

*Why do subjects and researchers disagree  
about the how and why of subjects' behavior?*

## *Face Validity from Multiple Perspectives*

*Louise H. Kidder*

As a rule, we do not ask subjects to draw their own conclusions about the effects of a treatment, the meaning of an action, or the cause of a behavior. Instead, we as researchers provide feedback to subjects and inform them about the causes or meanings of their actions. There are instances, however, in which subjects draw their own conclusions, and these reveal some interesting disagreements between researchers and their subjects. I will examine the sources of disagreement and discuss why research can seem to be valid from one perspective and invalid from another. I am concerned with disagreements not only between researchers and their subjects but also between researchers and the groups from which their subjects are drawn and to which the results will be generalized. Research results always require interpretation, and different constituencies are likely to make different interpretations.

The simplest assessment of validity, which uses no statistical techniques, is called *face validity*. When applied as a minimal criterion to research measures, it is assessed by experts who can look at a measuring instrument

I wish to thank Richard D. Ashmore, Howard S. Becker, Ronnie Janoff-Bulman, David Kipnis, Joseph E. McGrath, Ralph Rosnow, and Carolyn Wood Sherif for helpful comments on earlier drafts of this chapter.

and say whether it seems like an appropriate measure on its face. If we apply this same minimal criterion to research conclusions, we encounter a new problem—choosing the experts. There are numerous groups of potential experts who can assess the face validity of research results and conclusions, and they are likely to disagree, especially about research that has implications for social policy or action. The potential for disagreement is clearest in applied or evaluation research. Program funders, administrators, recipients, and onlookers often have different criteria for what constitutes success and for identifying the relevant outcomes. Even with basic research, however, there are different parties to the research process, notably the researcher and the subjects, to whom the results are generalized, and they may disagree about the causes, effects, or interpretations of events (Becker, 1967).

Applied to measurement, face validity is a simple criterion, and as a result, it tends to be treated lightly, it is seldom discussed, and it is often dismissed as superficial. Applied to research conclusions, face validity is a more complex issue, resembling construct validity, although it lacks a statistical solution. For instance, two parties who disagree about the “cause” of women’s alleged “fear of success” in research like Matina Horner’s (Horner, 1968) early studies are disagreeing about what the phenomenon should be called. One may call it an intrapsychic fear or motive to avoid success, while the other may call it a realistic appraisal of what happens when sex-role stereotypes are violated. The former term points to an intrapsychic cause; the latter, to a cultural pattern. Therefore, when different parties assess research conclusions and judge them valid or invalid on their face, they can disagree about the proper name of either the cause or the effect, and in so doing they are assessing the construct validity of a process.

To examine how and why different parties can fail to agree about the face validity of research conclusions, I have identified three points of disagreement. The first point of disagreement is about the causes of behavior. Where we look for causes in the complex set of causal chains depends upon our point of view, and researchers and subjects can differ predictably in where they look for causes. The second point of disagreement is about effects. Which of the many possible effects of a given cause that we select will determine whether we conclude that a given cause produced positive, negative, or null results? The third point of disagreement is about the meanings of events, and this involves simultaneous disputes about causes and effects.

### **Disagreements About Causes**

*Researchers and Subjects as Observers and Actors.* Some disagreements about causes are predictable. Actors and observers have different perspectives on social interactions, and they differ predictably in where they locate the causes

of behavior (Jones and Nisbett, 1972). Research subjects and investigators generally occupy the roles of actors and observers, respectively. Like actors, subjects have access to more information about themselves than the researchers do. Their life histories form the background of their current actions, and changes in environmental conditions stand out in relief against that background. Thus, they attend to the environment, locate causes for their actions outside themselves, and make situational attributions. For researchers, their subjects stand out in relief against a background of situational forces. Consequently, they make causal attributions that locate the sources of behavior in the persons.

In Milgram's (1963) studies of obedience, the subjects and the researchers differed predictably in their attribution of causality and responsibility. The subjects indicated during the debriefing that they held the experimenter responsible for the shocks that they had given. During the experiment, subjects said such things as "You're going to keep giving him, what, 450 volts every time?" They pointed to the experimenter as the cause of the shocks. One subject, worried about the learner's health, repeatedly asked the experimenter, "You accept all responsibility?" From the experimenter's point of view, the subjects were responsible for obeying orders and for inflicting harm on the learner. From the subjects' point of view, the experimenter was responsible, because the experimenter insisted that the subjects inflict harm. In the negotiations that occurred during the debriefing, neither party questioned the joint role of the commands from the experimenter and the obedience from the subjects, but they disagreed over the relative emphasis. Subjects held the experimenter more responsible than themselves and vice versa. Both exhibited a self-serving bias in their attributions (Miller and Ross, 1975), and they made the predictable attributions of actors and observers. As actors, the subjects located the cause of their behavior in the situation, that is, in the experimenter's insistence, and as observers, the experimenters located the cause of the subjects' behavior in them as persons, that is, in human nature. Milgram concludes his film of the obedience studies with the note that "human nature cannot be counted upon" to insulate us from the powers of government to command its citizens to commit acts of violence.\*

The subjects in Milgram's experiments disagreed with the researcher not only because they occupied different positions but also because they were deceived on two counts. The subjects were led to believe that the learner was receiving a shock when the subject pressed the lever when in fact no shock was delivered, and they were told that, while the shocks could be painful, they were not dangerous. The filmed account of the experiments shows that sub-

\**Obedience*, copyright 1965 by Stanley Milgram, distributed by the New York University Film Library. Extracts quoted with permission.

jects believed the first allegation — that the learner was receiving a shock — but that they were uncertain about the second. They had conflicting information about the dangers involved. The experimenter said that the shocks could be painful but that they were not dangerous, while the words printed on the shock apparatus said, "Danger . . . Severe Shock . . . XXXX . . . 450 Volts." The subjects' own eyes and ears told them that the shocks were both painful and dangerous, but the experimenter said that there was no danger. Caught between conflicting messages, the subjects negotiated with the experimenter and held him responsible for what happened. They acknowledged that their fingers were on the shock levers and thus that they were a proximal cause of the learner's discomfort, but they located the ultimate cause and therefore the final responsibility with the experimenter (Harvey and Rule, 1978; Kidder and Cohn, 1979).

The causal attributions that experimenters make are not necessarily trait attributions. Most social psychologists, in fact, are probably not trait theorists, for they examine the effects of environmental conditions. Milgram manipulated several environmental conditions and found that subjects disobeyed an authority's commands when they had confederates who also disobeyed and that they administered fewer shocks when they had to force the learner's hand onto the shock plate (Milgram, 1963).

There is often a difference, however, between the external causes that researchers identify and the external causes that subjects identify. For instance, in cognitive dissonance research (Festinger and Carlsmith, 1959), subjects who participated in a dull experiment and then received a small sum of money for telling someone else that the task was really interesting subsequently rated the task as more interesting than subjects who told the same lie for a large sum. The researchers concluded that external circumstances were the cause of the subjects' ratings: The \$1 payment was external justification for telling a lie, so subjects sought internal justification and called the task truly interesting. The subjects also concluded that something outside themselves made them call the dull task interesting, but they did not identify the small payment as the cause. They pointed to the stimulus properties of the task; the subjects found redeeming features in an otherwise dull task. Dissonance theorists point to the small payment and the subsequent dissonance arousal as the cause of the subjects' reports. Subjects point to features of the stimulus to explain their reports (Jones and Nisbett, 1972).

Nisbett and Wilson (1977) report a series of studies in which the subjects and the researchers identified different causes for the subjects' behaviors. For instance, researchers asked subjects to select one of four nightgowns and one of four identical pairs of nylon stockings. When subjects were asked why they had selected a particular gown or pair of stockings, they referred to qualities of the gowns or the stockings. The researchers, having alternated the positions

of the items, demonstrated that the subjects chose not according to appearance but according to position—the further that an item was to the right, the more frequently that item was selected. The subjects attributed their behavior to properties of the stimulus—that is, to quality or appearance—and the researchers attributed it to external circumstances—that is, to position. No subject noticed a position effect. The researchers concluded that the subjects were wrong and that their verbal reports could not be trusted. I reach a different conclusion. The subjects were fooled into thinking that they could discriminate among four indiscriminably attractive gowns or four identical pairs of stockings. Thus, it was the situation, rather than their verbal reports, that could not be trusted. That rigged situation resembles real-life situations in which consumers are given a choice among essentially identical products, but it is not representative of most choice points in people's lives, where they select stimuli rather than positions. The subjects, therefore, were wrong about the cause of their selection because the situation itself was rigged and unrepresentative.

In the preceding examples, the researchers and their subjects disagree about the causes of the subjects' behavior for two reasons. First, researchers have access to more information and to different information when they deceive their subjects. As a result, they attribute subjects' behavior to causes that the subjects are not always aware of, such as the insufficient justification of a small reward or the position of a product. Second, even in situations where there is no deception, subjects and researchers have access to different information because they occupy different positions as actors and observers. Deception limits the face validity of research because it creates a rigged situation, but even when subjects are not deceived, they are likely to differ from researchers in where they locate causes.

*Researchers and Subjects with Different Theories and Ideologies.* Some disagreements between researchers and the people whom they study resemble disagreements among researchers. They can hold different ideological and theoretical beliefs. Although we often design research to study the effects of one or two variables, we acknowledge that a complete understanding of the origins of any effect requires that we examine multiple causal chains (Brickman, Ryan, and Wartman, 1975; Kidder and Cohn, 1979). By selecting one of many possible independent variables, we place our emphasis at one point on one of many causal chains. These different causal chains represent complementary rather than competing explanations. For instance, if we study the effect of daytime versus nighttime police patrols on the incidence of crime, we have selected only one of many possible causes of the crime rate. Other causes include the cohesiveness of neighborhoods, levels of unemployment, use of personal and home protection devices, and so on. Each of these factors occupies a position in a separate causal chain.

Although these alternative chains represent complementary rather than competing explanations, they imply different solutions that compete in the political arena for money, time, and effort. For instance, if we study the effects of police patrols on the crime rate and we find that police patrols have an effect, we imply that the appropriate solution is to introduce more police patrols. However, if we study the influence of neighborhood organization or youth employment on the crime rate and we find an effect, we imply that the appropriate solution is to provide more youth employment programs or to promote neighborhood organization (Kidder and Cohn, 1979).

Several black psychologists have made similar points about the white researcher in black society. There are ideological and political differences between researchers who give person-centered explanations of race differences and researchers who give situation-centered explanations of race differences (Clark, 1973; Gordon, 1973; Nobles, 1973). Both as psychologists and as subjects of research, these black psychologists object to the search for inner causes, whether they be genetic or psychological, because such research does not lead to social change. Locating the causes for social problems within individuals permits us to blame victims rather than social conditions. Gordon (1973, p. 94) calls instead for a black psychology that combines analysis and action.

Similarly, feminists call for replacing genetic or intrapsychic explanations of sex differences with sociocultural explanations (Sherif, 1979). In the late 1960s, Horner (1968) postulated a motive to avoid success in women to explain why women did not aspire or achieve as men did. Since then, many women have said that Horner misidentified the cause (Parlee, 1980; Robbins and Robbins, 1973; Sherif, 1979; Unger, 1979; Wallston, 1979). They disagree with Horner's intrapsychic location of the cause in part because it does not fit their own experience (Parlee, 1980) and in part because it is more useful to view intrapsychic variables as symptoms or consequences than as causes. Horner's intrapsychic explanation also creates no force for social change. As Robbins and Robbins (1973, p. 137) have put it, "As long as it is believed that the lack of success of many professional women is primarily due to their psychological disability, expressed as a 'motive to avoid success,' there will be little incentive to redress some of the tangible external barriers—such as admission quotas, slower rates of promotion, and reluctance to grant tenure—which have stood just as surely in the way of professional advancement."

The absence of face validity imputed to Horner's findings led other researchers to improve on Horner's design. Her study confounded the gender of subjects and stimulus persons. Women responded to a female stimulus cue, Ann, who was first in her class in medical school. Men responded to a male stimulus cue, John, who was also first in his class. When women told stories predicting that Ann would subsequently fail or lead an unhappy life alone or

commit suicide, Horner concluded that they were projecting their own motives onto Ann and she deduced from this that women have a motive to avoid success. However, another explanation is that Ann's out-of-role behavior, not women's motives, caused them to tell sad stories about Ann. Subsequent researchers disentangled the gender of subjects and stimulus persons and demonstrated that both men and women fear for someone who succeeds in a role traditionally reserved for the opposite sex (Monahan, Kuhn, and Shaver, 1974). The failure of Horner's original interpretation was a failure of face validity. It did not ring true to some women and men, who subsequently developed sociocultural rather than intrapsychic explanations (Sherif, 1979).

The critics of intrapsychic explanations for apparent race and sex differences have called for causal analyses that would lead to social action. Intrapsychic explanations of self-hatred blame black people, and intrapsychic explanations of women's career aspirations blame women. Both deflect attention from social change to personal change. Critics of intrapsychic explanations do not deny that social and cultural conditions can have intrapsychic correlates, but they regard those correlates as consequences, not as causes.

In summary, when the subjects of research disagree with researchers about causes, they may do so for two reasons: First, they occupy different positions, as actors rather than observers, and they have access to different information. Second, they disagree about which of the many contributing causes are the important ones, whose discovery suggests solutions to problems.

### **Disagreements About Effects**

Research conclusions can lack face validity from the perspective of either the subjects or the population to which findings will be generalized if the reported effects of a treatment seem implausible or irrelevant. These differences of opinion can arise from several different sources. First, if the researcher examines manifest functions or public performances, while the subjects focus instead on latent functions or private performances, or vice versa, researcher and subjects are likely to reach different conclusions. Second, as with causes, subjects and researchers have access to different information about effects. The differences arise not from their positions as actors and observers but because they gather different types of information at different times. Finally, subjects and researchers can disagree about the relevance or importance of various effects and use different criteria for what constitutes a successful treatment.

*Latent Versus Manifest Functions, Private Versus Public Actions.* Reading about gender differences in reward allocations, I found myself objecting to the repeated finding that men allocate rewards in proportion to contribution, while women allocate rewards equally (Leventhal and Lane, 1970). Men who distributed rewards between themselves and a less-deserving part-

ner took more for themselves, following the rule of equity, and women who distributed rewards gave themselves and their partners identical amounts, following the rule of equality. These sex differences were reported repeatedly, with adults, with children, with real partners, and with fictional others. In spite of the repeated evidence, the effect of subjects' gender was not plausible to me because it did not fit either with my personal experience or with my observations as a teacher.

I had students work on group projects and asked them how they wanted to distribute the group grade. I explained that if I gave a grade of B to a group project, the members could divide it equally, so that everyone received a B, or they could divide it proportionally, so that some received A's, some received B's, and some received C's, provided that the group average remained B. The groups always agreed that the grade should be divided equally. However, individual students sometimes came to me in private to protest. They felt that they had contributed more to the project than other students and believed that they deserved an A, while others deserved a C. In private, they proposed an equitable, not an equal, distribution. The students who came to me were always women. They were not, therefore, conforming to the research findings that women prefer equality, while men prefer equity, at least not when they expressed their opinions privately. When I subsequently asked students to let me know in writing, anonymously, but with their gender indicated, how they preferred to allocate awards, I found women choosing equity. That is, they reported having done more of the work, so therefore they deserved more of the reward (Kidder, Belletirrie, and Cohn, 1977).

I then designed an experiment to test the effect of anonymity on reward allocations by men and women to themselves and to a less-deserving partner. In one condition, the subjects believed that their allocations would be public, and in the other condition, they were assured that their decisions would remain private. In the public condition, men chose equity and women chose equality. In private, each did the opposite. The public behavior conformed to the findings of previous literature, but the private choices did not. The apparent sex effects in the earlier literature disappeared in private and were even reversed. The sex effects reported earlier had no face validity for me because they contradicted my personal experience and my observations of the choices that men and women made in private. By making the public-versus-private condition a variable, I uncovered the out-of-role behavior that I had observed informally, and I found different sex effects.

Research that uncovers effects can have face validity for some people but not for others. Nancy Henley's research on the functions of touch uncovers some latent functions which contradict the cover story that touching is a sign of affection (Henley, 1977; Parlee, 1980). Henley's work identifies instances "when the act of touching has a different meaning: When touching is asym-

metrical . . . the act of touching represents a statement or affirmation of power differences" (Parlee, 1980, pp. 1-2). By observing the conditions under which one person touches another and that touch is not reciprocated, Henley noted that touching can serve to bestow low status rather than to convey affection. These two effects of touch—bestowing low status and conveying affection—occur in different contexts, and they have different causes. Identification of effects, therefore, is not always separable from identification of causes, particularly when the two involve people's intentions. Disagreements over such effects are disagreements about the meanings of events, and as such they are questions about construct validity.

Just as causal attributions sometimes contain a self-serving bias, interpretations of effects can serve the purposes of actors or observers (Miller and Ross, 1975). Some people may choose to believe that asymmetrical touching affirms affection rather than power. There is no research on how actors and observers or how doers and those who are "done unto" perceive effects. Consciousness-raising movements among low-status groups (black people, women, the physically disabled) suggest that low-status persons do not automatically perceive asymmetrical acts as affirmations of power differences but that they learn to assign such meaning to asymmetrical touch or to other behavior.

*Availability of Information About Effects.* The timing of the research process makes it almost inevitable that researchers will have more of one kind of information about the effects of a treatment than will their subjects. Individual subjects can assess the effects of a treatment upon themselves immediately. The researcher can assess the effects upon a large group of subjects, but only after all have participated. Even in research that provides for full and accurate debriefing, subjects cannot know what the overall effects of a treatment are until long after they have participated. Thus, if a researcher asks subjects to assess the effects of a treatment, they cannot make a fully informed judgment, because all the data are not available.

This inevitable difference between what subjects know when they participate and what a researcher knows at the end of a study accounts for some reported discrepancies between researchers' and subjects' descriptions of effects. The most dramatic discrepancy appeared in the evaluation of a delinquency prevention program, known as the Cambridge-Somerville experiment, begun in 1939 and evaluated thirty years later (McCord, 1978). Boys between the ages of five and thirteen were randomly assigned to a delinquency prevention treatment group or to a control group. Boys in the treatment group received scholastic tutoring, medical and psychiatric attention, opportunities to attend summer camp, access to youth activities, and biweekly visits from social workers. Boys in the control group received no special services. The treatment lasted five years.

In a thirty-year follow-up of men who had been in the treatment and

control groups, McCord found that men in the treatment group fared worse on every measure where there was a significant difference. Equal numbers of men in the treatment and control groups had been convicted of committing a crime, but of those men, a higher proportion of men in the treatment group had committed more than one crime. Men in the treatment group showed a higher incidence of self-reported alcoholism, ulcers, asthma, and high blood pressure. Equal numbers of men in the treatment and control groups had died by the time of the follow-up, but men in the treatment group had died at an earlier age. Of those who were living, men in the treatment group had lower occupational status on the average than men in the control group. The treatment, therefore, had uniformly negative effects.

There was one exception to this otherwise negative picture. When asked what they thought about the program, two-thirds of the men in the treatment group said that it had been helpful. They said such things as "It helped prepare me for manhood," it gave "better insight on life in general," and it helped them to have "faith and trust in other people" (McCord, 1978, p. 287). According to the subjects' assessment, the treatment helped. By every objective standard, it hurt. Why do they disagree?

The researcher had a better vantage point from which to assess the multiple effects of the treatment and access to longitudinal and cross-sectional archival data. If the subjects had been informed about the higher criminal arrest rate, the alcoholism rates, the lower occupational standing, and the earlier deaths of men in the treatment group, they, too, would probably have concluded that the program had hurt more than it had helped. For this reason, rather than ask who is right, I will examine what the subjects could have meant when they said that the program had helped. Their positive assessment resembles the almost universal finding of evaluations of compensatory education programs: "Regardless of the type of program, duration, or actual results, parents are enthusiastic" (McDill, McDill, and Sprehe, 1969, pp. 43-44).

It is possible that men in the Cambridge-Somerville experiment and parents of children in compensatory education programs are enthusiastic about the fact that they have been offered services that they normally would not receive. They may be responding more to the process than to the outcome (Kidder, 1981a). Perhaps it is the process that they appreciate. Men in the Cambridge-Somerville experiment would not have had access to the summer programs, counseling, or youth activities if they had not been in the treatment group. Without compensatory education programs, participating parents would not have access to organized preschool care for their children. Thus, they may be enthusiastic about the programs not because the programs produce the intended effects but because they provide goods and services that participants would not have had otherwise.

We can imagine similarly discrepant evaluations by researchers and participants in other social programs, such as the negative income tax experi-

ments. Researchers have studied the effect of the program on the number of hours worked. The program is considered a success if recipients of negative income tax benefits do not work less than they have before. I have found no reports of what the recipients had to say about the programs, but I can extrapolate from discrepancies found in evaluations of various negative income tax experiments. The program appears to have different effects on salaried workers and on nonsalaried self-employed farmers (U.S. Department of Health, Education and Welfare, 1978).

Salaried workers are more likely to reduce the number of hours worked than nonsalaried farmers are, particularly in those conditions of the experiment where participants lose their negative income tax benefits at a high rate as they earn increased amounts of money. In some programs, recipients lost as much as 70 percent of their benefits when their salaries rose above a minimum level. In effect, they were taxed on their benefits as if they were in a 70 percent tax bracket. Therefore, they had little incentive to increase the hours worked, because they rapidly lost their benefits. For nonsalaried self-employed farmers, however, the extra hours worked did not produce an immediate increase in their personal earnings. By investing time and labor on their farms, they could eventually realize higher earnings, but these would be delayed, and the workers suffered no loss in their immediate negative income tax benefits. Therefore, self-employed farmers increased the number of hours they worked, and salaried workers decreased the number of hours worked. The same treatment had different effects.

By one criterion, the number of hours worked, the program succeeded among farmers and failed among salaried workers. It is possible, however, that by another criterion, the program succeeded among both groups. What might that other criterion be? If we examine the behaviors of both groups, it appears that both groups tried to maximize the financial returns on their work. For the salaried workers, there was a point of diminishing returns. At that point, they decreased their hours worked, possibly in exchange for more leisure time or for unreported work. For the nonsalaried farmers, there was no immediate point of diminishing returns as long as they could defer their earnings. The goal for both groups appeared to be maximizing returns, not the number of hours worked. Asked to evaluate the negative income tax experiments, both groups could have called it successful by their own criteria. Thus, it is possible for researchers and their subjects to have access to different information, adopt different criteria for success, and attach different meanings to the resulting "effects."

### **Disagreements About Meanings of Events**

The face validity of conclusions rests on the construct validity of measures of both causes and effects. If researchers and subjects disagree about the

effects of a treatment like the negative income tax, they are disagreeing about what working means. For a salaried employee, working more means receiving fewer benefits and realizing a small return. For a nonsalaried farmer, working more means receiving the same benefits and realizing a delayed return on the work. For a researcher, working more is better than working less, and presumably it means the same in all instances if that is the criterion used to evaluate the negative income tax experiments. Therefore, researchers and subjects could disagree about the success of the taxation, because they disagree about whether working more is good or bad.

This example illustrates another point: When people disagree about the effects of an experiment, they may also be disagreeing about the causes. For instance, one interpretation of the failure of negative income tax experiments is that the payments produced a disincentive for work. A different interpretation is that the high taxation made it less worthwhile to work. Was it the payment of benefits or the loss of those benefits that made some people work less? When recipients earned money and their income rose above a baseline, they lost their negative income payments at a specified taxation rate. If the rate was 70 percent, they received only 30¢ for every dollar earned above the baseline income level, because their benefits were reduced at the rate of 70¢ for every dollar earned above the baseline. Therefore, in identifying the effects of the experiment, we are also identifying the cause, and in the present instance, the cause could be either payment of benefits or loss of benefits.

The subjects in Milgram's obedience studies who disagreed with Milgram about the meanings of their actions disagreed simultaneously about both causes and effects. Milgram's data have generally been interpreted as showing that subjects obeyed rather than defied the commands. The subjects, however—even those who pressed the maximum-shock lever—described themselves as defying rather than obeying authority. The filmed record of the experiments contains exchanges like the following, which occurred in a debriefing session:

*Milgram:* The purpose of the experiment was to see how you would react to Mr. Williams' [the experimenter's] authority — whether you would obey it or defy it.

*Subject:* I defied it! I mean, who the hell is he?

The subject believed he had defied authority because he had not willingly complied but had been coerced. Milgram, however, portrayed this subject as obedient. Another debriefing session contains the following exchange:

*Subject:* He [the experimenter] kept throwing him [the learner] 450 volts.

*Milgram:* Who actually pressed the switch?

*Subject:* I was, but he kept insisting.

*Milgram:* Why didn't you stop?

*Subject:* I *did* stop, but he kept telling me to go on.

In such cases, Milgram concluded that the subjects obeyed authority and did not stop, and the subjects concluded that they had defied authority or had stopped.

Rather than trying to determine who is right, I have chosen to examine why the subjects and the researcher disagreed. In the exchanges presented here, they disagree about the effects of commands from authority in part because they disagree about the causes of the subjects' behavior. The subjects said that they defied authority and stopped even when they had pressed the maximum-shock lever. They considered themselves defiant, because at some point they switched responsibility for the shocks from themselves to the experimenter. In one sequence of the film, a subject and the experimenter negotiate responsibility as follows:

*Subject:* The guy's hollering in there.

*Experimenter:* Continue, please, go on.

*Subject:* The guy has a heart condition in there. You want me to go on? He can't stand it. What if something happens to him?

*Experimenter:* The experiment requires that you continue.

*Subject:* I refuse to take the responsibility. Who's going to take the responsibility if something happens to that gentleman?

*Experimenter:* I'm responsible. Continue, please.

*Subject:* [After a few more shocks] Oh, no. No, I'm not going to kill that man.

*Experimenter:* The experiment requires that you continue.

*Subject:* We can't go on if the learner doesn't answer.

*Experimenter:* If the learner doesn't answer, go on.

*Subject:* Can't you check in and see if he's alright, please? [to the learner:] Answer, please. Are you alright in there? [to the experimenter:] Something's happened to that man in there. You accept all responsibility?

*Experimenter:* The responsibility's mine, correct. Please go on.

*Subject:* You're going to keep giving him, what, 450 volts every time?

Later, during the debriefing, the subject said to Milgram, "He [the experimenter] kept throwing him 450 volts." The subject had negotiated the respon-

sibility with the experimenter and concluded that it was the experimenter, not himself, who inflicted the shocks on the learner. The subject's interpretation of both the causes of his behavior and the effects of commands from authority differed from Milgram's interpretation. The subject felt that he had stopped and that it was the experimenter who had continued; the subject felt that he had defied authority, not obeyed it.

### **Drawing Conclusions with Face Validity from Multiple Perspectives**

Having identified conclusions that lack face validity from the perspective of a given constituency, I felt obliged to say whether or how we can do research that appears valid from multiple points of view. The question partially provides its own answer. We can conduct research that has validity from more than one constituent's point of view by asking questions and eliciting information from multiple perspectives. For instance, to evaluate the effects of a social program like compensatory education, we can use multiple criteria derived from the points of view of parents, children, teachers, and program administrators. Some criteria may have zero or even negative correlations with one another, because they represent divergent rather than convergent perspectives. The more applied and the more politically sensitive that a research topic is, the greater the potential for disagreement about causes and effects, which makes it all the more urgent that we elicit multiple perspectives. Research on intergroup conflict provides a useful paradigm. The rival groups of Rattlers and Eagles in the Robbers' Cave experiments had different perspectives on who was to blame for the rivalry. Neither group had an exclusive hold on the truth; both holds were necessary to understand the sequence of events (Sherif, 1976; Sherif and others, 1961).

Hastorf and Cantril's (1954) analysis of Dartmouth and Princeton students' perspectives on a hotly contested football game shows that the rivals disagreed not only in their descriptions of how "rough and dirty" the game was but also in their count of how many rules were broken by each team. "The 'game' actually was many different games . . . each version of the events that transpired was just as 'real' to a particular person as other versions were to other people" (Hastorf and Cantril, 1954, p. 132). Therefore, one solution to the problem of maximizing face validity for multiple constituencies is to include more than one perspective in the inquiry. Using the intergroup conflict paradigm, we can take our inquiry a step further and ask how the divergent perspectives developed and how they relate to the circumstances of the relationship (Carolyn Sherif, personal communication).

A second answer to the question of how to do research that has face val-

idity from other people's perspectives is contained in Herbert Blumer's description of how to "catch the process of interpretation through which they construct their actions" (Blumer, 1962, p. 62). He says that "to catch the process, the student must take the role of the acting unit." The presume danger of taking the role of the acting unit is that the researcher may lose his or her objectivity. Blumer counters this by saying that the failure to take the subject's point of view can lead to errors: "To try to catch the interpretive process by remaining aloof as a so-called 'objective' observer and refusing to take the role of the acting unit is to risk the worst kind of subjectivism—the objective observer is likely to fill in the process of interpretation with his own surmises in place of catching the process as it occurs in the experience of the acting unit which uses it" (Blumer, 1962, p. 62). Misplaced objectivity, therefore, can be a source of disagreement between subjects and researchers, because it forces researchers to guess about subjects' interpretations. Participant observation provides a means for researchers to learn what the subjects' interpretations are and reduces the need for guessing.

Participant observation is often suspect because it seems overly subjective. I have shown elsewhere (Kidder, 1981a) that, although participant observation appears to be subjective, it has much in common with quasi-experimental research. Participant observers can make causal analyses and rule out threats to internal validity in much the same way as quasi experimenters. By gathering data over an extended time, a participant observer can develop a time series design, and by gathering longitudinal data on multiple groups, individuals, or institutions, a participant observer can develop multiple time series. The long series of observations enables the participant observer to test rival explanations and to rule out threats to internal validity like a quasi experimenter. The subjective quality of participant observation, therefore, does not preclude causal analysis.

As a rule, participant observers do not use the objective language of independent and dependent variables, but that objective language can disguise rather than eliminate bias. As Carolyn Sherif (1979) points out, intrapsychic explanations for sex differences have been perpetuated by research that appears to be "objective." She adds that, by "glorifying the experiment," we imply that "knowledge is to be gained by studying parts, elements, or variables," and as a result, "much of what goes on is simply ignored" (Sherif, 1979, p. 102). An experiment is no less an interpersonal event than participant observation is, and the artifacts arising from demand characteristics and other unintended features have been amply documented (Rosenthal and Rosnow, 1969). Participant observation is not a panacea, but it can increase the face validity of research, because it enables a researcher to catch the process from a subject's point of view.

- Kidder, L. H., Bellettirre, G., and Cohn, E. S. "Secret Ambitions and Public Performances: The Effects of Anonymity on Reward Allocations Made by Men and Women." *Journal of Experimental Social Psychology*, 1977, 13, 70-80.
- Kidder, L. H., and Cohn, E. S. "Public Views of Crime and Crime Prevention." In I. H. Frieze, D. Bar-Tal, and J. S. Carroll (Eds.), *New Approaches to Social Problems: Applications of Attribution Theory*. San Francisco: Jossey-Bass, 1979.
- Leventhal, G. S., and Lane, D. W. "Sex, Age, and Equity Behavior." *Journal of Personality and Social Psychology*, 1970, 15, 312-316.
- McCord, J. "A Thirty-Year Follow-up of Treatment Effects." *American Psychologist*, 1978, 33, 284-289.
- McDill, E. L., McDill, M. S., and Sprehe, J. T. *Strategies for Success in Compensatory Education: An Appraisal of Evaluation Research*. Baltimore, Md.: Johns Hopkins University Press, 1969.
- Milgram, S. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology*, 1963, 67, 371-378.
- Miller, D. T., and Ross, M. "Self-Serving Biases in the Attribution of Causality: Fact or Fiction?" *Psychological Bulletin*, 1975, 82, 213-225.
- Monahan, L., Kuhn, D., and Shaver, P. "Intrapsychic Versus Cultural Explanations of the 'Fear of Success' Motive." *Journal of Personality and Social Psychology*, 1974, 29 (1), 60-64.
- Nisbett, R. E., and Wilson, T. D. "Telling More Than We Know: Verbal Reports on Mental Processes." *Psychological Review*, 1977, 84 (3), 231-259.
- Nobles, W. W. "Psychological Research and the Black Self-Concept: A Critical Review." *Journal of Social Issues*, 1973, 29 (1), 11-32.
- Parlee, M. B. "Developing Feminist Psychology." Paper presented at the 88th annual meeting of the American Psychological Association, Montreal, 1980.
- Robbins, L., and Robbins, E. "Comment on 'Toward an Understanding of Achievement-Related Conflicts in Women'." *Journal of Social Issues*, 1973, 29 (1), 133-137.
- Rosenthal, R., and Rosnow, R. (Eds.). *Artifact in Behavioral Research*. New York: Academic Press, 1969.
- Sherif, C. W. *Orientation in Social Psychology*. New York: Harper & Row, 1976.
- Sherif, C. W. "Bias in Psychology." In J. A. Sherman and E. T. Beck (Eds.), *The Prism of Sex*. Madison: University of Wisconsin Press, 1979.
- Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., and Sherif, C. W. *Intergroup Conflict and Cooperation: The Robbers' Cave Experiment*. Normal, Okla.: Institute of Group Relations, 1961.
- Unger, R. K. *Female and Male: Psychological Perspectives*. New York: Harper & Row, 1979.
- U.S. Department of Health, Education, and Welfare. "The Rural Income Maintenance Experiment." In T. D. Cook and others (Eds.), *Evaluation Studies Review Annual*. Vol. 3. Beverly Hills, Calif.: Sage, 1978.
- Wallston, B. S. "What Are the Questions in Psychology of Women? A Feminist Approach to Research." Presidential address for Division 35, American Psychological Association, New York City, 1979.

Louise H. Kidder is associate professor in the Department of Psychology at Temple University.