable. Once the confederate has been carefully trained to produce a convincing performance, this procedure may be easier to carry out than the accident technique. Like the accident technique, however, the introduction of the independent variable by confederates may arouse suspicion in some cases. In one experiment by Aronson and Darwyn Linder (1965), the subject heard herself being evaluated by the confederate in the context of a verbal-reinforcement experiment. This was described to the subjects as a standard part of the procedure. Although it succeeded for most subjects, a few found it odd that such personal material would be used to measure the effects of reinforcement on verbal behavior. Thus, impact and credibility are not automatically achieved by the simple introduction of a confederate into the procedure.

Often, it can be made to appear that the manipulation is not a regular part of the procedure but is something which the "other subject" (really a confederate) just happened to do on this occasion. For example, Schachter and Singer (1962) attempted to manipulate euphoria by having a confederate waltz around

e room, shooting rubber bands, playing with hula hoops, and practicing hook
iots into a wastebasket with crumpled balls of paper. Presumably, the subject
iterpreted this behavior as a unique event unrelated to the experimental pro-
edure. Similarly, Jack Brehm and Ann Cole (1966) attempted to manipulate
heir subjects' feeling of obligation toward another person. They accomplished
his by having the confederate go out to buy a soft drink from the vending ma-
chine and "thoughtfully" bring one back for the subject too. This "unique"
event, although it appeared to be unrelated to the experiment, succeeded in
making the subject feel somewhat uncomfortable toward the confederate. John
Darley and Bibb Latané (1968) used a confederate very effectively in an ex-
periment designed to discover the factors that affect people's willingness to
help others in an emergency. Subjects believed that they were members of a
group discussion that took place over an intercom system (in order to protect
the privacy of the other members, since they were supposed to discuss per-
sonal problems). The confederate (posing as another subject) hesitantly re-
vealed that one of his worst personal problems was that he was prone to ner-
vous seizures. Later on, when it was the confederate's turn to talk again, he
began to stutter and gasp, and his voice grew louder and more incoherent as he
stammered out that he was on the verge of a seizure and begged for help.
Clearly, the subjects perceived the event as exceptional, something completely
beyond the bounds of the experiment; as one subject said to herself, "It's just
my kind of luck, something has to happen to me" (1968, p. 381).

The Whole Experiment as a Treatment A third method of having the em-
pirical realization of the independent variable perceived as separate from the
experiment is to use the whole experiment (as perceived by the subject) as the
treatment, measuring the dependent variable at some later time. This tech-
nique is rather difficult in practice, since subjects may think that some of the
experimental events are intended to affect them; when the procedure is carried
out well, however, they are unlikely to perceive the whole experience as the
treatment. For example, Carlsmith and Gross (1969) performed an experiment
designed to investigate the effects of hurting someone on subsequent compli-
ance. In their experiment they induced the subject to administer electric
shocks to a confederate. The entire procedure was presented as a learning ex-
periment, with the subject playing the part of the teacher and administering
shocks whenever the confederate made an incorrect response. After perform-
ing this task, the subject was given an "explanation" of the experiment and
was told that it was over. Shortly thereafter, the confederate asked a favor of
the subject, who responded to this request without realizing that it was part of
the experiment.

THE ISSUE OF STANDARDIZATION

Ideally, the empirical realization of an independent variable is forceful enough
to have maximum impact and clear enough to generate the intended interpre-

tation in all subjects. There is no list of specific techniques for achieving this
ideal. However, some important general guidelines can be established. At the
heart of the question is one crucial, yet frequently misunderstood, point: It is
extremely important for all subjects to be in the same psychological state as a
result of the manipulation of the independent variable. This does not necessar-
ily mean that all subjects should be exposed to the identical independent vari-
able. This *does* mean that the experimenter's skill and wisdom should be used
to make sure that all subjects arrive at a similar understanding of the instruc-
tions (or the implications of the "event" manipulation). To achieve this goal,
the experimenter should take considerable latitude in delivering the instruc-
tions. This is a tricky issue and is one that may raise doubts in the minds of
many investigators. Our point is this: In their zeal for standardization, many
experimenters make an effort to have all instructions to the subjects tape-
recorded, printed, or computerized, attempting to make sure that all subjects
are exposed to identical stimuli. Such an effort is admirable, but in practice it
ignores the fact that people are different, and as a consequence, the same
instructions do not mean the same thing to all subjects. More prosaic, yet
more important, subjects differ greatly in their ability to understand instruc-
tions. For example, one of the most common mistakes the novice experi-
menter makes is to present the instructions too succinctly; consequently, a
large percentage of the subjects fail to understand some of the important
points. To ensure that all subjects understand what is going on in an experi-
ment (especially one as complicated as most social psychological experiments
are), a good deal of *redundancy* is necessary.

More important than simply providing redundancy, however, is ensuring
that each subject *fully understand* all the instructions and events that occur in
the experiment. The first thing the experimenter should try to do is to make
these instructions and events as simple as possible without making them un-
interesting. If the experimental situation is still complex, as is usually the case,
the experimenter should make sure that the subject is not missing or misinter-
preting anything. The experimenter can do this only by a combination of clear
instructions, questions, pauses, and probes and by repeating or paraphrasing
key parts of the instructions until satisfied that the subject is completely clear
about all of them. The point seems self-evident, but is has been our experience
that many experiments have failed precisely because the instructions were
never made clear and redundant enough to get through to all the subjects.

An example may point up the difficulties. In an experimental investigation
of compliance (Carlsmith & Gross, 1969), the experimenters were interested in
finding out whether compliance with a request is more likely shortly after a
person has hurt someone, and if so, whether the person is more likely to com-
ply with a request made by the individual who has been hurt or with one made
by a bystander who witnessed the event. As the experiment was set up, there
were three participants: a teacher (in some conditions the subject, in others a
confederate), a learner (always a confederate), and a witness (in some condi-
tions the subject, in others a confederate). It was important that the subject
know which of the two confederates was the learner and which played the

ther role. However, the learner-confederate could only be kept blind to what ondition the subject was in by being kept ignorant of whether the subject in ny given session was the teacher or the witness. Accordingly, the three par- icipants were separated by partitions so that they could not see one another; the subject was brought into the experimental room alone and was given an explanation of the experiment, with heavy emphasis on which person (sitting in which seat) would be learner and which would play the other role. Despite strong attempts to make these instructions redundant, 12 of the first 15 subjects missed enough of the information so that they did not know which confederate was in which role. Luckily, postexperimental probing revealed this weakness in the procedure, which would have vitiated the whole experiment had it not been discovered, since the subject would not know whether the person making the request had been hurt or not. The identity of the two confederates was a key fact of the experiment, so in order to pound that fact into the subjects' heads, the experimenters revised the procedure to make the distinction between the confederates so blatant that no one could mistake them. The confederates' clothing differed greatly; the subject was introduced to only one of them before the experiment began; and the instructions were made "overly" redundant, so that finally the identities of the two confederates were impressed on the subjects' minds. The point of this is not that subjects are thick-headed, but that experimenters are often so familiar with their own procedures that they are unable to put themselves in the position of someone who is hearing the experiment described for the first time.

There should be little argument about the merits of using simple and redundant instructions to ensure that the subject understands the experiment. What we are suggesting goes beyond that, however. First, although instructions should be clear and repetitious, it is unwise to make them too repetitious, or the bright subjects may become bored or annoyed. However, some subjects are bound to miss the point, even when the instructions are exceptionally clear and repetitious. Thus, the same set of instructions may bore some subjects and baffle others. How can this be avoided? It does not take much expertise to pick out the subject who cannot grasp the meaning of the instructions. Often, such a subject would like to ask for more information but is too timid to interrupt, and so sends out a variety of obtrusive nonverbal cues signaling perplexity. Usually, the experimenter will notice these cues; in fact, in most cases it may be difficult *not* to notice them. But some experimenters may feel that they should ignore this information and should instead continue to follow their standardized scripts, so as not to introduce extra variability into the experimental situation. We feel that when situations of this sort arise, the experimenter should not try to maintain a mechanical "standardized" format. If, in the course of delivering the instructions, the experimenter sees a vacant or uncomprehending expression on the subject's face, efforts to get through to the subject should be increased, even at the expense of departing from a standard set of instructions. Certainly, if the subject did manage to overcome timidity and make a verbal request for clarification, few experimenters would refuse to give it. We are simply suggesting that the same procedure be followed when the subject's uncertainty is obvious, albeit unspoken. Otherwise, one advantage of the use of standardized instructions, reduced random error variance, will be offset by an increase in that same error variance caused by some subjects' failure to comprehend task instructions.

The situation is analogous to teaching, in that the experimenter has a certain quantum of information to get across. Few teachers would continue to read a prepared lecture after all the students had dropped their pens and were staring with blank faces; instead, they would go back and try to find out where the difficulties lay, in an effort to attain their original goal of communicating the information. The experimental situation is similar: There is no reason for an instructor to disregard feedback from the person being instructed, just because their relationship is one of experimenter to subject. Of course, the experimenter should keep a record of exactly what was said to each subject, in the interests of replicability and a fuller understanding of the behavior of any subjects who deviate markedly from the rest.

Again, we anticipate that many experimenters will disagree with us, suggesting that standardization is the hallmark of an experiment. We agree, but exactly what is it that should be standardized? What the experimenter says, or what the subject understands? We feel that the more variability there is in the subjects' comprehension of the experimental operations, the more likely it will be that the changes caused by the independent variable will be obscured. Of course, by allowing the experimenter to depart from a standardized script, one may increase the possibility of introducing a systematic bias. But if proper techniques are employed to eliminate bias, this ceases to be a problem. (Some of these techniques are discussed in Chapter 9.) In particular, if the experimenter who is giving the instructions is unaware of the subject's experimental condition, there is no way in which variations in the presentation can systematically bias the results. Similarly, in many cases the attempt to make sure the subject fully understands the situation takes place before the introduction of the specific experimental treatment and thus cannot bias the results. Clearly, we do not advocate any sort of flexible presentation in contexts in which such variations could introduce systematic error.

With the issue of standardization, as with many of the other issues we have discussed, we are faced with a tension between two important desiderata: On the one hand, it is critical for the experimenter, when writing up the research, to describe accurately the nature of the treatment. On the other hand, it is important that all subjects be in the same psychological state as a result of the experimental treatment. The problem may be easy to solve when the treatment is clear and simple, for example, when the treatment consists of an experimenter staring at some subjects and looking away from others, or touching some subjects and not touching others. But when the treatment calls for, say, sexual arousal of the subject, the situation is more complex.

We may take the case of sexual arousal as a polarizing example, one that moves beyond the presentation of instructions and into procedures that might entail the presentation of dramatically different stimuli to different subjects. Suppose that we want to study the effects of sexual arousal on liking. Our independent variable calls for sexually arousing the subjects in one condition. We could pick a selection of salacious passages from pornographic novels, collect them into a booklet, and have all the subjects in the sexual arousal condition read the booklet from cover to cover. In this case each subject is getting the same version of the independent variable. The effectiveness of the treatment depends on whether the particular selection of passages we have chosen has a widely generalized eroticism that will succeed in "turning on" all the subjects in the arousal condition. As an alternative, we could accumulate a large variety of potentially arousing material—books, pictures, movies, and so on—in which the themes as well as the media were diversified, and try to find the stimulus that was most arousing for each subject. Subjects' tastes might differ widely, and if we used this sort of procedure, not all subjects would be exposed to exactly the same stimulus. But presumably, if we were successful in our "something for everybody" strategy, all subjects *would* become sexually aroused. The first procedure certainly is superior on the dimension of "stimulus standardization." Although that is desirable when feasible, it may be more important to see to it that all subjects are in the state required by our conceptual independent variable, even though different amounts or even kinds of stimulation are required to produce this state. In other words, often our conceptual independent variable is really some sort of response, like sexual arousal, which we believe will act as an internal stimulus for subsequent behavior. In cases like this it may be more important to strive for standardization of this response, our real independent variable, than for standardization of the external stimulus intended to produce that response.

Consider the analogous situation in animal research. If we want to study the effects of arousal on some performance and we decide to manipulate arousal by administering an appropriate drug, we do not give the same amount to each subject. Typically, the variation comes about because we correct for some individual difference factor, such as the animal's weight. Thus, although we are administering different dosages to each animal, we are in fact administering a standard amount per kilogram of body weight. But at other times the level of the external stimulus might be set according to some aspect of the animal's performance. For example, we might introduce just enough rewarding brain stimulation to bring the animal to a threshold level of bar pressing so that the subject presses a bar on 50 percent of the occasions that the stimulation is received. Animals in one condition are at threshold, and in another condition they are not; the independent variable is the level of bar pressing, and it is the same for all animals in a given condition. The brain stimulation is the external stimulus that is designed to bring the animals to the appropriate level of the independent-variable dimension, and it is not necessarily the same for all animals. It may take different amounts of brain stimulation to bring different animals to threshold.

Some of the resistance to such flexible administration of the independent variable in social psychology stems from the fact that at the present state of development of our techniques, we lack accurate procedures (such as rate of bar pressing) for deciding when a subject has reached a standardized level on one of our complicated variables; therefore, we often have no clear criterion for deciding how the manipulation may need to be altered in order to bring any given subject to the intended level. Daniel Katz, for example, objected to these suggestions primarily on the grounds that our means of assessing the presence of the intended internal state are subjective and that by using such techniques, experimenters "no longer rely upon the variables created and manipulated but upon the psychological equivalence of effects produced" (1971, p. 277). Assessing the effectiveness of the treatment for individual subjects is not *necessarily* a matter of the experimenter's subjective impressions, however. For example, in order to determine amount of sexual arousal, we might attach a (male) subject to a phalloplathismograph, give him some arousing material to read or look at, measure his response, and if necessary, give him another installment, continuing to show him arousing stimuli until some criterion response has been measured. Of course, such a manipulation check may be difficult to justify within the context of some social psychological experiments.

The technique preferred will depend to some extent on the precise nature of the conclusion the experimenter wishes to draw. If interested solely in discussing the effects of reading a particular salacious short story (as an editor of *Playboy* or *Penthouse* might be), the experimenter would prefer the standardized stimulus presentation. If, however, the experimenter wishes to make inferences about sexual arousal per se, a group of sexually aroused experimental subjects is needed. Of course, most people would agree that an erection is better evidence for sexual arousal than the presentation of a standard set of stimulus materials. Presumably, if we could find as clear physiological evidence for criterion levels of other psychological states as we can for sexual arousal, there would be little resistance to the idea of varying the stimulus presentation in order to achieve these levels in all subjects.

Physiological responses are not the only type of response that can be used to indicate the effectiveness of the manipulation of the independent variable, however. The goal is to bring subjects to a given state, and as long as that state is specifiable in terms of some other measurable response such as verbal reports, overt behavior, or nonverbal cues, we see no reason to insist on using a standard stimulus presentation, rather than a standard response, as the criterion. Any measure that can be used as a manipulation check can be used as a criterion; the difference is simply that we are suggesting that the experimenter use the information as soon as it is available, rather than wait and find out that the manipulation was ineffective when it is too late to do anything about it.

As an example, consider the internal state of guilt, for which there is no known physiological measure. In the Carlsmith and Gross experiment (1969), guilt was operationalized by having the subject administer painful shocks to another person in the context of a teacher-learner experiment. The stimulus situation was relatively standardized; the teacher-learner interaction consisted f 15 trials, on exactly 9 of which the learner made a mistake, and the subject was required to administer a standard shock. The subject believed that each of the shocks was of the same voltage. The authors might instead have used a situation similar to the Milgram (1973) situation, in which the voltage of each shock delivered as punishment was higher than the last, and their empirical realization of the conceptual variable "guilt" might have been delivery of a shock of 300 volts or some other specific number. In this case, differing degrees of pressure might be necessary to bring different subjects to this level, but the level itself would be taken as the indication that the desired amount of guilt had been aroused. Whenever the subject reached that level, the induction could be terminated and the dependent variable measure introduced.

Katz also stated that "if one could assess clearly and accurately the psychological states of individuals, and if we knew how to produce them, then we would have no need for all the previous language about experimentation" (1971, p. 277). In the first place, the fact that we may be able to find a way to assess a given state in a given context does not imply that we can "assess clearly and accurately the psychological states of individuals" in general. The techniques suggested in this section are not intended to be used in every experiment, but only in situations in which some means of assessment is available. We do not insist that this criterion measure have guaranteed accuracy. Indeed, the issue of "accurate" assessment of psychological constructs raises issues that are much more general than the specific recommendations we are making (see, for example, Schneider, Hastorf, & Ellsworth, 1979); we do feel, however, that some form of criterion assessment is often better than no assessment, and for this reason we do not insist that our measures be perfect.

Second, in most experiments social psychologists have to make some assumptions about their abilities to produce and assess psychological states: not only in experiments in which an effort is made to standardize responses to the independent variable, but also in experiments in which response standardization is less important. The creation of any independent variable treatment designed to create a given state is an attempt to produce that state, and in a way it seems more ambitious for the social psychologist to presume to have found a technique so effective that it will work for everyone than it is to aim for such a technique but to realize that it might fall short in some cases. Likewise, the problem of assessment is not limited to assessing the success of the manipulation, but rather occurs in all social psychological experiments when the dependent-variable measure is taken.

The great difficulty with using the subject's response as a criterion, of course, is that we cannot always specify exactly what we did. But sometimes we can. To return to our example on sexual arousal, we might have a standard series of literary selections that we present one at a time until the subject is aroused, stopping at different places for each subject when a certain level of arousal is reached. Obviously, it is desirable to be able to specify in exact detail just what was done with a subject. Further, as techniques become better and we become more and more certain how to manipulate variables, we may

also expect to develop measures that will give a clear indication of whether the independent variable has influenced the subject in the manner intended.

Until that time comes, we should assume that there will be variability in the subject's understanding of the treatment, whether the treatment consists of a tape recording, a printed set of instructions, or an interaction with the experimenter. We are certainly not recommending that the experimenter *seek* to present variable instructions. The experimenter should start out by attempting to construct a standard operating procedure. Pretests should be conducted, and if the subjects show little consistency in their interpretations, the experimenter should modify the treatment until finding a standard procedure that does produce consistent responses. But even with all this careful preparation, not all subjects will see the stimulus in a way that conforms to the experimenter's expectations. Our recommendation is that the experimenter recognize that variability exists in the subjects' understanding of the instructions and that on some occasions a standardized manipulation may not fit. On these occasions, experimenters would do well to increase stimulus variability deliberately in order to decrease the variability in the subjects' understanding of the stimulus situation. Of course, these variations must be describable when the experimental procedure is written up and cannot be used unless adequate techniques for eliminating experimenter bias are employed.

Pilot Testing the Independent Variable

How can we be sure that we have a good empirical realization of our independent variable? How do we know if subjects are attending to the features of the situation we expect they will? How do we know that the complex series of events encountered by the subject really is arousing what we think or hope it is? We have discussed some of the answers to this question at an abstract theoretical level; here, we attempt to give some practical advice on how to determine whether a treatment is producing the intended conceptual variable. The most general technique for finding out just what a treatment is doing to people is to run some **pilot tests**. Pilot tests are not necessarily formal rehearsals of the experiment; one need not run pilot subjects through the entire experimental procedure from start to finish. The purpose of these "trial runs" is to check out the experimental paradigm for the existence of unforeseen technical problems—to get the bugs out of the procedure—and any technique that is useful for discovering these rough spots can be a suitable "pretest."

The discoveries made during pretesting and the consequent adjustments and modifications of the procedure represent one of the most important sequences of events in constructing an experiment. Informative as they are, these discoveries and revisions seldom, if ever, appear in the final, published version of an experiment. Thus, in this section we will rely even more heavily than usual on our own research and experiences.

During pretesting the experimenter can conduct long, probing interviews with the subject. Often, the subject is capable of providing valuable hints as to where the weaknesses in the treatment occurred, what parts of it were misunderstood, and where it evoked reactions different from those the experimenter intended. For example, in one study (Aronson & Carlsmith, 1962), the experimental treatment was designed to give the subject an expectation of success or failure on a test. Since the experimenters wanted the test to refer to an ability that was vague enough so that they could give the subjects convincing false feedback about their performance, they decided to call it a test of "social sensitivity." On each trial of the "test," the subject was to look at photographs of three men and pick out which one was Jewish. But when they tried the test out, pilot subjects became upset, arguing that it was impossible to recognize Jews on the basis of appearance. Those who did try it failed to build up any expectations about their future performance, even when they had attained consistently high (or low) scores on four blocks of 20 trials each. Because the subjects believed that judging a Jew from appearance was entirely a matter of chance, their perceptions of how they would do on the next series of trials was unaffected by the feedback about how they had done on the last series. On the basis of the information gained from these pilot subjects, the experimenters changed the test so that the subjects were asked to pick out schizophrenics rather than Jews, and from then on the experiment ran smoothly. Apparently, subjects are less threatened by the implications of picking the faces of schizophrenics than of Jews, and they readily accepted the level of ability that was experimentally communicated to them.

If deception is used, the pretest subject is the best source of information about the effectiveness and credibility of the cover story. Such interviews can, of course, be conducted while the experiment is actually being run, but it is usually during pilot testing that the most valuable information is obtained, since the experimenter still has the opportunity to make extensive alterations in the procedure without invalidating the experiment.

If one is particularly interested in the adequacy of the empirical realization of an independent variable which corresponds to some internal state, it is a good idea to interview the pilot subject right after the treatment, without continuing the whole experiment. If one waits until the experiment is over and then attempts to question the subject about the dependent variable as well, the subject may find it difficult or impossible to describe the effects of the experimental treatment. Studies of cognitive dissonance offer a good case in point. The general hypothesis in all dissonance studies is that subjects will do whatever they can to reduce the dissonance. Suppose that the experimenter is running pretests in order to find out if the manipulation really did produce dissonance. If subjects are not questioned until the end of the experiment, those who have succeeded in reducing all the dissonance might well report that the manipulation of the independent variable aroused no dissonance or discomfort whatever. Take, for example, the study by Festinger and Carlsmith (1959). In this study, subjects in the dissonance condition spent an hour working on a tiresome, apparently purposeless task and then were paid $1 to tell another subject that the task was fun and exciting. The treatment designed to produce

the dissonance was the payment of an inadequate sum for the false description of the task. According to the prediction, the subjects would reduce this dissonance by changing their opinion of the task in a favorable direction. If subjects were not questioned until the end of the experiment—when they had already changed their attitudes so as to reduce most of the dissonance which had been created—they might well report that they felt no discomfort about saying that the experiment was interesting, in return for payment of $1. If, however, they were interviewed immediately after agreeing to give the false description, they might be able to report more accurately on whether the treatment had caused the expected reaction.

The dissonance example provides an illustration of the major difficulty with the use of introspective reports: Too often, subjects are unable or unwilling to explain just what the effects of some treatment have been. Following an experiment, it is not at all uncommon for subjects to deny any feelings of the kind the experimenter expected to arouse, although their behavior throughout the experiment was just what would be anticipated if they had experienced precisely those feelings. This basic fact, of course, is the reason why psychologists have turned from introspection to behavioral research. For example, in the study on severity of initiation, the women who had undergone a severe initiation expressed favorable opinions of the boring group, as the theory predicted. When the experimenter explained his hypothesis to these women, many of them made such comments as: "Gee, that's a fascinating experiment, and I'm sure severe initiations really work that way on some people, but I loved that group because they said such good things."

As Timothy Wilson (1985) points out, the mental processes that *guide* our behaviors are different from the mental processes we employ when we consciously attempt to *understand* or *explain* our behaviors. When subjects attempt to reconstruct their inner experiences rationally, and tell the experimenter about them, they may fail to capture the elusive thoughts or feelings that gave rise to their original behaviors. Subjects' retrospective self-reports may thus at best be an imperfect approximation of their perceptions and feelings at the time an event occurred. As a consequence, even when subjects are asked to report their feelings immediately after the administration of the independent variable, they may not be able to give accurate descriptions. People are not always able to articulate subtle psychological states, and often they come up with plausible but inaccurate accounts of the influences on their behavior (Nisbett & Wilson, 1977). In addition, in some situations, subjects may feel inhibited about revealing their feelings, even if they can formulate them accurately. Some subjects may feel that an interview with a psychologist is an opportunity that should not be wasted on reports of their immediate responses to the experimental situation. In a variety of situations, introspective reports—even if taken immediately—may not be a trustworthy means of assessing the effectiveness of the independent variable.

A more difficult but far better technique for checking whether the experimental technique is producing the desired internal state is to run a number of pilot subjects in a separate experimental paradigm, in which the dependent variable is some behavior presumed to be a direct indication that the intended subjective state has been aroused. In other words, instead of manipulating one variable and measuring its effect on some other variable, we temporarily ignore the relationship between the independent variable and other variables and instead concentrate on the relationship between the treatment and some measure of the state or response that would logically seem to indicate the presence of our intended independent variable. Returning to the discussion with which we opened this chapter, our eventual prediction is that there is a relationship between a certain kind of internal stimulus and some kind of behavior. In the final experiment we will administer a treatment designed to elicit the internal state, assume that it has been produced, and measure to see if it in turn produces the predicted behavior. What we are suggesting as a means of checking on the appropriateness of the treatment is that we carry out a pilot experiment with the same manipulation and see if the internal state has been elicited. In other words, we perform a simple, one-stage experiment, using as our dependent variable some behavior that we believe is a direct indication of the internal state.

Consider again the experiment by Aronson and Carlsmith (1963), in which children were asked not to play with an attractive toy. In one condition the admonition was put in the form of a mild threat; in the other, in the form of a severe threat. Although intuitively the two threats seemed to differ along a dimension of severity, it would have been desirable to substantiate that intuition by providing independent evidence of this difference. One way of doing this would be simply to ask the child how severe the threat was, but the children were so young that they probably would not have given useful answers to such an abstract question. A better technique would be to run other subjects in a pilot test in which the forbidden toy was made more desirable, so that a number of children would in fact disobey that admonition. With the toys actually used in the Aronson-Carlsmith experiment, even the mild threat was severe enough so that no child went against it. If the toys were made so desirable that some children played with the forbidden toy in spite of the threat, and if this disobedience was more common when the threat was mild than when it was severe, we would be confident that the severe threat was really more severe than the mild threat.

It should be obvious by now that pretesting does not consist of a set of formal procedures. The pretesting stage of an experiment provides an opportunity to become familiar with the situation and even to allow one's intuitions their say: about ways of testing the hypothesis that will "fit" the situation, about procedural changes that will make the experiment coherent and meaningful, about the kinds of variables that are likely to influence the subjects in the particular subculture from which they are to be drawn. For example, if one wants subjects to engage in some counterattitudinal behavior, pretesting provides an opportunity to find out what the subject's attitudes are and therefore what is counterattitudinal. If one wants to expose nursery school children to

mild and severe threats, the pretest stage is the time to get to know the children and find out what kinds of things are threatening to them. It is a time to observe the experiment from the subject's point of view, to develop an intuition of what the situation "feels like," and to tinker with it until it "feels right." Part of the knack of being a good experimenter is an ability to seek out relevant information in the pretest stage and an openness to hunches and intuitions gained in the process of watching one's own experiment develop.

The Advantages of a Live Experimenter

In an attempt to avoid bias and gain a kind of control over the stimuli presented to the subjects, many experimenters have turned to the use of "canned" operations in the form of tape recordings, printed instructions, and computer-presented instructions. There are many situations in which the use of these techniques is justified and even essential. There are situations in which such methods can have a great deal of impact, such as the Darley-Latané (1968) experiment, in which the confederate's "nervous seizure" was really a tape recording. However, we should not lose sight of the fact that a live experimenter is not simply a bias-producing machine but instead frequently is a necessary ingredient in the experimental process. In this chapter and elsewhere (see Chapter 4) we have already discussed the important role an experimenter can play in making sure all subjects understand the instructions. In addition, the live experimenter can often succeed in "selling" a cover story to a degree that cannot be matched by canned instructions. Many times, when we read about the experimental procedure of a deception experiment as reported in a journal article, we are struck by the simplicity and transparency of the subterfuge and are amazed that it was successful. But the success of a cover story in disguising the true purpose of an experiment cannot be judged solely by looking at the words spoken by the experimenter. The manner of delivery makes a crucial difference. An explanation is often made more plausible by the physical presence of the experimenter who, through earnest demeanor and maintenance of eye contact, can frequently succeed in convincing the subject that the "experimental problem" described in the cover story is not only a legitimate object of scientific inquiry but is even an interesting and exciting area of investigation.

Moreover, the live experimenter can detect not only the fact that the subject is not understanding a set of instructions or is becoming inattentive but also that a subject is beginning to appear incredulous. Thus, the experimenter may repeat, answer questions, and deviate slightly from the prepared script in order to allay any doubts that may occasionally creep into the mind of a subject. Just as some subjects are, by nature or experience, brighter or more wide-awake than others, some subjects are more suspicious than others. For this reason, it would be absurd for the experimenter to stick blindly to the prepared script when a slight change in wording or emphasis might allay the subject's suspicions. Again, what we are advocating is not the abandonment of control but rather the attainment of a richer kind of control through an attempt to have all subjects in approximately the same state of mind when the independent variable is introduced. Let us reemphasize the fact that the use of a live experimenter often raises the possibility of bias in an experiment. We are well aware of this, and we believe that an experimenter should keep a careful record of any deviations from the script and should start all over again with a better manipulation if these deviations (especially those designed to allay suspicion) become frequent. But there are many ways of avoiding bias that do not sacrifice the advantages of the "personal touch" of the human experimenter who, in our opinion, should not become another victim of automation. (The problem of bias and techniques for avoiding it are discussed in Chapter 9.)

Parenthetically, it should be pointed out that "facing up" to a subject has other advantages as well. Most important of all, it lets the experimenter see what is going on. By shutting the subject up in a room with a tape recorder, a booklet, or a microcomputer, the experimenter too is shut off from information that could be of the utmost importance in interpreting the outcome of the experiment. For example, in attempting to convey a set of instructions to a subject in a face-to-face interaction, the experimenter may come to realize that these instructions are not viable; there is nothing quite like a yawn in the face to convince the experimenter that the instructions are dull and unmotivating. If the instructions are presented on a tape recording, the experimenter might never see the yawn and might run the whole experiment without realizing that the subjects are totally indifferent to the treatment. Similarly, if deception is used, there is nothing like the skeptical yet pitying look of an incredulous subject to convince us that we had best go back to the drawing board. As mentioned above, a talented experimenter can occasionally "sell" a rather incredible cover story. At the same time, there are some experimental situations which are inherently so transparent that a very sophisticated and elaborate cover story is required. In dealing with our own research assistants, we have found that we can waste a great deal of time and energy trying to convince a novice experimenter that his or her cover story is inadequate. A pilot trial on one subject is far more convincing.

Thus, the live, two-way exchange between subject and experimenter is an important learning experience. Few things are as unnerving (and therefore as educational) as being stuck for an hour with a subject who doesn't believe a word you're saying and couldn't care less. One way of avoiding such experiences is to build better and more convincing experiments. A good social psychology experiment never bores a subject, unless boredom is the conceptual variable.

Who's Running This Experiment, Anyway?

We have placed great emphasis on the experimenter's ability to interact with the subject. Our suggestion is that in the course of presenting the instructions to the subject, the experimenter should be certain that the intended information is being communicated and that the subject understands the instructions. This inevitably involves asking whether the subject understands or has any questions. There is a danger inherent in this strategy. If given the opportunity, some subjects attempt to wrest control of the session away from an unwary experimenter. When the experimenter asks if they understood, such subjects may take the floor and begin to ask questions which have nothing to do with the experiment, discuss previous experiments they have been in, and ask questions pertinent to aspects of the procedure which are yet to come. If this occurs, the experimenter stands in danger of either invalidating the experiment by engaging in long and friendly conversation with the subjects or offending and angering them by cutting them off too short. Thus, although we advocate a flexible procedure, this flexibility should operate within a limited range. In most experiments wide variation in the experimenter-subject rapport will increase the error due to variability among subjects' perceptions of the experiment; at worst, it could conceivably interact with experimental treatments in a manner which would make significant results meaningless, due to systematic error.

A good general rule in such a situation is for the experimenter to answer only those questions that clarify aspects of the procedure already covered but never to indulge the subject's ramblings and never to allow a subject to reorganize the sequence of instructions merely by asking a question pertaining to material that would have been covered a few minutes further along in the presentation. While setting the stage (see Chapter 6), the experimenter often has a chance to forestall future inappropriate comments from the subject; the best way to do this is by means of a short prelude to the main instructions, in which the experimenter explains the importance of control and uniformity in experimental situations and requests the subject to ask questions only for clarification. The experimenter should also state willingness to chat with the subject after the experiment, when these constraints will no longer apply. Most subjects can understand this and are not offended. In spite of this preface, occasionally a subject will deviate. If this occurs, the best way to field irrelevant chatter or anticipatory questions in the course of the experiment is to repeat, politely but firmly, that it is necessary to achieve a high degree of experimental uniformity and that, consequently (1) it would be preferable to shelve this discussion until after the experiment, or (2) the question raised will be answered in a few moments.