SELF-PERCEPTION THEORY[1]

Daryl J. Bem

STANFORD UNIVERSITY
STANFORD, CALIFORNIA

[1] Development of self-perception theory was supported primarily by a grant from the National Science Foundation (GS 1452) awarded to the author during his tenure at Carnegie-Mellon University.

I. Introduction

Individuals come to "know" their own attitudes, emotions, and other internal states partially by inferring them from observations of their own overt behavior and/or the circumstances in which this behavior occurs. Thus, to the extent that internal cues are weak, ambiguous, or uninterpretable, the individual is functionally in the same position as an outside observer, an observer who must necessarily rely upon those same external cues to infer the individual's inner states.

These two propositions constitute the heart of the author's self-perception theory and, accordingly, the central topic of this review. This article will trace the conceptual antecedents and empirical consequences of these propositions, attempt to place the theory in a slightly enlarged frame of reference, and, hopefully, clarify just what phenomena the theory can and cannot account for in the rapidly growing experimental literature of self-attribution phenomena.

Self-perception theory was initially formulated, in part, to address empirically certain questions in the "philosophy of mind" (e.g., Chappell, 1962; Ryle, 1949). When an individual asserts, "I am hungry," how does he know? Is it an observation? An inference? Direct knowledge? Can he be in error, or is that impossible by definition? How does the evidential basis for such a first-person statement (or self-attribution) differ from the evidential basis for the third-person attribution, "He is hungry"?

Such questions have traditionally been subjected to purely analytic rather than empirical analyses, and psychologists have generally been willing to leave such exercises to the philosophers. In earlier times, when debates about introspection were in vogue, psychologists did discuss such matters, but those discussions, too, were primarily philosophical rather than empirical in tone. In fact, it appears that the only discussion in the literature which treats such questions as substantive psychological problems is B. F. Skinner's "radical-behavioral" analysis of "private events" and their role in a science of human behavior (Skinner, 1945, 1953, 1957). It was Skinner's analysis which inspired the present "self-perception theory"; accordingly, it is with Skinner's analysis that this review begins.

A. The Ontogeny of Self-Attributions

In order to identify and label things in his environment, a child must initially have someone else around who will play the "original word game" of pointing and naming, someone who will teach the child to distinguish between objects and events that appear similar and to label

them with different descriptors. The skills of self-description would appear to emerge from the same procedure, both with respect to overt behavior ("I seem to be eating more today") and to the effects of stimuli on the individual ("It gives me goose pimples"). The problem, of course, arises when the stimuli or events to be described are "private" internal states to which nobody but the individual himself has access, for it then becomes difficult for the verbal community to make differential reinforcement of the appropriate descriptive response directly contingent upon the presence or absence of the stimuli which are to be labeled. What, then, can be done?

As Skinner has noted, there are a few cases when an appropriate descriptor can still be acquired without explicit training. For example, some private stimuli are generated from covert behavior which was at one time overt or which accompanied parallel overt behavior to which descriptors could be attached. Thinking sometimes has this property: "I said to myself said I . . ." A second resource which is sometimes available is metaphor or stimulus generalization. An individual, for example, can easily learn to identify "butterflies in the stomach," and at least one child has generated his own metaphor by describing a foot which had fallen asleep as "feeling like gingerale when I hold the glass to my face." Some descriptors of emotional states (e.g., "feeling blue" or "down in the mouth") may have similar historical origins. But these special cases cover a very limited domain of self-descriptive statements, and most of the time a child must still be explicitly taught how to describe his internal states in the same way he is taught to describe his outer environment: someone must be able to "point and name." In training a child to describe pain, for example, an observer, at some point, must teach him the correct response at the critical time when the appropriate private stimuli are impinging upon him. But the observer himself must necessarily identify the "critical time" on the basis of observable stimuli or responses and implicitly assume that the private stimuli are, in fact, accompanying these public events. The description "it hurts" may thus be established in a child's repertoire by saying: "Don't cry; I know it hurts to bump your head." The descriptor itself can then generalize to a large class of private "painful" stimuli even though it was originally an observable response (crying) and a public stimulus variable (bumping the head) which set the occasion for the observer's inference that the child was experiencing pain.

If the resources available to the community for setting up descriptions of private events seem inadequate—and it is proposed that this list of resources is exhaustive—it should be recalled that the result is also often inadequate. Not only does the community remain ignorant

of whether the complaint of a headache is valid, for example, but the same deficiencies which generate public mistrust lead the individual himself to faulty self-knowledge. "There appears to be no way in which the individual may sharpen the reference of his own verbal repertoire in this respect [Skinner, 1953, p. 261]."

Thus, far from having direct and unerring knowledge of our internal states, Skinner's analysis implies that we have virtually no knowledge at all until we have been explicitly trained. Internal identifications that we have not been taught remain internal identifications that we cannot make. In everyday life we are spared from confronting our incompetence in this regard only because people know better than to call upon us to make internal discriminations which we are typically not taught ("Is it your spleen or your liver which is causing you the discomfort, Mr. Jones?"). It should be noted that Skinner's analysis thus reverses the usual practice in psychology of assuming the existence of awareness *a priori* and then treating any residual unawareness as problematic. For Skinner, awareness, not unawareness, is the phenomenon which normally requires analysis, and the uniqueness of his conceptual contribution resides precisely in his explicit recognition of this fact.

B. The Self-Perception Postulates

It was the above analysis which first suggested that many of the self-descriptive statements which appear to be exclusively under the control of private stimuli may in fact still be partially controlled by the same accompanying public events used by the training community to infer the individual's inner states in the first place (Bem, 1964, 1965). Meanwhile, the underlying corollary that private stimuli probably play a smaller role in self-description than we have come to believe—either as self-observers ourselves or as psychologists—was already receiving support from the experimental work on emotional states by Schachter and his colleagues (Schachter, 1964).

For example, in their now classic experiment, Schachter and Singer (1962) manipulated the external cues of the situation and were able to evoke self-descriptions of emotional states as disparate as euphoria and anger from subjects in whom operationally identical states of physiological arousal had been induced. These subjects required the internal stimuli of arousal in order to make the gross discrimination that they were emotional, but the more subtle discrimination of *which* emotion they were experiencing was under the control of the external cues, the "emotional" behavior of a stooge. In another experiment conducted within Schachter's theoretical framework, Valins (1966) was able to manipulate attitudes toward stimulus pictures of seminude females by

giving his male subjects false auditory feedback which they could interpret as their heartbeat, thus showing that any internal stimulus control of attitude statements could be overridden by external cues.

The controlling external cues in these and similar experiments have usually resided in the social or physical setting in which the individual was placed or in verbal instructions given to him by the experimenter. But that is not the only possible source of such cues. To us as observers, the most important clues to an individual's inner states are found in his behavior. When we want to know how a person feels, we look to see how he acts. Accordingly, it seemed possible that when an individual himself wants to know how he feels, he may look to see how he acts, a possibility suggested anecdotally by such statements as, "I guess I'm hungrier than I first thought." It was from this line of reasoning that the first postulate of self-perception theory was derived: Individuals come to "know" their own attitudes, emotions, and other internal states partially by inferring them from observations of their own overt behavior and/or the circumstances in which this behavior occurs.

The second postulate of self-perception theory suggests a partial identity between self- and interpersonal perception: To the extent that internal cues are weak, ambiguous, or uninterpretable, the individual is functionally in the same position as an outside observer, an observer who must necessarily rely upon those same external cues to infer the individual's inner states.

Because self-perception theory is conceived of as a "behaviorist's" theory, it is important to emphasize that neither the interpersonal observer nor the individual himself is confined to inferences based upon overt actions only. Social psychologists (e.g., Asch, 1952) have long been critical of behavioral analyses of social interaction (e.g., the Miller-Dollard analysis of imitation; see Miller & Dollard, 1941) precisely because they feel that there is something more to interpersonal perception than responding to the overt behavior of another individual. In particular, so the criticism goes, behavioral analyses fail to explicate how it is that individuals are able to take account of one another's meanings, motives, intentions, and the like.

This criticism is often illustrated by citing cases in which identical behaviors may have different "meanings," meanings which observers have no difficulty discerning. For example, an individual might attribute "suddenly aroused determination" to his friend were he to see him chasing a mouse into the room with a raised broom. But he would attribute "fear" if an identical hasty entrance were to be followed rather than preceded by the mouse. Were no mouse present, he might well classify his friend's action as anger directed at him. In all three cases

the overt behaviors observed are the same, and it is not particularly illuminating simply to say that the individual is responding to the "intent" of his friend or to the "meaning" of the action, since it is precisely the intent and meaning which require explication. In this example, it is clear that the meaning of the action resides in the mouse, that is, the intent or meaning is inferred from the stimulus conditions that appear to be controlling the observed behavior. To a radical behaviorist, this *is* the "intent" or "meaning" of the behavior. This, then, is the "something more" of interpersonal perception: the ability to respond not only to the overt behavior of others, but to respond as well to the controlling variables of which their behavior appears to be a function. The first postulate of self-perception theory embraces this observation in its proposal that self-attributions are made from the individual's observations of his "own overt behavior and/or the circumstances in which it occurs," for most important among those "circumstances" are the apparent controlling variables of that behavior.

A less whimsical, more pertinent example of interpersonal perception in which the apparent controlling variables of the behavior can provide a basis of inference is provided by the individual who attempts to infer the actual beliefs or attitudes of a persuasive communicator. Is the communicator being paid? If so, how much? Did the communicator have a free choice in what to say, or was he coerced? To the extent that the communicator appears to be free from the control of these kinds of explicit reinforcement contingencies, that is, to the extent that he does not appear to be "manding" reinforcement (Skinner, 1957), he will be judged to be a credible communicator, and his statements will typically be described as his "actual" beliefs and attitudes.

If one now applies the postulates of self-perception theory to this same example, one arrives at the hypothesis that the communicator himself might infer his own beliefs and attitudes from his behavior if that behavior appears to be free from the control of explicit reinforcement contingencies. This hypothesis was, in fact, the first prediction to be derived from self-perception theory (Bem, 1964, 1965), and tentative evidence for its validity appeared to reside already in the forced-compliance experiments conducted within Festinger's (1957) theory of cognitive dissonance.

Consider, for example, the classic experiment by Festinger and Carlsmith (1959). In this experiment, 60 undergraduates were randomly assigned to one of three experimental conditions. In the $1 condition, the subject was first required to perform long repetitive laboratory tasks in an individual experimental session. He was then hired by the experimenter as an "assistant" and paid $1 to tell a waiting fellow student (a

stooge) that the tasks were enjoyable and interesting. In the $20 condition, each subject was hired for $20 to do the same thing. Control subjects simply engaged in the repetitive tasks. After the experiment each subject indicated how much he had enjoyed the tasks. The results show that the subjects paid $1 evaluated the tasks as significantly more enjoyable than did subjects who had been paid $20. Subjects paid $20 did not express attitudes significantly different from those expressed by the control subjects. This phenomenon is now known as the reverse-incentive effect.

As implied above, self-perception theory interprets such results by considering the viewpoint of an outside observer who hears the individual making favorable statements about the tasks to a fellow student, and who further knows that the individual was paid $1 [$20] to do so. This hypothetical observer is then asked to state the actual attitude of the individual he has heard. If the observer had seen an individual making such statements for little compensation ($1), he can rule out financial incentive as a motivating factor and infer something about the individual's attitudes. He can use an implicit self-selection rule and ask: "What must this man's attitude be if he is willing to behave in this fashion in this situation?" Accordingly, he can conclude that the individual holds an attitude consistent with the view that is expressed in the behavior: He must have actually enjoyed the tasks. On the other hand, if an observer sees an individual making such statements for a large compensation (e.g., $20), he can infer little or nothing about the actual attitude of that individual, because such an incentive appears sufficient to evoke the behavior regardless of the individual's private views. The subject paid $20 is not credible in the sense that his behavior cannot be used as a guide for inferring his private views. The observer's best guess, then, is to suppose that the individual's attitude is similar to that which would be expressed by anybody who was selected at random and asked for his opinion—the attitude of a control subject, in other words.

Self-perception theory asserts that subjects in the Festinger-Carlsmith experiment (and other experiments utilizing this paradigm) are themselves behaving just like these hypothetical observers. They survey their own behavior of making favorable statements about the task and then essentially ask themselves (implicitly): "What must my attitude be if I am willing to behave in this fashion in this situation?" Accordingly, they produce the same pattern of results as the outside observers: low-compensation subjects infer that they must agree with the arguments in their communication, whereas high-compensation subjects discard their behavior as a relevant guide to their "actual" attitudes and express

the same attitudes as the control subjects. In short, if one places the hypothetical observer and the communicator into the same skin, the findings obtained by Festinger and Carlsmith are the result. The final attitudes of the actual subjects in the experiment are thus viewed as a set of self-attributions made by the individual on the basis of his own behavior in the light of the contextual constraints in which this behavior appeared to be occurring.

Judges made blind ratings of the persuasive communications delivered by subjects in the Festinger-Carlsmith experiment and found that $1 communications were no more persuasive than $20 communications; in fact, the trend was in the other direction. This illustrates again that it is not the behavior *per se*, but the behavior in conjunction with its apparent controlling variables which provides the crucial information for either interpersonal or self-attributions.

The crux of the self-perception interpretation in such studies is that the individual's own behavior will be used by him as a source of evidence for his beliefs and attitudes to the extent that the contingencies of reinforcement for engaging in the behavior are made more subtle or less discriminable. Monetary inducement is not the only way of manipulating this parameter of self-credibility. For example, several cognitive dissonance studies have manipulated the degree to which the individual appears to have the choice of engaging in the behavior or refusing to do so. Thus if an outside observer sees an individual freely choosing to make opinion statements (e.g., "the tasks were fun and interesting"), he is more likely to take such statements at face value, to infer that they reflect the individual's true opinions, than if the individual was required or coerced into making such statements. The same analysis applies to the amount of justification given a subject for engaging in the counter-attitudinal behavior: The more justification given, the less likely it is that the behavior will be used by an observer or the individual himself to infer what he believes. When the operations of forced-compliance studies are submitted to a careful analysis in terms of the discriminative stimuli presented to the subject, then those conditions in which outside observers would base an inference of the individual's attitudes on the overt behavior are found to be the same conditions in which the individual's own attitudes are affected.

II. Experimental Evidence for the Theory

As we shall see in Section III, several other experiments and paradigms from the cognitive dissonance literature are amenable to self-perception interpretations. But precisely because such experiments

are subject to alternative interpretations, they cannot be used as unequivocal evidence for self-perception theory. In particular, the stimulus operations in such studies have several functional properties. For example, monetary compensation not only manipulates the parameter of self-credibility (according to self-perception theory), but that of incentive and reinforcement as well, a fact which has led to replication failures, replication reversals, and a profusion of interpretations. [For several discussions, see Abelson, Aronson, McGuire, Newcomb, Rosenberg, and Tannenbaum (1968, pp. 801–833).] Similarly, manipulations of justification, choice, commitment, and so forth involve a veritable jungle of complex and ambiguous stimulus operations, a feature of dissonance studies which has been a major target of criticism (e.g., Chapanis & Chapanis, 1964). It thus seemed desirable to design a "self-credibility" experiment in which the controlling stimuli would be "raised from birth" in the laboratory so that they would have no functional properties other than those postulated by the theory to be relevant to the self-credibility of the induced behavior. Secondly, it seemed desirable to eliminate the variance in the overt behavior which arises from permitting the subject to compose his own persuasive communications. Finally, reflecting the author's Skinnerian bias, it was decided to design experiments for testing self-perception theory in which every subject is his own control and provides a complete replication of the entire design. The cartoon experiment (Bem, 1964, 1965) was the first of these studies.

A. The Cartoon Experiment

The dependent variable in the cartoon study was the subjects' attitudes toward a series of magazine cartoons. The induced behavior which served as the source of self-attribution was the overt statement, "This cartoon is very funny" or, "This cartoon is very unfunny." The self-credibility parameter was manipulated with two colored lights whose functional properties were established through the preliminary training procedure described below.

The experimental session was disguised as a tape-recording session for the preparation of experimental stimulus materials. Each subject first underwent a training procedure in which he answered simple questions about himself. After each question was asked, a tape recorder was turned on, automatically illuminating a colored light. The subject was instructed to answer the question truthfully whenever the light was amber. Whenever the light was green, he was to make up a false answer to the question and say it aloud into the tape recorder. In this way, the subject learned that he could believe himself whenever he spoke in the presence of the amber light, but could not believe himself in the presence

of the green light. The lights were reversed for half the subjects. After this training procedure, the subject was shown a series of magazine cartoons which he had rated as "neutral," that is, as neither funny nor unfunny in a previous session. For each cartoon, he was instructed by the experimenter either to say, "This cartoon is very funny," or, "This cartoon is very unfunny." The tape recorder was turned on just before he made each statement so that one of the colored lights was on while he spoke, even though he was not instructed to attend to the lights in this portion of the session. Sometimes the "truth" light was illuminated, other times the "lie" light. After the subject had made each statement and the recorder (and light) had been turned off, he was asked to indicate his actual attitude toward the cartoon on an attitude scale. An awareness questionnaire was administered at the end of the session.

As the self-perception hypothesis predicted, the subjects changed their attitudes significantly more when they made their statements in the presence of the "truth" light than when they made their statements in the presence of the "lie" light. For example, if a subject had said "This is a very funny cartoon" in the presence of the "truth" light, then he subsequently rated this cartoon to be funnier than if he had made the statement in the presence of the "lie" light. As a parallel to the Festinger-Carlsmith dissonance experiment, the "truth" light replicated the controlling functional property of the $1 inducement, signifying to the individual that his behavior could be used as a guide to his "true" attitude; the "lie" light abstracted the function of the $20 inducement, signifying to the individual that his behavior was irrelevant to his true attitude. It should also be noted that no subject was aware of any attitude change nor of any effects of the two lights on his subsequent cartoon ratings.

B. The False Confession Experiment

The cartoon study demonstrated that the self-attributions known as attitude statements could be brought directly under the control of an individual's own verbal behavior and the accompanying stimulus conditions in which that behavior occurs. The false confession experiment (Bem, 1966) was an attempt to retain the same independent variables, but to extend the evidence for the theory to a different kind of dependent variable: the recall of prior events. In particular, the false confession study attempted to verify the possibility that a false confession can effectively distort an individual's recall of his past behavior if the confession is made in the presence of cues previously associated with telling the truth. A second possibility tested was that cues pre-

viously associated with lying can create self-disbelief in *true* statements, leading again to distortions in recall of the actual behavior.

The experiment utilized the same experimental setting as the cartoon experiment, but instead of rating cartoons prior to the experimental session, subjects were given a list of 100 common nouns and an alphabetical list containing 50 of those nouns. Their task was to cross out each word on the master list which also appeared on the alphabetical list. Thus each subject was required to scan the alphabetical list for each of 100 words and then either to cross out the word or to not cross out the word. This was the behavior which the subject was later asked to recall.

Each subject then went through the preliminary training session, described above, in which he learned to make true statements whenever the amber recording light was on and false statements whenever the green recording light was on. He was then required to make statements about the 100 nouns. Sometimes he was required to state aloud that he had crossed out a word and sometimes to state that he had not crossed out a word (e.g., "I did not cross out the word *tree*"). Half the statements he had to make were true, and half were false. Again the colored lights were connected to the tape recorder so that either the amber or the green light was illuminated as he made his "confession." After each statement was made and the recorder and light turned off, the subject indicated whether he recalled crossing out the word or not crossing it out. In addition, he indicated on a five-point scale how confident he was in the accuracy of his recall. The complete design for each subject was thus a $2 \times 2 \times 2$ in which the word had either been crossed out or not, the statement was either true or false, and it was made in the presence of either the "truth" light or the "lie" light. In addition, each subject was asked to recall his behavior on a set of control words about which no overt statements had been made.

The results confirmed the expectations of the self-perception hypothesis: In the presence of the "lie" light, the false confessions had no effect. Subjects were as able to recall their previous behavior just as accurately as they were for the control words not appearing in their confessions. But in the presence of the "truth" light, false confessions did produce significantly more errors of recall and less confidence in recall accuracy than either false confessions in the presence of the "lie light" or no confession at all. In addition, there was weak support for the complementary possibility that more recall errors would be produced by the "lie" light when the overt statement was true than would be produced by the "truth" light—where the light confirms rather than contradicts the validity of the statement.

In a replication of this experiment, Maslach (1971) reconfirmed the major finding that the truth light produced more errors of recall following false confessions, but failed to find that the "lie" light distorted recall for true statements. Instead, she reports that the "truth" light produced more errors of recall in both the false and true confession conditions, although the effect was stronger for false confessions. This opens up the possibility that the subjects are exercising more care in responding after a "lie" light statement, an hypothesis supported by auxiliary data. Unfortunately, because Maslach's design did not contain a set of control words, it is not possible to know which cells are deviating significantly from a recall baseline and hence which combination of independent variables is producing the effects. Nevertheless, her pattern of results casts a cloud over the false confession experiment as a pure demonstration of a self-perception effect.

C. The Pain Perception Experiment

Both the cartoon and false confession studies employed the overt behavior of self-persuasive statements and self-credibility cues as the independent variables. The pain perception study (Bandler, Madaras, & Bem, 1968) was designed to move beyond this conceptual paradigm and also to make contact with an extensive literature on pain perception. It has long been known that an individual's perception of pain is only partially a function of the pain-producing stimulus. This is apparent from the wide cultural differences in labeling stimuli as painful (e.g., childbirth; Melzack, 1961), from research on placebo effects (Beecher, 1959, 1960), and from the phenomena of hypnotic analgesia (Barber, 1959, 1963) and masochism (Brown, 1965).

Recent research has also revealed a number of "cognitive" operations which can influence an individual's inference that a particular stimulus is painful. For example, Nisbett and Schachter (1966) showed that the judged intensity of shock-produced pain and the willingness to tolerate such pain can be manipulated by supplying the individual with an alternative explanation for the physiological arousal he is experiencing. And, working within a cognitive dissonance framework, Zimbardo, Cohen, Weisenberg, Dworkin, and Firestone (1969) demonstrated that individuals who have volunteered to continue participation in an experiment using painful electric shocks and who were given little justification for volunteering to do so reported the shocks to be less painful and were physiologically (GSR) less responsive than individuals who were given no choice about continuing or who were given much justification for volunteering to continue.

As noted earlier, these parameters of choice and justification can

be embraced within the self-perception framework. If an outside observer sees an individual choosing freely to continue enduring shocks with little justification, the observer might well assume that the shocks are not very painful. If on the other hand, the individual appears to have been given little or no option to endure the shocks, the observer could infer nothing about their painfulness from the individual's behavior. Again, if one places this hypothetical observer in the same skin as the subject, then the pattern of results displayed in the experiment by Zimbardo *et al.* (1969) is the result.

There is another kind of behavior which an observer can use to judge the painfulness of the shock: Does the individual attempt to escape from the shock or does he endure it? It seems reasonable to suppose that an observer would judge shocks from which the individual tried to escape as more uncomfortable than shocks the individual was willing to endure. Self-perception theory, then, predicts that the individual who observes himself freely choosing to escape a shock should rate it as more painful than a shock he chooses to endure. On the other hand, if he is not permitted to choose the action, but is simply instructed by the experimenter to escape the shock, he should not rate it as any more painful than a control shock. The study by Bandler *et al.* (1968) attempted to verify this hypothesis.

Each subject in the experiment received a series of shocks to his hand. Just after the onset of each shock, one of three colored lights was illuminated on a box in front of him. He was told that if the red light came on, he should press a button which he held in his hand and this would terminate the shock. In order to enhance feelings of choice, he was also told, "However, if the shock is not uncomfortable, you may elect to not depress the button. The choice is up to you." This condition was labeled the escape condition. In the green-light condition, subjects were instructed to not depress the button (the no-escape condition) unless the "shock is so uncomfortable that you feel you must . . . Again the choice is up to you." In the yellow-light condition, subjects were told that "we are interested in recording only the time that it takes you to press the button once the yellow light comes on. Therefore, please press the button as soon as the yellow light is illuminated. Your depression of the button *may* or *may not* turn off the shock."

Following each shock the subject rated the discomfort it produced on a 7-point rating scale. The 30 critical shocks were all of equal intensity, and, if not terminated by the subject, had a duration of 2 seconds. Thus, for 10 shocks paired with the escape light, the subject pressed a button and terminated the shock. For the 10 shocks paired with the no-escape light, the subject did not press the button. (Despite

the implied choice, subjects overwhelmingly complied with the instructions.) And, finally, for the 10 shocks paired with the "reaction-time" trials, the subject pressed the button as soon as he could after the light was illuminated. For five of these trials, pressing the button terminated the shock. For the remaining five reaction-time trials, pressing the button had no effect on the shock. It should be noted that the subject's overt behavior was the same in the reaction-time condition as it was in the escape condition: He pressed the button when the light was illuminated. But the subject was not given the implied choice of pressing or not pressing and, as the instructions make clear, pressing the button did not necessarily terminate the shock. Thus in the reaction-time condition, the button press should no longer be seen by the subject as a self-determined escape response, and he should not infer his discomfort from it. Discomfort ratings should therefore be significantly lower in the reaction-time condition than in the escape condition.

The results supported the self-perception hypothesis. Subjects rated shocks as significantly more uncomfortable when they escaped them than when they endured them, the same inference an outside observer might have drawn. This was true even though the endured shocks were necessarily longer than the escaped shocks. (Within the reaction-time condition, the endured shocks were rated as slightly more uncomfortable than the briefer terminated shocks.) Moreover, the subject had to perceive that he had some choice in the matter. Pushing the button in the reaction-time condition was not used as a guide for inferring the discomfort of the shock. Ratings of shocks in the reaction-time condition were significantly lower than those in the escape condition and were not significantly different from ratings of shocks in the no-escape condition.

This experiment has since been replicated by Corah and Boffa (1970) using white noise as the aversive stimulus and a between-subjects design for the choice parameter. That is, half of their subjects were given the implied choice of escaping or not escaping, just as Bandler *et al.* had done, but the remaining subjects were not given such a choice. The results of Bandler *et al.* were confirmed. When given choice, subjects rated bursts of noise they escaped as significantly more aversive than bursts of noise they endured. When the choice was omitted, endured noise was slightly more aversive than the escaped noise, a replication of the Bandler *et al.* finding within the reaction-time condition.

A second replication of the Bandler *et al.* experiment was conducted by Klemp and Leventhal (1972), who controlled for stimulus duration by contrasting the shocks the subject "chose" to escape with the reaction-time shocks he escaped (i.e., without choice) and the shocks he

"chose" to endure with the reaction-time shocks which were not terminated. These investigators found that the self-perception effect—chosen escaped shocks rated more painful than unchosen escaped shocks—held only for subjects who had a high tolerance for shock. An opposite pattern was found for low tolerance subjects: After choosing to escape a shock, they rated it as less uncomfortable than a shock which is escaped without choice. Klemp and Leventhal suggest, among other possibilities, that low-tolerance subjects may be more frightened, i.e., they may show more of the bodily components of fear, and fear might intensify the judged discomfort of shock. If being able to choose is critical in reducing this fear (as other research implies), then lower discomfort ratings in the choice conditions would be expected.

This finding is also consistent with an earlier finding by Nisbett and Schachter (1966), who actually manipulated fear in a self-attribution experiment involving electric shock. The predicted self-perception effect occurred only in the low-fear condition; there were no significant differences between conditions in the high-fear treatment.

It will be recalled that the basic self-perception postulate states that self-perception effects based upon observations of overt behavior occur "to the extent that internal cues are weak, ambiguous, or uninterpretable." The findings of Klemp and Leventhal and of Nisbett and Schachter would appear to exemplify cases in which the internal stimuli of fear override any information potentially available to the individual from external sources, including overt behavior.

III. The Reinterpretation of Cognitive Dissonance Phenomena

If a person holds two cognitions that are inconsistent with one another, he will experience the pressure of an aversive motivational state called cognitive dissonance, a pressure which he will seek to remove, among other ways, by altering one of the two "dissonant" cognitions. This proposition is the heart of Festinger's theory of cognitive dissonance (Festinger, 1957), a theory which received more widespread attention from personality and social psychologists during the decade of the Sixties than any other contemporary statement about human behavior.

In 1965, it was suggested that self-perception theory might be able to account for the major phenomena of cognitive dissonance theory (Bem, 1965), a suggestion which was elaborated more systematically in Bem (1967b). The basic theme of that reinterpretation has already been spelled out in Section I, B, with respect to the Festinger and Carlsmith (1959) experiment. Now that the experiments providing independent evidence for the self-perception postulates have been de-

scribed, it is appropriate to return to the conceptual reanalysis of the various dissonance paradigms.

A. The Forced-Compliance Studies: The Psychology of Insufficient Justification

The Festinger-Carlsmith study exemplifies the experimental procedure known as the forced-compliance paradigm. In all of these experiments, an individual is induced to engage in some behavior that would imply his endorsement of a particular set of beliefs or attitudes. Following his behavior, his "actual" attitude or belief is assessed to see if it is a function of the behavior in which he has engaged and of the manipulated stimulus conditions under which it was evoked. Thus in the Festinger-Carlsmith experiment, subjects were induced for either $1 or $20 to tell a waiting fellow student that the repetitive tasks were fun and interesting, and then their own attitudes were assessed. As noted earlier, the low-compensation subjects expressed significantly more favorable attitudes toward the tasks than the high-compensation subjects or control subjects, the reverse-incentive effect.

Dissonance theory interprets these results by noting that all subjects initially hold the cognition that the tasks are dull and boring. In addition, however, the experimental subjects have the cognition that they have expressed favorable attitudes toward the tasks to a fellow student. These two cognitions are dissonant for subjects in the $1 condition because their overt behavior does not "follow from" their cognition about the task, nor does it follow from the small compensation they are receiving. To reduce the resulting dissonance pressure, they change their cognition about the task so that it is consistent with their overt behavior: they become more favorable toward the tasks. The subjects in the $20 condition, however, experience little or no dissonance because engaging in such behavior "follows from" the large compensation they are receiving. Hence, their final attitude ratings do not differ from those of the control group. It is the motivational force provided by the drive toward consistency or dissonance reduction which changes the attitudes of the low-compensation subjects.

As noted earlier, self-perception theory considers the subject in such an experiment as simply an observer of his own behavior. Just as an outside observer might ask himself, "What must this man's attitude be if he is willing to behave in this fashion in this situation?" so too, the subject implicitly asks himself, "What must my attitude be if I am willing to behave in this fashion in this situation?" Thus the subject who receives $1 discards the monetary inducement as the major motivating factor for his behavior and infers that it must reflect his actual

attitude; he infers that he must have actually enjoyed the tasks. The subject who receives $20 notes that his behavior is adequately accounted for by the monetary inducement, and hence he cannot extract from the behavior any information relevant to his actual opinions; he is in the same situation as a control subject insofar as information about his attitude is concerned. Thus in the self-perception explanation, there is no aversive motivational pressure postulated. The dependent variable is viewed simply as a self-attribution based on the available evidence, which includes the overt behavior of the communication and the apparent controlling variables of that behavior. A similar analysis can be applied to other experiments conducted within the forced-compliance paradigm.

The advantages of the self-perception explanation emerge more clearly in forced-compliance experiments in which dissonance theory does not provide an applicable account of the findings. Three such experiments can be mentioned here.

1. Pro-Attitudinal Advocacy

Kiesler, Nisbett, and Zanna (1969) conducted a forced-compliance study specifically designed to rule out the dissonance explanation by inducing subjects into committing themselves to argue a position with which they initially agreed. Such a situation presumably does not arouse the "dissonance" produced by counterattitudinal advocacy. In addition, Kiesler *et al.* manipulated the "meaning" of the behavior by having a confederate subject express his willingness to argue the issue assigned to him because (*a*) he believed strongly in it (belief-relevant condition), or (*b*) the experiment was scientifically valuable (belief-irrelevant condition). The confederate's issue was not the same one assigned to the actual subject. The results produced a self-perception effect: Subjects in the belief-relevant condition were found to be more favorable to the position they were to argue than either belief-irrelevant subjects or control subjects who were not committed to the behavior.

2. Rejection of Alternative Action

A second study in which subjects were to give a speech advocating a position they already endorsed was conducted by Zanna (1970). This study was specifically designed to show that the individual's inferences about his beliefs may be based not only on acts he performs but also on alternative acts he rejects. After committing themselves to giving a speech which advocated a moderate position on an issue, subjects were given the opportunity to alter the speech so that it advocated a position more extreme, an opportunity which was arranged so that all subjects

would nevertheless choose to reject it. Control subjects were not given the opportunity to alter the speech. A belief-relevance manipulation similar to that in the Kiesler *et al.* (1969) study was introduced for all subjects. The results supported the self-perception analysis: Subjects who had observed themselves decline a chance to make their speech more extreme attributed to themselves an attitude toward the issue less extreme than the self-attributions made by control subjects. In further support of the self-perception postulate that such self-attributions are like interpersonal attributions, Zanna also conducted a successful parallel experiment which showed that observers generate the same pattern of results as the actual subjects. This kind of study is now known as an interpersonal simulation (Bem, 1968c), and will be discussed at length in Section III, C.

Harvey and Mills (1971) conducted an experiment conceptually similar to the Zanna study in which subjects were given the opportunity to substitute a different speech for the one they had initially given. As in the Zanna study, the opportunity was offered in such a way that all subjects rejected it. Again the results showed that subjects who were given the opportunity and rejected it expressed final attitudes more in line with the speech they had given than subjects who were not given that opportunity. However, because the speech in this study was counter-attitudinal, both dissonance theory and self-perception theory can predict this result (notwithstanding the contention of Harvey and Mills that this study favors dissonance theory over self-perception theory). A dissonance-theoretic prediction that attitude change would be greater if the subject's initial attitude were made more salient for him was not confirmed; in fact, the differences were in the opposite direction. As we shall see in Section III, D and E, self-perception theory suggests that making an initial attitude more salient should *diminish* the degree to which self-attributions can be altered by observations of overt behavior.

3. *The Effects of Repeated Advocacy*

Cohen (Brehm & Cohen, 1962, pp. 97–104) conducted an experiment to see what the effects of repeated "dissonances" might be on attitude change. Although dissonance theory might predict that dissonance pressure would build up, yielding greater final attitude change, Brehm and Cohen note that it is also possible for a "tolerance for dissonance" to build up at the same time. Thus they ventured no specific predictions. In this study, subjects in the predissonance conditions were required to write four separate essays against positions they currently held; subjects in the preconsonance conditions wrote four essays in favor of positions they currently held. All subjects were then induced to write a fifth essay which

was counter-attitudinal. The fifth essay was preceded by a manipulation of justification. Half of the subjects were given little justification for writing the fifth essay; half were given a great deal of justification. Following the writing of the final essay, attitudes were assessed on that one issue.

The pattern of results did not lend itself to any simple dissonance or "tolerance" of dissonance interpretation. Within the predissonance conditions, there was no attitude change at all. But within the preconsonance conditions, the usual effect of justification was found: significantly greater attitude change when little justification was given than when a great deal of justification was given. Another way of stating the results is that attitude change occurred only in the preconsonance-low justification condition.

The self-perception analysis of this pattern of findings is straightforward. In fact, the experiment is remarkably like the cartoon and false confession experiments described in Section II. Each subject receives pretraining which informs him either that he can believe what he says (preconsonance condition) or that he cannot (predissonance condition). Thus preconsonance subjects arrive at the fifth essay with their self-credibility intact, a credibility which can be maintained (low justification) or destroyed (high justification) for the crucial fifth essay. The predissonance subjects arrive at the fifth essay with their self-credibility already destroyed, and hence further manipulations have no effect, and no attitude change is observed.

4. Combining Sources of Credibility Information

The experiment on repeated "dissonances" makes explicit an underlying assumption of self-perception theory concerning the way in which different sources of information combine. Self-perception theory states that if external contingencies seem sufficient to account for the behavior, then the individual will not be led into using the behavior as a source of evidence for his self-attributions. This suggests that if any elements of the situation imply that the behavior is not credible or is not otherwise relevant to his private views, then the presence of other cues implying credibility does not help. Thus, if either predissonance training *or* high justification was present in the repeated dissonance study, the behavior was not used as the basis for self-attribution. Another way of saying this is that the sources of self-credibility combine multiplicatively: If any source implies irrelevance or low credibility, then self-attributions will not occur.

More direct support for this combination rule is provided in an experiment by Linder and Jones (1969) which is a clever hybrid between a

self-perception study and a dissonance study. Their subjects were trained to make true and false statements in the presence of two colored lights following exactly the procedures used in the cartoon and false-confession studies. Thus subjects learned that they could believe what they said in the presence of the "truth" light, and that they could not believe themselves in the presence of the "lie" light. All subjects then had to read counter-attitudinal essays in the presence of one of the two lights. Half of the subjects were permitted a "free" choice whether or not to comply; the remaining subjects were given no choice. Final attitudes were then assessed. The results show that significantly greater agreement with the position advocated in the essay appeared only when essays were read under free choice conditions in the presence of the "truth" light. In other words, if either the "lie" light was on *or* if the subject had no choice, self-attributions based upon the behavior did not occur. The rule for combining conflicting sources of credibility information appears to be multiplicative.

5. The Forbidden Toy Studies

One variation of the forced-compliance paradigm which has been employed with increasing frequency is the forbidden toy procedure. In this situation, children are asked not to play with an attractive toy and are given either a mild or a severe threat of punishment for disobeying. Following a period in which all the children do resist the temptation, children under mild threat have been shown to devalue the prohibited toy (Aronson & Carlsmith, 1963). Unlike the studies of counter-attitudinal advocacy, which frequently fail to replicate, the forbidden toy finding has proved extremely reliable, producing both verbal derogation of the previously forbidden toy (Aronson & Carlsmith, 1963; Carlsmith, Ebbesen, Lepper, Zanna, Joncas, & Abelson, 1969; Lepper, Zanna, & Abelson, 1970; Ostfeld & Katz, 1969; Turner & Wright, 1965) and actual behavioral avoidance of the toy up to 6 weeks after the initial session (Freedman, 1965; Pepitone, McCauley, & Hammond, 1967).

The dissonance theory account of this effect is that a severe threat provides high justification for the child not to play with the toy, rendering the temptation period nondissonant; the mild threat, however, is insufficient to justify his not playing with the toy, producing dissonance which can be resolved by convincing himself he really doesn't like the toy. The self-perception account is also straightforward. The child, when asked his toy preferences after the temptation period reviews his behavior toward the toy and the apparent controlling variables of his avoiding the toy. If he has refrained from playing with the toy under severe threat, he can still infer that he may like the toy, but if he has

refrained under mild threat, then he could conclude that he must not like the toy. Again, it is as if subjects use the implicit self-selection rule, "What must my attitude be if I am willing to behave in this fashion in this situation."

Both theories can also handle certain variations of the procedure. For example, Lepper *et al.* (1970) show that if the child is told prior to the temptation period that other children did not play with the toy when asked, then mild-threat subjects no longer devalue the toy. Within the dissonance theory framework, this consensual information gives the child a "consonant" reason for not playing with the toy, and hence no dissonance is aroused. Similarly, in self-perception theory, the information prevents the subject from using the implicit self-selection rule; because everyone behaves the same way, the child cannot use such normative behavior to infer something unique about his attitudes from it. [Kelley (1967) has pointed out the importance to the self-attribution process of having subjects believe their behavior is distinctive if it is to be used as the basis of inference.]

But both theories require added assumptions to account for the second finding of Lepper *et al.*: Mild-threat subjects still do devalue the toy if the consensual information is given after rather than before the temptation period. This implies that the attitudinal attribution takes place during the temptation period and is irreversible. Dissonance theory does not commit itself with regard to when dissonance reduction takes place, but the implication that it is a continuing process which begins immediately and which, once accomplished, is resistant to reversal is consistent with the spirit of the theory. In self-perception theory, however, it has always been implicitly assumed that the attribution probably occurs when the experimenter asks the individual for his final opinion, at which time he reviews the immediate past for the answer. The Lepper *et al.* experiment implies that the child seeks an inferential account for his behavior earlier, during the temptation period. Something prompts him to ask, "How come I'm not playing with this toy?" But even if this assumption is added to the self-perception account, there is still nothing within the theory which would predict that the child's attribution should resist change when new information comes in (e.g., the consensual information). In this instance, then, self-perception theory requires more "patching up" to account for the results than does dissonance theory. Interestingly, however, the forbidden toy paradigm has simultaneously become the vehicle for demonstrating two nondissonance, self-perception phenomena (Lepper, 1971). These will be discussed in detail in Sections IV, A and B.

These, then, are the major phenomena which have emerged from the

forced-compliance paradigm of cognitive dissonance theory. As we shall see in Section VII, there are hidden intricacies in these phenomena which outrun all current theories, including the present analysis. At this juncture, however, this section may be taken to constitute the psychology of insufficient justification as seen from the perspective of self-perception theory.

B. The Free Choice Studies

After the forced-compliance paradigm, it is the free choice paradigm which has received the most attention within dissonance theory (Brehm & Cohen, 1962). In these studies, a subject is permitted to make a selection from a set of objects or courses of actions. The dependent variable is his subsequent attitude rating of the chosen and rejected alternatives. Dissonance theory reasons that any unfavorable aspects of the chosen alternative and any favorable aspects of the rejected alternatives provide cognitions that are dissonant with the cognition that the individual has chosen as he did. To reduce the resulting dissonance pressure, the individual exaggerates the favorable features of the chosen alternative and plays down its unfavorable aspects. This leads him to enhance his rating of the chosen alternative. Similar reasoning predicts that he will lower his rating of the rejected alternatives. These predictions are confirmed in a number of studies (see Brehm & Cohen, 1962, p. 303; Festinger, 1964).

A number of secondary predictions concerning parameters of the choice have also been confirmed. In an experiment by Brehm and Cohen (1959), school children were permitted to select a toy from either two or four alternatives. Some children chose from qualitatively similar toys; others chose from qualitatively dissimilar alternatives. The children's postchoice ratings of the toys on a set of rating scales were then compared to initial ratings obtained a week before the experiment. The main displacement effect appeared as predicted: Chosen toys were displaced in the more favorable direction; rejected toys were generally displaced in the unfavorable direction. In addition, however, the displacement effect was larger when the choice was made from the larger number of alternatives. This is so, according to dissonance theory, because "the greater the number of alternatives from which one must choose, the more one must give up and consequently the greater the magnitude of dissonance [Brehm & Cohen, 1959, p. 373]." Similarly, the displacement effect was larger when the choice was made from dissimilar rather than similar alternatives because "what one has to give up relative to what one gains increases [p. 373]," again increasing the magnitude of the dissonance experienced.

Again, self-perception approaches the phenomenon by considering an observer trying to estimate a child's ratings of toys; the observer has not seen the child engage in any behavior with the toys. Now compare this observer with one who has just seen the child select one of the toys as a gift for himself. This comparison parallels, respectively, the prechoice and the postchoice ratings made by the children themselves. It seems likely that the latter observer would displace the estimated ratings of the chosen and rejected alternatives further from one another simply because he has some behavioral evidence upon which to base differential ratings of these toys. This is the effect displayed in the children's final ratings.

The positive relation between the number of alternatives and the displacement can be similarly analyzed. If an observer had seen the selected toy "win out" over more competing alternatives, it seems reasonable that he might increase the estimated displacement between the "exceptional" toy and the group of rejected alternatives. Finally, the fact that the displacement effect is larger when the alternatives are dissimilar would appear to be an instance of simple stimulus generalization. That is, to the extent that the chosen and rejected alternatives are similar to one another, they will be rated closer together on a scale by any rater, outside observer, or the child himself.

In sum, if one regards the children as observers of their own choice behavior and their subsequent ratings as inferences from that behavior, the dissonance findings appear to follow.

C. The Interpersonal Simulations of Cognitive Dissonance Studies

As we have seen, the reinterpretation of dissonance studies proceeds by utilizing the second postulate of self-perception theory: To the extent that internal cues are weak, ambiguous, or uninterpretable, the individual is functionally in the same position as an outside observer. In order to bolster the purely conceptual use of this postulate in the reinterpretation of the dissonance studies, an experimental methodology was devised for testing this notion more experimentally. The resulting experiments within this methodology have now become known as interpersonal simulations (Bem, 1965, 1967a, 1967b, 1967c, 1968a, 1968b, 1968c; Bem & McConnell, 1970). In these experiments, an observer-subject is either given a description of one of the conditions of a dissonance experiment or actually permitted to observe one of these conditions and then asked to estimate the attitude of the subject whose behavior is either described or observed. The prediction is that such observer-subjects should be able to reproduce the patterns of results

generated by actual subjects in the original experiments. It should be noted that observer-subjects are not asked to play amateur psychologist and predict the results of the experiment; rather, each observer-subject attempts to infer the attitude of a single "other." In the simulation, he "stands in" for the actual subject.

For example, the free choice study described in the prior section was simulated by giving college students a sheet of paper which informed them that an 11-year-old boy in a psychology experiment was permitted to select one of several toys to keep for himself (Bem, 1967b). The sheet informed the observer-subject which toy the child had chosen and from which alternatives he was permitted to choose. Each observer-subject then estimated the child's ratings of the chosen and rejected toys. These observer-subjects were randomly assigned to one of four conditions corresponding to the combinations of number of alternatives (two or four) and similarity of alternatives (similar or dissimilar). The toys listed were selected from the list reported in the original experiment (Brehm & Cohen, 1959) and were rated on the same rating scales. A group of control observer-subjects simply estimated toy ratings of a typical 11-year-old boy. The results showed that observer-subjects not only replicated the main displacement effect (chosen toy enhanced, rejected toy devalued), but also replicated the effects of the number of toys (displacement effect enhanced with more rejected alternatives) and the effects of the similarity of the toys (displacement effect diminished with similar alternatives).

A somewhat more elaborate simulation was conducted for the Festinger and Carlsmith (1959) study in which, it will be recalled, subjects who had received $1 for telling a stooge that a series of tasks were fun and interesting subsequently gave the tasks higher ratings than either subjects paid $20 or control subjects.

Seventy-five college undergraduates participated in an experiment designed to "determine how accurately people can judge another person." Twenty-five subjects each served in a $1, a $20, or a control condition. All subjects listened to a tape recording which described a college sophomore named Bob Downing, who had participated in an experiment involving two motor tasks. The tasks were described in detail, but non-evaluatively; the alleged purpose of the experiment was also described. At this point, the control subjects were asked to evaluate Bob's attitudes toward the tasks. The experimental subjects were further told that Bob had accepted an offer of $1 ($20) to go into the waiting room, tell the next subject that the tasks were fun, and to be prepared to do this again in the future if they needed him. The subjects then listened to a brief conversation which they were told was an actual recording of Bob and the girl who was in the waiting room. Bob was heard to argue rather imagi-

natively that the tasks were fun and enjoyable, while the girl responded very little except for the comments that Festinger and Carlsmith's stooge was instructed to make. The recorded conversation was identical for both experimental conditions in order to remain true to the original study in which no differences in persuasiveness were found between the $1 and the $20 communications. In sum, the situation attempted to duplicate on tape the situation actually experienced by Festinger and Carlsmith's subjects. All subjects estimated Bob's responses to the same set of questions employed in the original study.

The results of this simulation showed that interpersonal observers were able to replicate the reverse-incentive effect. Observers of $1 subjects estimated "Bob's" attitude to be significantly more favorable toward the tasks than did observers of either $20 subjects or control subjects. Observers of $20 subjects did not differ in their estimates of Bob's attitude from observers of control subjects. An extended version of this same simulation (Bem, 1967b) replicated an interaction effect between the monetary compensation and the length of the communication found in the original study. Similar simulations conducted by the author (Bem, 1965) have replicated results from a forced-compliance experiment which utilized small compensations of 50¢ and $1 (Brehm & Cohen, 1962, p. 73) and results from an experiment on hunger ratings by Brehm and Crocker (Brehm & Cohen, 1962, pp. 133–136).

A particularly persuasive simulation was performed by Alexander and Knight (1971) in order to replicate a complex pattern of results found in a forced-compliance study by Carlsmith, Collins, and Helmreich (1966). In their experiment, Carlsmith *et al.* had shown that one obtains the classical reverse-incentive effect when the subject actually has a face-to-face encounter with the waiting stooge, but that an incentive effect emerges (more money, more attitude change) if the subject simply writes an anonymous essay stating that the tasks were fun and interesting. Alexander and Knight were able to simulate this interaction effect between the monetary inducement and the mode of counter-attitudinal behavior by modifying and extending Bem's use of tapes about "Bob Downing."[2]

There have also been several simulations which have failed to

[2] The successful interpersonal simulation of the Carlsmith *et al.* (1966) findings is actually the least important aspect of the Alexander-Knight article, and it trivializes their contribution to present it in this context. The simulation was actually the first step toward demonstrating their theory that the results of many experiments can be predicted from a knowledge of the most socially desirable response in the situation. Their interpretation of the forced-compliance studies in particular is an alternative to both the dissonance and self-perception analyses. For both methodological and theoretical reasons, the Alexander-Knight article is an important document for social psychologists.

replicate the original studies. Some of these failures have come from simulations specifically modified in ways designed to disconfirm the self-perception analysis (e.g., R. A. Jones, Linder, Kiesler, Zanna, & Brehm, 1968; Piliavin, Piliavin, Loewenton, McCauley, & Hammond, 1969). For reasons to be discussed below, such failures are not, in fact, informative with respect to the validity of the theory. Several other failures are also not informative in this regard because they yield only unsystematic noise rather than systematic differences between observer-subjects and actual involved subjects. (Most of these appear in unpublished manuscripts sent to the author over the years.)

The major reason such failures are uninformative is the unreliability of the dissonance phenomena themselves, particularly the reverse-incentive effect, the design most commonly simulated. Indeed, one has learned to fear that an experimenter's unintended belch might destroy or reverse the effect (for discussions, see Abelson *et al.*, 1968, pp. 801–833). It is only semifacetious to suggest that serious doubts about the self-perception analysis could be raised if the simulations were too reliable or more robust than the phenomena themselves, that is, if the simulations were not also sensitive to unknown and uncontrolled parameters. For example, Harris and Tamler (1971a) have shown that one of the simulations fails if it is conducted the way most of the theory's critics have conducted their simulations, but it succeeds if the observer knows that the assessment of final attitudes was made by an experimenter different from the one who manipulated the independent variable, hardly a predictable contingency.

Such variability of outcome suggests that simulations should probably always be conducted directly in conjunction with the experiment being simulated, either simultaneously or with stimulus materials and consequent behaviors actually generated from the experiment itself. Only a few simulations have been conducted in this way.

For example, Harris and Tamler (1971a, 1971b) actually conducted the dissonance experiment which separated the attitude assessment from the initial experimental setting and confirmed that the dissonance effect could be obtained in this way. Materials for their successful simulation were then carefully patterned after the experiment itself. As noted above, they demonstrated that the simulation failed if this single item of information was omitted.

A second example is provided by Zanna (1970) who conducted both the original experiment reported in Section III, A, 2, on the rejection of alternative actions as a source of self-attribution and an associated interpersonal simulation. The simulation successfully replicated his finding that rejection of the opportunity to make a more extreme speech results

in the attribution that the subject is more moderate. He also found, as predicted, that observer-subjects who were told that a subject rejected the opportunity to make the speech more moderate judged him to be more extreme than did control observer-subjects, a prediction that was not actually tested in the original experiment because of time limitations.

R. G. Jones (1966) has reported a study in which observer-subjects listened to tapes of the original sessions. Again observers' attributions replicated the inverse functional relation between monetary compensation and final attitude found with involved subjects.

The strategy of conducting both the original experiment and simulation simultaneously pays off even when the simulation fails. For example, in one study of this type, Wolosin (1969) uncovered some important systematic differences between observers and actors in their perceptions of the actor's freedom to comply with the experimenter's request. As we shall see in Section V, such a finding has important implications well beyond the controversy over the simulations, and Wolosin has wisely decided to follow up this auxiliary finding (Wolosin, 1971; Wolosin & Denner, 1970). At the same time, however, it should be noted that even the care and rigor of Wolosin's study did not guarantee that negative results on the simulation would provide an unequivocal disconfirmation of self-perception theory, because none of the main effects or predicted interactions on the *major* dependent variables (ratings of thirst, willingness to undergo water deprivation, and actual water consumption) were significant for the involved subjects themselves. Thus most—albeit not all—of the comparisons that Wolosin hoped to make between observers and involved subjects required him to argue that the columns of random noise generated by two types of observers do not fluctuate in harmony with the column of random noise generated by involved subjects.

D. The Controversy over the Interpersonal Simulations

Although there have been some simulations which have failed to reproduce the pattern of results found in the original dissonance experiments, the controversy surrounding the simulation methodology has focused on the simulations which are successful. In particular, the controversy has centered around the information that the observer-subject ought or ought not to be given concerning the original situation (Bem, 1967a, 1968c; Elms, 1967; R. A. Jones *et al.*, 1968; Mills, 1967; Piliavin *et al.*, 1969). Most of the critics have objected specifically to the fact that observer-subjects are not told the original subject's premanipulation attitude. The conceptual reason for this deliberate omission can be understood by considering some of the epistemological features of the simulation methodology.

Self-perception theory asserts that an individual's attitude statements and an observer's judgments about the individual's attitudes are "output statements" from the same internal "program." Both the individual and observer are assumed to use a self-selection rule: "What must my [this man's] attitude be if I am [he is] willing to behave in this fashion in this situation?" To test this isomorphism, we run a simulation of a self-judgment situation, the dissonance experiment, but instead of writing our own program, we plug in an interpersonal judgment program that the culture has written for us. This program is embodied in our interpersonal observer, who "stands in" for the original subject.

But before we can actually run such a simulation, we must first abstract the relevant "input statements" from the situation being simulated: we must decide how to describe the situation to the observer. This requires some theoretically guided assumptions. For example, if the dissonance experiment subject actually arrives at his final attitude by using the self-selection rule, as the theory implies, then it follows that any conflicting initial attitude he may have had prior to the experiment must no longer be very salient for him. That is, the self-perception analysis implies that the data of his incoming behavior "update" his information on his attitude, replacing any prior information to the contrary. Insofar as the individual himself is concerned, his postmanipulation attitude is, in fact, the same attitude which motivated him to comply in the first place; phenomenologically, there is no attitude "change" as such. If an interpersonal simulation is to constitute a valid test of the isomorphism between the subject and an observer, then the theory dictates that a conflicting "initial" attitude of the original subject must not be part of the "input" description for the observer any more than it is for the subject himself. The observer, too, is postulated to be using the self-selection rule to infer the original subject's postmanipulation attitude.

It should be noted that this set of assumptions about what input information an observer-subject in the simulations should receive is self-correcting. If the wrong input statements are selected, then the simulation will not succeed in producing output statements which match the output of the original experiment. Thus, R. A. Jones *et al.* (1968) reconfirmed that the simulations produce the "dissonance effect" outputs when the inputs dictated by the self-perception model are employed, but they found that the simulations fail when a conflicting "initial" attitude is introduced into the description given to the observer-subject. After an intensive analysis of observer-subject's inferential processes under several variations of the simulations, Piliavin *et al.* (1969) reported that when the "Bem" inputs are utilized, observer-subjects do, in fact, utilize the self-selection rule and replicate his results. But when additional infor-

mation, including a reference to the subject's initial attitude, is introduced into the description, the observer-subjects become amateur psychologists and revert to hypotheses about attitude *change*. They are no longer stand-ins for the original subjects and, accordingly, they fail to reproduce the dissonance effects.

Such results underscore the importance of conducting simulations directly in conjunction with the experiment being simulated. For example, in the combined experiment and simulations of Harris and Tamler (1971a, 1971b) mentioned above, these investigators first verified that they could obtain the reverse-incentive effect even if the subject's initial attitude were made salient for him just prior to the forced-compliance manipulation (reinstatement condition) *and* if the final attitude was assessed by a different experimenter in a "different experiment" (separation condition). They then conducted a $2 \times 2 \times 2$ simulation in which two levels of monetary incentive were crossed with the reinstatement *vs.* nonreinstatement variable and the separation *vs.* nonseparation variable. Under nonseparation conditions [which parallel the R. A. Jones *et al.* (1968) and Piliavin *et al.* (1969) designs], these investigators obtained an incentive effect when reinstatement was employed, and a reverse-incentive effect when it was not. But under separation conditions, both reinstatement and nonreinstatement conditions generated the reverse-incentive effect predicted by self-perception theory. When the actual experiment was simulated exactly, the simulation worked.

It is important to note just what a successful simulation means. It implies the same thing that a successful computer simulation implies, namely, that the process model embodied in the "program" is functionally equivalent to the process being simulated, and, further, that the selection of the input statements was not in error. The simulation becomes a plausibility demonstration, a sufficiency test: The process model embodied in the program is sufficient—but not necessarily "true" or unique—for generating the output statements observed in the situation being modeled by the simulation.

But there are weaknesses in this methodology when it is applied to this context. Abelson (1968) has noted some of the validation problems connected with the simulation methodology in general, and his discussion of the "degrees of freedom" problem is particularly relevant to interpersonal simulations. Abelson stated the problem this way:

> If a simulation could be "right for the wrong reasons," that is, fit the data by virtue of compensating errors, then in what sense can a good fit be regarded as support for the theory underlying the simulation model? . . . Most cognitive simulations are so rich in qualitative detail that it is very easy for them to fail . . . Because it is so hard to obtain good data fits, anything which comes close is impressive, and any

cognitive model yielding an apparently perfect fit to a wide range of data would indeed deserve serious theoretical recognition.

With social simulations, however, the issue is probably more cogent. If the outcome variables of the model are few while the number of parameters to be juggled is great, there can always be the lingering suspicion that a good fit was too easy to achieve and thus not strongly supportive of the model [pp. 343–344].

This, then, is the main reason why the interpersonal simulations provide only weak support for the self-perception theory. There are too many "degrees of freedom," input variables that can be "juggled" (like the initial attitude), while the complexity of the output predictions rarely exceeds a prediction about the ordering of two or three means. Abelson suggests some of the paths open for strengthening simulation arguments, however. The most obvious remedy, of course, is "to design the simulation so as to generate as large a number of outcome variables as possible. The more outcomes that can be validated, the merrier—and the more convincing the underlying theory [Abelson, 1968, p. 344]."

This path toward strengthening the self-perception theory of dissonance phenomena was followed in three interpersonal simulations (discussed in Section III, C) which not only replicated main effects of dissonance experiments but reproduced either an interaction effect which dissonance theory itself had not anticipated or the secondary effects of additional parameters in the experiments (Alexander & Knight, 1971; Bem, 1967b). Although these extended simulations do not go very far toward eliminating the "degrees of freedom" problem, they are illustrative of the method.

A second remedy is "to show, if possible, that the fit was not so easy by changing the model in various ways and demonstrating consistent lack of fit [Abelson, 1968, p. 343]." In effect, this is what some of the critics have done. They have altered some of the input assumptions of the model and demonstrated that the simulations fail (R. A. Jones *et al.*, 1968; Piliavin *et al.*, 1969). As Abelson has remarked:

Ironically, what [Bem's] detractors should now really be doing if they must still simulate is to replicate [his] outcome with clearly bad descriptions to the observer, rather than to reverse [his] outcome with purportedly good descriptions [R. P. Abelson, personal communication, January 22, 1969].

The third and most important way of strengthening simulation arguments which suffer from the "degrees of freedom" problem, however, is to return to the original situation and demonstrate that the assumed isomorphism between the inputs of the original situation and the inputs of the simulation actually exists. This is what Harris and Tamler (1971a, 1971b) have done. This too, is what was done in an experiment by Bem and McConnell (1970) which demonstrated that

subjects in dissonance experiments cannot, in fact, recall their initial attitudes at the time of the final attitude assessment, that they see their postmanipulation attitudes as the same attitudes which motivated them to comply in the first place, and that they do not experience any attitude change phenomenologically. Hence, initial attitudes are not salient for involved subjects, and should thus not be made salient for observers in the simulations.

This process of moving back and forth between the simulation and the actual situation is precisely the one which cognitive theorists have attempted to follow and, in fact, it is this interaction between simulation and direct experimentation which constitutes the heuristic utility of the simulation methodology. A simulation reveals an underlying assumption or implication of the model which was not originally observed or even anticipated. The theorist can then return to the original situation armed with a new hypothesis. Thus, the hypothesis that subjects in dissonance experiments would perceive their postexperimental attitudes to be identical to the pre-experimental attitudes came to light through the debate between the author and his critics over the simulations. Even though the hypothesis was logically implied by the original analysis, it remained unarticulated until the "countersimulations" of R. A. Jones *et al.* (1968) raised it explicitly. It is important to note, however, that once one of the isomorphisms is questioned, the issue can be resolved only by returning to the original situation. A simulation will not suffice: "No 'as if' methodology, including the technique of interpersonal simulation, is an adequate substitute for the intensive study of the actual situation being modeled [Bem, 1968c, p. 273]."

E. The Possibility of a Crucial Experiment

At the end of their experiment, Bem and McConnell (1970) state that "If the past history of controversies like this is any guide, it seems unlikely that a 'crucial' experiment for discriminating between [dissonance theory and self-perception theory] will ever be executed. At this juncture each theory appears capable of claiming some territory not claimed by the other, and one's choice of theory in areas of overlap is diminishing to a matter of loyalty or aesthetics (p. 30)." Admittedly, this was partially a ploy to bring the whole thing to a halt. But it didn't work, for the controversy over the salience of initial attitudes itself suggested a possible "crucial" test to Snyder and Ebbesen.

In the typical forced-compliance experiment, the subject's initial attitude is in conflict with the induced behavior. What should happen, Snyder and Ebbesen wondered, if the initial attitude is made more salient to the subject? Self-perception theory predicts that this will

diminish the degree to which the final attitude attribution will be based upon the induced behavior and hence will diminish the amount of attitude change observed. Dissonance theory, however, predicts just the opposite: Making the initial attitude salient makes one of the two "dissonant" cognitions salient, thus increasing the amount of dissonance aroused. This in turn should produce *more* attitude change. If, on the other hand, the behavior is made more salient, then both theories predict more attitude change. For self-perception theory, this would make salient the very source of evidence upon which the final attribution is to be based; for dissonance theory, this is simply making the second of two dissonant cognitions more salient, again arousing more dissonance. Finally, consider the possibility that both the initial attitude and the behavior were to be made more salient. Under the multiplicative rule for combining different sources of credibility information (Section III, A, *4*), self-perception theory predicts that this will be equivalent to making the initial attitude alone salient; hence, attitude change will diminish. Dissonance theory here predicts the greatest attitude change of all since both "dissonant" elements are made salient, thus maximizing the amount of dissonance aroused.

With these differential predictions in hand, Snyder and Ebbesen attempted the crucial test by duplicating exactly the experiment by Bem and McConnell (1970), a standard forced-compliance experiment in which subjects were given a choice or no choice to write a counter-attitudinal essay. First they verified that they could replicate the standard effect as Bem and McConnell had done: Subjects given their choice expressed final attitudes more in line with their counter-attitudinal essays than those expressed either by no-choice subjects or control subjects who wrote no essays at all.

When initial attitudes alone were made salient, choice subjects did not differ from no-choice subjects, and neither group differed from the no-essay control group. Considered alone then, this column of results supported self-perception theory and runs counter to the dissonance theory prediction. When the behavior alone was made salient—where both theories predict enhanced attitude change—the column of results looked like the standard condition: There was more attitude change in the choice than in the no-choice condition which, in turn, did not differ from the control condition, but the effect was not significantly larger than in the standard condition. Neither theory is supported strongly by this condition, but neither was this a condition designed to discriminate between the two theories. Finally, no attitude change was observed in either condition when both the attitude and the behavior were made salient. Again, this supports self-perception theory (multiplicative rule) and not dissonance theory.

A contrast applied to the entire pattern of results shows that self-perception theory accounts for more of the between-cells variance than dissonance theory. But on the other hand, both theories leave a significant amount of the variance unexplained. In particular, there were differential amounts of attitude change among the several no-choice conditions which neither theory anticipates. Snyder and Ebbesen propose a self-perception theory of their own to account for the pattern of results, a theory which proposes that the subject's perception of how counter-attitudinal his behavior appeared to him mediates the attitude change effects. Whatever the merits of the Snyder-Ebbesen theory turn out to be, the "crucial" test between dissonance theory and the original self-perception interpretation of forced-compliance effects remains equivocal. Nevertheless, contrary to the prediction of Bem and McConnell (1970), somebody did manage to force the two theories to confront each other.

It is particularly unfortunate that the Snyder-Ebbesen experiment is equivocal, for the remaining Zeigarnik is likely to tempt others to go forth and do likewise. There are many more interesting paths off the self-attribution road to be explored, and the demise of the precious controversy between dissonance theory and self-perception theory is a consummation devoutly to be wished. It is resolved that, henceforth, nothing more on this subject shall be heard from this quarter.

IV. Other Self-Perception Phenomena

In addition to the cognitive dissonance experiments and the initial studies specifically designed to provide support for self-perception theory, there are a number of other effects which have begun to emerge which fit more or less into the self-perception explanatory framework. Three of these will be reviewed here.

A. Misattribution Effects

Before self-perception theory had been enunciated, Schachter's (1964) work on emotional states was already providing evidence that a person's attribution of his internal states was a joint function of both internal physiological cues and external factors of the situation (e.g., Schachter & Singer, 1962). It follows from this fact that one can manipulate an individual's self-attributions by manipulating those external factors.

Valins (1966) took this possibility a step further when he showed that an individual's self-attributions could also be influenced by giving him false feedback about his autonomic arousal. Thus, as noted earlier, Valins was able to manipulate attitudes toward stimulus pictures of

seminude females by giving his male subjects false auditory feedback which they could interpret as their heartbeat. Since then, a number of studies have adopted this general procedure. For example, Valins and Ray (1967) asked snake-phobic subjects to look at slides picturing snakes; subjects were given false feedback about their heartbeat designed to imply that they were not afraid of snakes. Subsequently, these subjects were able to approach the snake more closely than control subjects.

Berkowitz, Lepinski, and Angulo (1969) and Berkowitz and Turner (1972) have showed that they could influence the amount of instrumental aggression displayed by subjects by first giving those subjects false feedback about how angry they were. And finally, extinction of the GSR can be facilitated (Loftis & Ross, 1971) or retarded (Koenig & Henriksen, 1972) by giving subjects false information about their arousal. Although there are some intracacies in these studies to be discussed later, the general point to be made at this juncture is that false feedback concerning arousal can lead individuals to misattribute emotional states to themselves.

Misattribution can also be created by manipulating not only the apparent degree of arousal, but the apparent sources or causes of arousal (cf. Nisbett & Valins, 1971). For example, Nisbett and Schachter (1966) were able to obtain higher shock tolerance from subjects who believed that their arousal was produced by a pill rather than the shock. Using a very similar technique Ross, Rodin, and Zimbardo (1969) reduced fear of electric shock by persuading their subjects to attribute their arousal to loud noise. And finally, Storms and Nisbett (1970) gave insomniacs placebo pills and told them that the pills would produce either arousal or relaxation. As predicted, "arousal" subjects reported that they got to sleep more quickly than they had on nights without the pills—presumably because they attributed their arousal to the pills and, as a consequence, worried less about their insomnia, a worry which seems to exacerbate insomnia. Similarly, "relaxed" subjects reported that they got to sleep less quickly than usual—presumably because they worried that their emotions were unusually intense since their arousal level was high even after taking an arousal-reducing agent.

As Nisbett and Valins (1971) point out, all such studies can be seen as special cases of the underlying assumptions of self-perception theory even though the source of cues for the self-attributions are not the individual's overt behavior *per se*. But even this gap was closed in a study by Davison and Valins (1969), who actually manipulated the subjects' behavior and its apparent controlling variables to produce misattribution. In their study, Davison and Valins asked subjects to

take a series of electric shocks of steadily increasing intensity and told them to report when the shocks first became painful and when they could no longer tolerate them. Following this series, subjects were given a placebo which they were told might change their skin sensitivity. A second shock series was then administered, but the intensity of all shocks was surrepititiously halved so that subjects ended up taking nearly double the number of shocks than in the first series before reporting pain or asking the experimenter to terminate the series. All subjects were then told that the experiment was over. Some subjects were merely thanked and told that in a few minutes, after the drug had worn off, they would participate in another experiment. Other subjects were told that the drug was really a placebo. Davison and Valins reasoned that the placebo subjects would thus have to attribute their high-shock tolerance on the second series to themselves, their actual ability to withstand shock, since they now knew the drug could not have been responsible, whereas the subjects who had not been "debriefed" would make no such attribution of ability to themselves. This hypothesis was confirmed when a third shock series was administered as part of a "different" experiment. Placebo subjects took significantly more shock on the third series than they had on the first, whereas subjects who continued to believe that their performance on the second series was due to the drug did not.

Bowers (1971) conducted an experiment conceptually quite similar to the Davison-Valins experiment, except that he altered picture preference rather than shock tolerances through misattribution of behavior. His subjects were shown pairs of postcard pictures and asked to state their preference for one of the two pictures in each pair. Each picture also had a four-digit code number printed on its face, and each paired comparison required the subject to choose between a landscape and a portrait. Immediately prior to the series of trials, subjects in two of the experimental conditions were hypnotized and told they should always select pictures whose code numbers contained the digit 7. They were then given amnesia for the suggestion and for the fact that they had been hypnotized. The series of paired comparisons was then begun.

As in the Davison-Valins experiment, three series of trials were given. The first 20 trials determined the subject's baseline for preferring either portraits or landscapes. None of the pictures in these trials had a code number containing the digit 7. During the next 90 (treatment) trials, 60 of the pictures contained the digit 7 paired with the type of picture (portrait or landscape) the subject had least preferred in the baseline trials. Thus, all previously hypnotized subjects saw themselves choosing either portraits or landscapes for at least two-thirds of the

treatment trials with no explanation for their behavior. Bowers reasoned that these subjects would be led to infer that they must, in fact, now prefer their previously nonpreferred type of picture. But for half of these subjects, obvious explicit verbal reinforcement was administered during the treatment trials each time the subject expressed his preference for the nonpreferred type of picture. This, according to Bowers, should provide a suitable explanation to the subject for his preference change and thus prevent the self-attribution from taking place. Several types of control groups who had not been hypnotized were also run.

The test of this self-attribution hypothesis was sought in the subjects' preferences on a final set of 40 trials in which none of the pictures contained the digit 7. As predicted, subjects whose behavior had been manipulated through the hypnotic suggestion and who were left with no alternative explanation for their preference change persisted in choosing the initially nonpreferred type of picture significantly more than the reinforced group or any of several control groups.

It is clear why these various phenomena have earned the label of misattribution effects. On the other hand, it is easily overlooked that every effect discussed so far in this review is, without exception, a "misattribution" effect. If individuals actually accurately discriminated the variables controlling their behavior, then none of the predicted self-attributions would have occurred. For example, as Kelley (1967) has pointed out about forced-compliance studies, subjects in low-justification conditions must have an "illusion of freedom," must fail to apprehend the forces which induced them to comply, if they are to draw the predicted self-attributions from the behavior in which they engage. [The perception of freedom has itself now become a major research topic. For a review and analysis, see Steiner (1970).]

Self-perception theory may thus appear deficient because it does not attempt to account for this pervasive unawareness of the actual controlling variables. Or, as a Stanford colleague is fond of saying, self-perception theory appears to explain everything about why induced compliance leads to attitude change except why induced compliance leads to attitude change. But this apparent deficiency emerges only if one assumes that awareness is the normal state of things, and that unawareness is the phenomenon to be explained. As noted in Section I, A, the unique contribution of the radical-behavioral analysis of self-referring statements resides precisely in its explicit assumption that unawareness is the given and awareness the problematic. From this perspective, what needs to be explained—and what Skinner's analysis explains—is how individuals learn to respond to the apparent controlling variables we have purposely made salient in the laboratory, not why

they fail to discriminate the actual controlling variables we have intentionally obscured. Under such conditions, the radical behaviorist's answer to the question, "Why do subjects have this 'illusion' of freedom?" is "Why not?"

It is consistent with Western man's fascination with the unconscious that misattribution seems somehow "sexier" and more mysterious than veridical attribution. There are, of course, problematic cases of unawareness involving repression and motivated distortion. But these are challenge enough without adding the extra theoretical burden of having to explain the pseudo-problem of why individuals are not also perfect information processers.

B. The Self-Attribution of Dispositional Properties

All of the studies reviewed so far in this article demonstrate that external sources of stimuli can exercise control over an individual's attributions of his transitory states or his attitudes. Some studies are beginning to be reported, however, which suggest that it might be possible to change longer-standing attributions that the individual might make about himself by manipulating his behavior and apparent controlling variables appropriately. Interestingly, the first real clue that this might be possible was discovered almost accidentally by Freedman and Fraser (1966), who were investigating the so called "foot-in-the-door" phenomenon in which a person who can be induced to comply with an initial small request is then more likely to comply subsequently with a larger and more substantial demand.

In their study, Freedman and Fraser had two undergraduate experimenters contact suburban housewives in their homes, first with a small request and later with a larger more consequential request. The housewives were first asked either to place a small sign in their window or to sign a petition on the issue of either safe driving or keeping California beautiful. Two weeks later, a second experimenter returned to each home, asking all subjects to place a large and rather unattractive billboard promoting auto safety on their front lawn for several weeks. The second request thus involved both an action and an issue which were either similar to or different from those involved in the first request. A control group was contacted only about the second request.

The results showed a very strong foot-in-the-door effect. Subjects who had complied with the earlier trivial request were much more likely to comply with the larger one 2 weeks later. The remarkable finding, however, was the striking generality of the effect. It did not matter which issue or action had been involved in the initial request; the compliance generalized equally in all four conditions to the later

larger request. Thus, even the small action of signing an innocuous petition to "Keep California Beautiful" increased the subsequent probability of the subject's agreeing to place a large billboard reading "Drive Safely" on her lawn when asked to do so 2 weeks later by a second unrelated experimenter. In fact, the failure to find a generalization gradient as a function of similarity between the initial and final requests tends to rule out several plausible theoretical explanations of the effect. Thus, Freedman and Fraser (1966) arrive *post hoc* at what is essentially a self-perception explanation:

> What may occur is a change in the person's feelings about getting involved or about taking action. Once he has agreed to a request, his attitude may change. He may become, in his own eyes, the kind of person who does this sort of thing, who agrees to requests made by strangers, who takes action on things he believes in, who cooperates with good causes [p. 201].

In thinking about the Freedman-Fraser study and self-perception theory, it occurred to Lepper (1971) that attributions of this kind might also be taking place in other experimental settings which had previously measured attitudinal attributions only. For example, in the forbidden toy paradigm it is found that children who comply under mild threat conditions devalue the forbidden toy. But the attribution, "I don't like that toy" is only one possible inference that a child in the mild-threat condition might draw. Lepper suggests that another inference might be that he is a particularly "good" boy, one who is able to resist temptation. Moreover, this is not an inference which would be drawn under conditions of severe threat. Such an attribution might then generalize so that the child would display increased resistance to temptation in a different situation, an exact analogue to the Freedman-Fraser finding.

Lepper tested this hypothesis in the classical forbidden toy setting. Two groups of second-grade children were forbidden from playing with an attractive toy under mild or severe threat of punishment, while a third, control group received no initial prohibition. Three weeks later, a second experimenter asked these subjects to play a game in which they could obtain attractive prizes only by falsifying their scores. As predicted from the self-perception analysis, subjects who complied with the initial prohibition under mild threat showed more resistance to temptation in this second situation than control subjects or subjects who had initially complied under severe threat.

Lepper also reports that subjects who had initially complied under severe threat tended to show even less resistance to temptation than control subjects. Although this finding is open to more than one interpretation, Lepper suggests that it might be an "overjustification" effect, wherein the child under threat of severe punishment infers that

he resists temptation only because of strong external forces. Hence, when such pressures are subsequently withdrawn, he is even less resistant to temptation than before. We turn now to a more detailed consideration of such overjustification effects.

C. Overjustification Effects

The self-perception analysis of insufficient justification essentially states that a person will infer that he was intrinsically motivated to execute the induced behavior to the extent that external contingencies of reinforcement appeared to be absent. Thus he infers that he "wanted" to do the activity, that he believes in it, or that it reflects his true opinions. An overjustification effect is predicted if one is willing to assume that to the extent that external contingencies of reinforcement are strongly apparent, the individual infers that he did not want to perform the activity, that he does not believe in it, or that it does not reflect his true opinions. Because behavior is "consonant" with the initial attitudes in most overjustification studies, dissonance theory does not apply, and some writers (e.g., Nisbett & Valins, 1972) have looked to such effects as clearer instances of self-perception phenomena.

If overjustification effects do occur, then they provide a possible affirmative answer to the old question of whether or not extrinsic reinforcement for an activity reduces the intrinsic motivation to engage in that activity. Performing the activity under strong contingencies of reinforcement leads to the attribution that the activity must not be enjoyable in itself and then perhaps to decreased motivation to engage in that activity (deCharms, 1968; Deci, 1971, 1972; Deci & Cascio, 1971; Lepper, Greene, & Nisbett, 1971). For a number of reasons, early experiments with monkeys do not provide an affirmative answer to this question, discussions in elementary textbooks, and elsewhere (e.g., deCharms, 1968) notwithstanding. Accordingly, new attempts to confirm this overjustification effect have now begun to appear. For example, a series of studies by Deci (1971, 1972; Deci & Cascio, 1971) does suggest that intrinsic motivation to solve puzzles is reduced when the activity is executed under either monetary reinforcement or, possibly, threat of punishment, but not if positive verbal feedback is the reinforcer. It is clear that the "meaning" of the individual's self-observed behavior is going to be a function of his past history with the particular reinforcement contingencies used. Thus, money has probably acquired the discriminative property of "buying" compliance more than verbal praise.

The definitive experiment in this area, however, appears to be an elegant study by Lepper *et al.* (1971), who carefully divorced the measurement of intrinsic interest from the situation in which the rewards

were administered in order to rule out alternative interpretations of the anticipated results. Children who met a criterion of intrinsic interest in a play activity during baseline observations in their classrooms were randomly assigned to one of three conditions. In separate individual sessions 2 weeks later, children in the Expected-Award condition engaged in the activity in the anticipation of receiving an extrinsic reward. Children in the Unexpected-Award condition engaged in the activity only for its own sake, but subsequently received the same reward. The No-Award control subjects neither expected nor received the reward, but otherwise duplicated the experience of the subjects in the other conditions. One to two weeks after the experimental sessions, the target activity was again introduced into the children's classrooms, and unobtrusive measures of intrinsic interest were again obtained. The results confirmed that children in the Expected-Reward condition now freely engaged in the activity less than did either children in the Unexpected-Award or No-Award conditions.

Although these effects have here been interpreted in terms of self-perception theory, there are some problems involved in doing so which will be discussed in Section VII where the role of other explanations (e.g., deCharms, 1968) will also be considered.

V. Some Differences between Self-Perception and Interpersonal Perception

Self-perception theory asserts that self- and interpersonal perception are similar in two ways. First, the processes of inference involved in attribution are the same, and second, both actors and observers share certain sources of evidence—overt behavior and its apparent controlling variables—upon which those attributions can be based. This leaves open at least four ways in which the self-attributions and interpersonal attributions can still differ.

The first difference is what might be called the *Insider vs. Outsider* difference. All of us have approximately 3–4 ft^3 of potential stimuli inside of us which are unavailable to others but which are available to us for self-attributions. The thrust of the Skinnerian analysis of self-attributions is not that we can make no discriminations among internal stimuli, but only that we are far more severely limited than we suppose in this regard because the verbal community is limited in how extensively it can train us to make such discriminations. Nevertheless, the Insider can often detect, for example, that he is "trying hard" to solve a problem, and can infer that the problem is difficult, whereas an Out-

sider lacking such internal information, might infer laziness and suppose the problem to be easy.

A closely related difference is the *Intimate vs. Stranger* distinction. Here it is our knowledge of our past behavior which guides our attributions, whereas others typically lack such historical information. If past experience has convinced the Intimate that he is intellectually capable, then he will dismiss an experimental task as unfair, irrelevant, or both when he fails it, but as fair and pertinent when he succeeds. The Stranger, however, might well infer that the individual is stupid if he fails but capable if he succeeds. The difference between Intimate and Stranger is that the Stranger does not have any past performance upon which to "anchor" a dispositional attribution, and hence, he is more likely to permit task performance to determine a dispositional attribution than is the Intimate for whom the present task is but a single datum point in a familiar history of intellectual competence. The individual himself has already achieved a relatively stable dispositional inference about his ability, and hence fluctuations in performance can more plausibly be attributed to the task. [This is essentially a partial restatement of Kelley's (1967) "analysis of variance" model for attributions.]

A third difference between self-perception and interpersonal perception stems from the *Self vs. Other* difference. It is here that motivational effects may enter as the Self seeks to protect his esteem or defend against threat. Presumably the several Freudian defense mechanisms are prototypic of processes which distort self-attributions so that they differ from interpersonal attributions. On the other hand, the jump to motivational explanations is probably made too hastily. For example, when subjects and observers differ in evaluating intellectual competence in success and failure conditions, it is tempting to infer that the subject is defensively trying to maintain his self-esteem. But he may simply be veridical. As the example in the above paragraph illustrates, such attribution differences can be frequently explained by looking at the Self's knowledge of his past history. In other words, the motivational *Self vs. Other* difference is probably too often invoked when it is the *Intimate vs. Stranger* difference which is operative.

Moreover, the evidence for esteem-maintenance processes in self-attribution is not nearly so strong as is often supposed. For example, Ross, Bierbrauer, and Polly (1971) conducted a study in which professional teachers and college students attempted to teach an 11-year-old boy the spelling of a list of commonly misspelled words. Contrary to theories of self-esteem maintenance or ego-defensiveness, participants tended to rate "teacher factors" as being more important when the child failed than when he succeeded, and "student factors" more important

in success than in failure conditions. Moreover, this pattern of attribution was considerably more pronounced for professional teachers than for college students, a result which seems to challenge the frequent assertion that esteem maintenance and ego-defensiveness become factors when outcomes are important or central to one's self concept.

This is not to say that motivational distortions do not occur in self-attribution. But Self should be innocent until proved guilty.

There is, finally, the possibility that there exists an actual difference in perspective between *Actor vs. Observer*, in which different features of the situation are differentially salient to them. In an excellent article which seems likely to become highly influential, E. E. Jones and Nisbett (1972) suggest that an actor's attention is focused outward toward situational cues rather than inward on his own behavior. For the observer, however, the actor's behavior is the figural stimulus against the ground of the situation. E. E. Jones and Nisbett (1972) present a well-reasoned argument for the existence of this perspective difference and some preliminary findings relevant to it. It is clear, however, that some ingenuity will be required to isolate a pure perspective difference empirically, uncontaminated by the other differences discussed above. At the time of this writing, there appears to be only one study which comes close to accomplishing this feat (Storms, 1971).

The major thesis of the Jones-Nisbett article is that the several differences mentioned above, including the perspective difference, conspire to create a pervasive tendency for actors to attribute their actions to situational requirements, whereas observers tend to attribute the same actions to stable personal dispositions of the actor. It is, of course, too early to evaluate the validity of this proposition; but it is sufficiently rich in its implications, and it is so likely that the various actor-observer differences will pull in opposite directions in some situations, that the full exploration of this single hypothesis is likely to set the direction of research in this area for the next few years.

VI. The Shift of Paradigm in Social Psychology

During the Sixties, it will be recalled, all thinking beings were characterized by chronic drives toward consistency and uncertainty reduction, vigilant forces which coaxed us all toward cognitive quiescence. Our affects, cognitions, and behaviors were held in homeostatic harmony, and our "evaluative needs" initiated emergency information searches whenever any internal state broke through threshold without clear identification or certified cause. In contrast, we are emerging into the Seventies as less driven, more contemplative creatures, thoughtful men and women whose only motivation is the willingness to answer

the question, "How do you feel?" as honestly and as carefully as possible after calmly surveying the available internal and external evidence.

There is, in short, a shift of paradigm taking place within social psychology, a shift from motivational/drive models of cognitions, behaviors, and internal states to information processing/attribution models of such phenomena. Self-perception theory is only one element in that shift, and thus it is appropriate at this juncture to place it within this larger context. We begin with a brief history of this transformation and the four separate lines of research which mark it.

First, of course, are the various cognitive consistency theories, whose formal history is usually traced to Heider's (1946) article, "Attitudes and Cognitive Organizations" (McGuire, 1966). The system Heider proposed employed the motivational constructs of Gestalt psychology, and he elaborated the theory in the well-known *The Psychology of Interpersonal Relations* (Heider, 1958), and explored the motivational aspects in a 1960 contribution to the *Nebraska Symposium on Motivation* (Heider, 1960).

During the decade of the Fifties, several other consistency formulations were developed, and collectively they set the dominant tone of the motivation/drive paradigm during the early Sixties. This era appeared to culminate with the publication of the massive source book of such theories in 1968 (Abelson *et al.*, 1968). Cognitive dissonance theory is the most prominent example of this paradigm.

The second research tradition involved in the paradigm shift is Schachter's (1964) work on the cognitive and physiological foundations of emotional states. Although Schachter's theorizing has not been associated with the cognitive consistency paradigm as such, it is rooted in the same tradition, and the major motivational concept "evaluative needs" rests upon Festinger's (1954) earlier theory of social comparison processes. This is the motivation which leads individuals to seek out an appropriate explanation and label for otherwise ambiguous internal states.

It is illustrative of the paradigm shift that this prominent motivational feature of Schachter's (1959) earlier work on affiliation and the initial research on emotional states (e.g., Schachter & Singer, 1962) has now receded very much into the background, and investigators who trained under Schachter tend increasingly to employ the vocabulary of self-attribution in their studies.

Some convergence between the cognitive consistency research and Schachter's work was achieved when dissonance-theoretic operations began to be applied to the manipulation of emotional and motivational states (e.g., Zimbardo, 1969).

The third relevant conceptual development has been self-perception

theory. As noted in this review, the theory derives from the very different tradition of radical behaviorism. However, as the paradigm shift has developed, the Skinnerian parentage of the theory has been increasingly muted in successive translations. Indeed, one purpose of this review has been to repay homage to the origins of self-perception theory and, in Section VII, C, below, to restate the need for some of the heuristic advantages that the stubborn functional orientation can bestow when behavioral mysteries threaten to become behavioral-science muddles.

The ways in which self-perception theory has attempted to assimilate both the Schachter tradition and cognitive dissonance theory have already been reviewed in detail. It should be noted that self-perception theory lacks any motivational construct other than an implicit assumption that individuals are willing to answer inquiries concerning their internal states.

The fourth major development in the move to the attribution paradigm is attribution theory itself. Once again history begins with Heider, who stated the major ideas in *The Psychology of Interpersonal Relations* (Heider, 1958). During the "consistency" era, this book was cited primarily for its formal balance theory, while Heider's rich but less formalized observations about person perception and attribution were relatively ignored. This was remedied by E. E. Jones and Davis (1965), who added a number of testable propositions and explicated some specific empirical consequences of the attribution hypotheses contained within the book. The resulting research tended to focus on an observer's attribution of an actor's intentions and attitudes (e.g., E. E. Jones & Harris, 1967) and would probably have proceeded independently of the other three traditions discussed above had it not been for the influential essay, "Attribution Theory in Social Psychology" by Harold Kelley in the 1967 *Nebraska Symposium on Motivation.* This essay integrated the Jones and Davis formulation and self-perception theory into a single theoretical framework along with some propositions about attributional biases, errors, and illusions. These latter considerations also afforded Kelley the opportunity to make his observations about the "illusion of freedom" found in dissonance experiments, thus providing an added flourish to the convergence of these several distinct lines of research and theory.

If Kelley can be seen as a final step in this shift from drive models to information-processing models, as this brief intellectual history implies, then it is somewhat ironic that his essay appears in a symposium on motivation. For despite Kelley's valiant try, the motivational flavor is very bland indeed:

Consideration of attribution theory is relevant for a symposium on motivation in several respects. The theory describes processes that operate *as if* the individual were motivated to attain a cognitive mastery of the causal structure of his environment. Indeed, Heider explicitly assumes that "we try to make sense out of the manifold of proximal stimuli . . . " And Jones and Davis state, "The perceiver seeks to find sufficient reason why the person acted and why the act took on a particular form." The broad motivational assumption makes little difference in the development and application of the theory, but it gives the theory a definite functionalistic flavor . . . and affords whatever motivational basis might seem necessary to support the complex cognitive processes entailed in attribution.

More important for the student of motivation are the specific processes and their consequences. Attribution processes are assumed to instigate, under certain conditions, such activities as information-seeking, communication, and persuasion. Thus attribution theory plays an important role in describing the motivational conditions for these significant classes of social behavior. Equally important is the relevance of attribution theory to the *perception* of motivation, both in others and in one's self [From Kelley, 1967, p. 193, by permission of The University of Nebraska, Lincoln, Nebraska]. [Emphasis in the original.]

It is an admirable attempt, but the strongest motivation to emerge from this quotation appears to be Kelley's need to understand why he was there. Presumably his fellow participants were thus provided with "sufficient reason why the person acted and why the act took on a particular form."

It is, finally, Kelley's article which has now set the stage for the analysis of the differences between self- and interpersonal perception, discussed in the previous section, and this appears to be the next phase of research in this area.

This, then, is where the paradigm shift currently stands and where it appears to be going. But there are still some sticky problems left in what has gone before, as we shall now see.

VII. Some Unsolved Problems

Up to this point, this review has attempted to fit as many phenomena as possible into a single framework with a minimum number of loose threads. This pedagogically motivated elegance, however, must now come to an end, for it has been purchased at the price of some fairly glib legerdemain. It is time to sneak backstage and see what the performance looks like from the wings.

A. The Conceptual Status of Noncognitive Response Classes

If one has managed to alter an individual's attitude or self-attribution, it is not unreasonable to expect that this will induce consequent changes in other response systems. For example, if one has increased a

person's favorability toward a dull task, he might be expected to work at the task more assiduously. Induce him to believe that he doesn't fear snakes and he will approach snakes more closely. Convince the child he is obedient, and he will resist temptation. Change a man's perception of his hunger, thirst, or pain, and he should consume more or less food or drink, or endure more or less aversive stimulation. Nor should such expectations be confined to instrumental or consummatory behaviors only, for there is a long history of evidence that beliefs, attitudes, and self-attributions can exercise influence over physiological responses as well (for reviews, see Frank, 1961; Zimbardo, 1969). It is therefore not unreasonable to expect physiological changes to follow upon induced self-attributions of internal states.

Happily, the experimental laboratory has blessed such expectations with some striking confirmations. Dissonance manipulations designed to enhance the perceived attractiveness of dull tasks do produce greater intensity of behavior on the task itself (for a review, see Weick, 1967). Behavioral observations of subjects in Schachter's experiments reveals them to be behaving "appropriately" in accord with their induced emotional states (e.g., Schachter & Singer, 1962; Schachter & Wheeler, 1962); and, dissonance manipulations designed to alter self-attributions of drive states like hunger, thirst, and pain do alter overt behaviors with respect to their respective stimuli and do produce striking physiological changes (Zimbardo, 1969).

In the present review, we have seen that false feedback designed to imply that subjects are not snake-phobic leads them to approach the snake more closely (Valins & Ray, 1967); feedback designed to manipulate self-attributions of anger produces changes in overt instrumental aggression (Berkowitz & Turner, 1972); and feedback designed to create misattributions concerning autonomic arousal alters resistance to extinction of the GSR (Koenig & Henriksen, 1972; Loftis & Ross, 1971). Attempts to encourage subjects to attribute fear-induced arousal to a pill or loud noise rather than shock produce greater shock tolerance (Nisbett & Schachter, 1966; Ross *et al.*, 1969), as does self-observed behavior designed to imply that the subject has higher shock tolerance than he initially thought (Davison & Valins, 1969). Induced obedience which implies to the child that he is "good" produces greater resistance to temptation (Lepper, 1971); and finally, overjustification for performing intrinsically interesting activities diminishes subsequent engagement in those activities (Deci, 1971, 1972; Lepper *et al.*, 1971).

But precisely because it has been "not unreasonable to expect" these phenomena to occur, and precisely because they have in fact occurred, the problematic nature of their conceptual status within the

various theories has been insufficiently appreciated. Thus, the "theoretical" predictions or explanations of these phenomena that one finds in the literature are rarely more sophisticated than the "it-is-not-unreasonable-to-expect . . ." statement with which this section opened three paragraphs above. The lucky theory within which the particular investigator is working then gets gratuitous credit for the "derivation." A related practice, also encouraged by the fact that the response classes seem to "hang together," is to treat the response classes interchangeably as if they were functionally equivalent; the self-attributions and the noncognitive responses are simply grouped together as the "effects" of the stimulus manipulations. Such practices are unfortunate for they can easily obscure important gaps in our understanding by causing us to pretend to knowledge that we do not in fact possess. It is thus important to explore how the various theories account for these noncognitive response classes.

In attribution models generally—and in self-perception theory in particular—cognitions or self-attributions are the dependent variables. Instrumental behaviors, consummatory responses, and physiological responses (real or falsified) are among the variables which can serve as antecedent or independent variables, the stimuli from which self-attributions of beliefs, attitudes, or internal states can be partially inferred by the individual. Attribution models are thus very explicit about the direction of the causal arrow, and they remain mute about any phenomenon in which the noncognitive response classes play the dependent variable role; as *dependent* variables, such response classes are extratheoretical. To state this another way, attribution models do not treat cognitions, overt behaviors, and physiological responses as functionally equivalent response classes, but rather, spell out in detail the mechanisms through which the cognitive response class can be under the partial functional control of the other two. How do attribution models account for noncognitive response classes? They don't! Self-perception theory can get us from the stimulus manipulation to the attribution. It cannot get us from the attribution to anything beyond that.

The consistency paradigm is in much the same position as the attribution paradigm with regard to the physiological response class. Thus an early prediction that physiological phenomena might emerge from dissonance-theory settings was precisely a speculation in the spirit of "it would not be unreasonable to expect . . . [Brehm & Cohen, 1962, pp. 151–155]." The positive empirical results which followed and confirmed that early hunch (Zimbardo, 1969) in no way altered the theoretical status of the hypothesis within the formal theory itself, nor does the invocation of dissonance reduction as a motivational explanatory

concept bridge the gap from the attribution changes to the physiological effects. It is, for example, no "explanation" to assert that an individual lowers his GSR in order to reduce dissonance until it is explained just how the individual goes about doing just that. Again, this gap is not to be confused with the prior gap from the stimulus operations to the attribution changes, a link with which the theory *is* prepared to deal. This is, of course, the same position in which self-perception theory finds itself; as noted above it, too, has a theory about the first link, but is reduced to handwaving about the second. Similarly, the effects of false feedback on GSR extinction reported by Koenig and Henriksen (1972) are *not* accounted for by any of the three theories they mention—modeling, Schachter's theory, or self-perception theory—and for the same reason: None of these theories contains the theoretical machinery for explaining physiological changes in a nontrivial way. For example, the "explanation" borrowed from Schachter's formulation, that "a state of arousal will be perceived as positive or negative depending upon the label which a person attaches to that state, and that he will then behave accordingly" (i.e., show higher resistance to extinction of the GSR) is, at best, a restatement of the data. It is in no sense an explanation of that second link. (See also Bem, 1972a.)

It is thus an important step forward simply to recognize that a detailed theoretical model is still needed to account for the cognitive control of physiological responses. One of the criteria for a successful theory in this domain will almost certainly be its ability to account simultaneously for the related physiological effects of placebo medication, hypnotically or cognitively induced anesthesia, and the associated phenomena of the "mind-body" problem. A start in this direction is provided by Zimbardo (1969, pp. 269–283) whose theoretical discussion at least outdistances the dissonance theory framework which guided the choice of stimulus manipulations.

When one turns from the physiological to the behavioral variables associated with self-attributions, the consistency paradigm appears to be on firmer ground. For example, although the theory of cognitive dissonance is, in literal terms, a theory about cognitions (like the attribution models), the concept of a general drive toward consistency extends itself more easily to instrumental and consummatory behaviors than to physiological responses. Thus, if an individual suffers inconsistency between something he believes and the cognition that he is not behaving in accord with that belief, a purely cognitive conflict, then it follows from the basic postulate of the consistency model that he can achieve drive reduction by altering either the belief or the behavior. The motivational construct within the theory provides a built-in "motor" force

behind a change in overt behavior.[3] If a dissonance manipulation makes an individual more favorable toward a dull task, a higher rate of performance on the task is a reasonably legitimate prediction from the theory. It is important to note that the behavior in this formulation is necessarily mediated by a prior belief, attitude, cognition, or attribution with which the behavior is brought into harmony.

A similar kind of consistency principle is also invoked to explain behavioral effects by many investigators who employ Schachter's theory of emotional states. As already noted above, Koenig and Henriksen (1972) remark that Schachter postulates that an individual "will then behave accordingly" after he has labeled his emotional state. Similarly, Berkowitz and Turner (1972) interpret Schachter as saying that an individual interprets his state and "then acts in a manner consistent with this interpretation." And although Berkowitz and Turner go beyond the Schachter formulation in their own analysis of the stimulus variables leading to instrumental aggression, they too come back to the same mechanism in order to get from the self-attribution of anger to the act of aggression: "Looked at from a larger perspective, the findings also provide yet another demonstration of the search for cognitive consistency. We want our actions to be in accord with our emotions, as we understand them . . ."

Interestingly, however, Schachter himself does not invoke such a principle of consistency in his own writings on the topic [including Schachter (1964), which other writers most frequently cite in this connection]. Rather, he treats self-attributions and overt behaviors as separate "indices" of the underlying "mood" he set about to produce; that is, Schachter's conceptual analysis treats the two response classes as functionally equivalent. Thus, from the very beginning, Schachter and his colleagues have routinely collected behavioral observations along with, or even in lieu of, self-report data of emotional state (e.g.,

[3] It will be noted that we thus grant legitimacy to a motivational concept for explaining cognitive and behavioral responses, but deny it legitimacy in accounting for physiological responses. The distinction, however, is not based upon response class membership *per se*, but upon the individual's ability to directly control the response in question. A motivational construct is still, at bottom, a way of saying that the individual "wants to" perform some response, even if unconsciously; being motivated is not sufficient, however, if he "doesn't know which string to pull." As recent work demonstrates, physiological responses can be brought under direct conscious control; presumably they then fall subject to motivational explanations in the same way that instrumental responses do. Nevertheless, we are still avoiding here the deeper epistemological problems concerning the explanatory legitimacy of motivational constructs generally. As a sometimes radical behaviorist, I am inclined to the view that their explanatory power is, in general, illusory.

Schachter & Singer, 1962; Schachter & Wheeler, 1962). Similarly, self-attributions of hunger appear in some of the Schachter obesity studies (e.g., Goldman, Jaffa, & Schachter, 1968), but the dependent variable has now become eating behavior *per se*, and the word "hunger" has faded quietly from view (e.g., Nisbett & Storms, 1972). Although several new conceptual distinctions have been introduced into this important research to keep abreast of the new findings, there has been no comparable conceptual distinction introduced to parallel or accompany the *sub rosa* shifts from one response class to another.

In sum, pure attribution models presume only to deal with the cognitive response class; additional machinery must be added if they are to deal with behavioral or physiological responses as *dependent* variables. Schachter's model, hovering somewhere between the information-processing/attribution paradigm and the motivational/drive paradigm, does not distinguish on the dependent variable side between the self-attributions and the "emotional" behavior. Just as the attribution models do, Schachter's model places physiological responses in the role of independent variables only; they are stimuli which partially determine the individual's perception of his emotional state. Finally, theories within the motivational/drive paradigm, particularly the theory of cognitive dissonance, cannot handle the physiological response class in anything other than a trivial fashion, but they do have a conceptual device for predicting or explaining any overt behavioral changes that are mediated by prior cognitions, attitudes, or attributions. We turn now to a closer examination of this proposed sequence of events from stimulus manipulation to attribution change to behavior change.

B. Do Attributions Mediate Behavior?

Increase a person's favorability toward a dull task, and he will work at it more assiduously. Make him think he is angry, and he will act more aggressively. Change his perception of hunger, thirst, or pain, and he should consume more or less food or drink, or endure more or less aversive stimulation. Alter the attribution, according to the theory, and "consistent" overt behavior will follow.

There seems to be only one snag: It appears not to be true. It is not that the behavioral effects sometimes fail to occur as predicted; that kind of negative evidence rarely embarrasses anyone. It is that they occur more easily, more strongly, more reliably, and more persuasively than the attribution changes that are, theoretically, supposed to be mediating them.

For example, in a well-controlled study by Grinker (1969) on eyelid conditioning, it was predicted that ". . . the dissonance aroused by

voluntary commitment to a painful stimulus will be reduced by lowering pain-avoidance motivation—that is, by perceiving the UCS to be less threatening or painful—and thus the conditioning level [will be less]." The study did obtain the predicted effects in conditioning, but "there were no significant differences between any groups on self-report measures of perceived pain, irritability, eye tearing, or apprehension, or on other questionnaire items designed to measure subjective response to the aversive aspects of the situation [Grinker, 1969, p. 132]." And in a closely related experiment, Zimbardo *et al.* (1969) were able to obtain predicted changes in learning performance, physiological measures, and pain perception. But the attributions of pain showed the weakest effects, and furthermore, the correlations between these cognitive attributions and the behavioral measures of learning they were supposed to be mediating were $-.01$ for one group and $+.11$ for the other.

Similarly, in the Davison and Valins (1969) shock study, experimental subjects were willing to take more shock than control subjects, as predicted, but they did not rate a set of sample shocks as any less painful than did controls. The snake study by Valins and Ray (1967) gives similar results: Experimental subjects were able to approach the snake more closely than control subjects, but they did not report themselves to be any less frightened of snakes than did the controls.

Finally, Weick (1967) reviewed all the studies designed to increase an individual's favorability toward a dull task. He found that increased effort on the task often occurred in the absence of the attitude change toward the task which was supposed to cause the increased effort. Weick concludes that

> initial cognitive enhancement of the task followed by increased effort simply does not occur often enough for us to be convinced that this is a reasonable explanation. Instead, it appears that the phenomenon in which we are interested may involve just the opposite sequence of events, namely behavioral change followed by occasional attempts to summarize the experience evaluatively.

What is one to make of such failures? One possible explanation for them is that the measures of attributions are not well designed or appropriate to the self-attribution which actually mediates the behavior. Another possibility is that subjects are hesitant to admit to some states like anger (Schachter & Singer, 1962). Although these methodological explanations may account for some of the negative findings, the same pattern of results—behavior changes in the absence of equally strong attribution changes—is found in some of the best designed and carefully executed experiments in the field (e.g., Grinker, 1969).

Another possibility is that the attributions do change as predicted and do mediate the behaviors, but that the attributions themselves are

unconscious (Brock & Grant, 1963; Zimbardo, 1969, p. 76). There have been arguments, of course, that inconsistency itself and the process of inconsistency reduction need not be represented in awareness (e.g., Tannenbaum, 1968), and the author himself has made a parallel claim that individuals need not be able to verbalize the cues they use in arriving at self-attributions (Bem, 1965, 1968b). But such claims can edge dangerously close to metaphysics, and the next retreat into invisibility—that one of the "dissonant" cognitions itself is unknown to the individual—should surely be resisted mightily until all other alternatives, save angels perhaps, have been eliminated. A related but more plausible explanation involving defensive denial processes has been proposed by Zimbardo (1969, pp. 269–273); his version of an unconscious cognition at least generates some empirical consequences, and there are some suggestive supporting data for such a process within the settings explored in that body of research.

A final lead is provided by Weick's suggestion, quoted above, that the attributions or attitudes may follow upon rather than precede the behaviors. This is, of course, the major postulate of self-perception theory and a phenomenon well known to dissonance theory. If this is, in fact, the sequence involved, then it would explain why the measured attribution changes are often less reliable and weaker than the behavioral changes since they are the third, rather than the second link in the chain, as originally assumed. For example, the self-reports of euphoria and anger in the classic Schachter and Singer experiment (1962) were obtained after the behavior had occurred and were, in fact, less reliable than the behavioral observations themselves.

A similar instance appears in the Berkowitz and Turner study (1972). The self-reports of anger were retrospective measures in which the subjects had to recall how angry they had been prior to engaging in the aggressive behavior. A study by Bem and McConnell (1970) would imply that such "recall" measures would be more highly correlated with the subject's current attribution (as altered by the intervening behavior) than they would be with the actual previous state he is attempting to recall. And in fact, the Berkowitz-Turner self-report data—designed to check the false feedback manipulation of anger—do appear to parallel the overt aggression displayed by the subjects more closely than they correspond to the meter readings themselves. Prophetically, Berkowitz (1968) himself has said elsewhere: "We generally assume as a matter of course that the human being acts as he does because of wants arising from his understanding of his environment. In some cases, however, this understanding may develop *after* stimuli have evoked the action so that the understanding justifies but has not caused the behavior [p. 308]."

Is, then, the Berkowitz-Turner study one of these cases of reverse sequence? Not necessarily, for it is still possible that the appropriate attributions were actually present prior to the behavior and did mediate it. All that can be said is that self-report measures collected after the behavior has occurred may not be a valid index of those attributions. The same holds true for the Schachter and Singer study (1962) and several other studies which have collected attribution data confounded by intervening behavior. This analysis, then, implies that some of the failures to find attribution changes may simply reflect the methodological practice of collecting the self-reports after other stimulus events, including overt behaviors, have intervened. But like the other explanations offered above, this cannot account for all the failures, for again some of the best studies are not subject to this criticism (e.g., Zimbardo *et al.*, 1969), and they still do not find attribution effects as strong as the behavioral ones they are supposed to be mediating.

If we are thus forced to the conclusion that, at least in some settings, attribution changes do not mediate the observed behavioral effects, then we find that a phenomenon which had been previously accounted for within the consistency paradigm has become "unsolved." That is, we are still left with the task of accounting for the behavior changes themselves. Several attempts to do so are already under way.

For example, Nisbett and Valins (1972) have proposed that the stimulus manipulations themselves may be insufficient to actually alter the attribution, but that they do cause the individual to question his current attribution sufficiently so that he "tests" the new attributional hypothesis by engaging in behavior to find out. As a result of engaging in the behavior, the hypothesis about his attribution may be confirmed, and he will accept the attribution as valid. Or the hypothesis may be disconfirmed, leading him to reject the attribution, and leaving the investigator with a set of results showing a behavioral effect and no attribution effect. Thus, false feedback implying that one is not afraid of snakes is not sufficient to create a stable attribution of "I'm not afraid," but it is sufficient to motivate a test of this possibility by approaching the snake. The process of handling the snake can then stabilize the new attribution via the self-perception process. This intriguing scenario is spelled out in greater detail in Nisbett and Valins (1972).

With regard to the task enhancement studies, Weick (1967) has suggested that the behavioral effects in these experiments might be accounted for by propositions drawn from frustration theory and cue-utilization theory; again the attribution effects—when they do occur—can be handled as postbehavior phenomena by either self-perception theory or dissonance theory.

It may be that still other cases of behavioral effects such as the overjustification phenomenon will be "re-solved" by using variations of motivational constructs like the need to be in control of one's self and environment (cf. deCharms, 1968; Zimbardo, 1969). And, as suggested earlier, the physiological effects should probably be split off and reunited with other physiological phenomena under cognitive influence rather than being grouped according to their independent-variable manipulations as dissonance or attribution phenomena.

It is clear that the door has now been opened for many mini-theories, for it is unlikely that any single process will account for the diverse phenomena which found themselves grouped together when consistency and attribution models converged. It is, of course, painful to have to deny to self-perception theory some of the effects with which it has been gratuitously credited by other investigators. At least its heuristic value for such phenomena remains intact even if its explanatory power is more limited than its friends had realized. Similarly, it may seem a shame to abandon the parsimony which obtained during the reign of the consistency theories, but it is now clear that some of that parsimony was illusory and was purchased at the cost of obscuring some important gaps in our knowledge. The fact that everything seems to be falling apart should probably be taken as an index of scientific advance.

C. The Strategy of Functional Analysis

If there is an underlying moral here, it is that response classes should be given independent conceptual statuses from one another and analyzed separately for the stimulus variables which control them. If they are observed to covary, one should first inspect the stimulus manipulations for overlapping functional properties which produce that covariation. Any theory which assumes that one response class should vary as a function of another ought to spell out in detail the mechanism of control. What are the stimulus properties of the response class which is presumed to exert functional control over the other? Finally, response classes should not be treated as functionally equivalent unless the theory explicitly dictates that they can be and/or experimentation vindicates the merger.

These prescriptions form a part of a more general strategy known as functional analysis, the strategy associated with the radical behaviorism within which self-perception theory was initially enunciated.

A functional analysis of a complex behavioral phenomenon proceeds by first inquiring into the ontogenetic origins of the observed dependent variables—treated as response classes in their own right rather than as reflections of underlying structures, processes, or internal states—and

then attempts to ascertain the controlling or independent stimulus variables of which those observable behaviors are a function.

In its purest form, such a strategy also assumes that the principles of behavior, that is, the most general orderly functional relations between stimuli and responses, are relatively simple and few in number. The complexity of human social behavior derives, it is assumed, from the complexity and variety of the environmental conditions under which these principles have been operative in the individual's past history. Thus, the radical behaviorist does not begin with the *a priori* expectation that he will discover new principles of behavior through the study of social behavior in its terminal complexity. Rather, he attempts to establish a complex behavioral phenomenon as a special case or a compound of previously substantiated functional relations discovered in the experimental analysis of simpler behaviors. The re-emergence from the analysis of the previously established functional relations becomes the vindication for the extrapolation into new domains and for the network of assumptions upon which the extrapolation rests. The spirit of the analysis, therefore, is frankly inductive, not only in its experimental execution, but in its formal presentation.

The radical behaviorist's usual insistence that the analysis also eschew any reference to internal physiological or conceptual processes, real or hypothetical is, of course, its most celebrated (and misunderstood) prescription. Whatever the heuristic value of such a restriction may be in other psychological areas, the tactic carries especial probative force in the analysis of self-perception precisely because the socializing community itself must necessarily train the individual's self-descriptive skills on the basis of observable stimuli and responses.[4]

Not all functional analyses comprise every one of these tactics, and orthodox radical behaviorists are not the only psychologists whose

[4] It is probably the relaxation of this restriction which has robbed latter-day self-perception theory of its radical behavioristic flavor. One does not remain a behaviorist in good standing with repeated references to "inferential processes" and hypothetical inner dialogues ("What must my attitude be if I am willing to behave . . ."). In order to reclaim membership, therefore, it should probably be said that such concessions to expositional clarity do not, in my view, add anything to the explanatory power of the theory; it remains formally equivalent to its earlier, albeit nearly incomprehensible, incarnation in the more rigid and arid vocabulary of radical behaviorism (Bem, 1964). But as this section is attempting to demonstrate, a choice of language is not without heuristic consequences. For private "thinking" purposes, functional analysis remains my preference; but for exposition purposes, English prose does not seem overly risky. Roman and arabic numeral systems are also formally equivalent, but performing long division in the former is reputed to be unwieldy.

analyses are informed by particular elements of the approach. For example, Leventhal (1970) has recently employed a similar strategy in analyzing the attitudinal and behavioral effects of fear-arousing communications, thereby bringing elegant order out of the chaotic and conflicting findings in this area. (In addition, Leventhal's analysis is another instance of a drive theory being replaced by an information processing orientation.) Berkowitz's (1965) analysis of aggression has a similar spirit and strategy behind it.

If it had been employed earlier, the functional approach would have led to very different kinds of analyses within the domain of self-attributions. For example, Schachter's (1964) own review of the literature on emotion reveals that physiological cues should have more functional control over the cognitive attributions than over "emotional" behaviors *per se*. Thus, sympathectomized animals continue to show emotional behaviors, and human "subjects with cervical lesions described themselves as acting emotional but not feeling emotional [p. 74]." These findings would seem to have implications for which response class is being used as the "index" of mood in the Schachter experiments. But as noted earlier, Schachter and his colleagues interchange the two response classes repeatedly and nowhere acknowledge a functional distinction between them.

A similar confounding of these same response classes appears in the insomnia experiment by Storms and Nisbett (1970), an experiment also conducted within the Schachter framework. As described in Section IV, A, insomniacs who thought the placebo pill would arouse them reported that they got to sleep more quickly than they had on nights without the pills, whereas subjects who thought the placebo to be a relaxant reported that they got to sleep less quickly than usual. The important point to note here is that the dependent variable is the subject's *report* of how much time had passed before he fell asleep. But when this experiment is cited, and even in the abstract of the article itself, it is reported that "arousal" subjects *got* to sleep more quickly and "relaxation" subjects *got* to sleep less quickly. And that's not the same thing! Estimates of time passage are themselves attributions which are subject to manipulation (cf. London & Monello, 1972). Perhaps the implied state of arousal is more interesting to introspect than the implied state of relaxation, making time appear to pass more quickly. If true, then "arousal" subjects would report getting to sleep more quickly even if it weren't the case. This alternative explanation is admittedly less plausible than the original, but the point to be made here is that the time it takes to fall asleep is a different response measure from the self-report of the time it takes to fall asleep, and both are subject to cognitive manipulations. Indeed, the therapeutic implica-

tions of this experiment could be quite different from those suggested by the authors unless we are prepared to assume that getting insomniacs to think they are falling asleep faster is the same as curing the insomnia.

For years, personality theorizing has been dominated by the "trait" assumption that there are pervasive cross-situational consistencies in an individual's behavior. After reviewing the literature, Mischel (1968) concludes that the empirical search for such consistencies or traits rarely generates a correlation above +.30, a finding of some disappointment if one's theory of human behavior anticipates +1.00. He, too, suggests a learning-theoretic functional analysis in which covariance of responses is sought in the overlap of situational conditions which evoke and maintain particular response classes. Under such a strategy, one constructs the consistencies from the ground up rather than assuming them *a priori,* and any increment over zero in the magnitude of the cross-situational correlations becomes a matter for some rejoicing (for discussion, see Bem, 1972b).

There is a parallel in social psychology. The decade of the consistency theories was dominated by the assumption that everything was glued together until proved otherwise (cf. Bem, 1970, p. 34). Since it is now proving otherwise, it is suggested that we try the opposite assumption that nothing is glued together until proved otherwise. It is a question of whether we should begin with expectations of +1.00 correlations or .00 correlations. The heuristic advantage of this strategy is not guaranteed, of course. But the difference in morale if +.30 correlations continue to come is in itself worth considering.

References

Abelson, R. P. Simulation of social behavior. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology.* Vol. 2 (2nd ed.). Reading, Mass.: Addison-Wesley, 1968.

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum, P. H. (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

Alexander, N. C., & Knight, G. W. Situated identities and social psychological experimentation. *Sociometry,* 1971, **34,** 65–82.

Aronson, E., & Carlsmith, J. M. Effect of severity of threat on the valuation of forbidden behavior. *Journal of Abnormal and Social Psychology,* 1963, **66,** 584–588.

Asch, S. E. *Social psychology.* Englewood Cliffs, N. J.: Prentice-Hall, 1952.

Bandler, R. J., Madaras, G. R., & Bem, D. J. Self-observation as a source of pain perception. *Journal of Personality and Social Psychology,* 1968, **9,** 205–209.

Barber, T. X. Toward a theory of pain: Relief of chronic pain by pre-frontal leucotomy, opiates, placebos, and hypnosis. *Psychological Bulletin,* 1959, **59,** 430–460.

Barber, T. X. The effects of hypnosis on pain. *Psychosomatic Medicine,* 1963, **25,** 303–333.

Beecher, H. K. *Measurement of subjective responses: Quantitative effects of drugs.* London and New York: Oxford University Press, 1959.

Beecher, H. K. Increased stress and effectiveness of placebos and "active" drugs. *Science,* 1960, **132,** 91–92.

Bem, D. J. *An experimental analysis of beliefs and attitudes.* (Doctoral dissertation, University of Michigan) Ann Arbor, Mich.: University Microfilms, 1964. No. 64–12,588.

Bem, D. J. An experimental analysis of self-persuasion. *Journal of Experimental Social Psychology,* 1965, **1,** 199–218.

Bem, D. J. Inducing belief in false confessions. *Journal of Personality and Social Psychology,* 1966, **3,** 707–710.

Bem, D. J. Reply to Judson Mills. *Psychological Review,* 1967, **74,** 536–537. (a)

Bem, D. J. Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review,* 1967, **74,** 183–200. (b)

Bem, D. J. Self-perception: The dependent variable of human performance. *Organizational Behavior and Human Performance,* 1967, **2,** 105–121. (c)

Bem, D. J. Attitudes as self-descriptions: Another look at the attitude-behavior link. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes.* New York: Academic Press, 1968. (a)

Bem, D. J. Dissonance reduction in the behaviorist. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968. Pp. 246–256. (b)

Bem, D. J. The epistemological status of interpersonal simulations: A reply to Jones, Linder, Kiesler, Zanna, and Brehm. *Journal of Experimental Social Psychology,* 1968, **4,** 270–274. (c)

Bem, D. J. *Beliefs, attitudes, and human affairs.* Monterey, Calif.: Brooks/Cole, 1970.

Bem, D. J. The cognitive alteration of feeling states: A discussion. In H. London & R. E. Nisbett (Eds.), *Cognitive alteration of feeling states.* Chicago: Aldine, 1972. (a)

Bem, D. J. Constructing cross-situational consistencies in behavior: Some thoughts on Alker's critique of Mischel. *Journal of Personality,* 1972, in press. (b)

Bem, D. J., & McConnell, H. K. Testing the self-perception explanation of dissonance phenomena: On the salience of premanipulation attitudes. *Journal of Personality and Social Psychology,* 1970, **14,** 23–31.

Berkowitz, L. The concept of aggressive drive: Some additional considerations. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 2. New York: Academic Press, 1965. Pp. 301–329.

Berkowitz, L. The motivational status of cognitive consistency theorizing. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968. Pp. 303–310.

Berkowitz, L., Lepinski, J., & Angulo, E. Awareness of own anger level and subsequent aggression. *Journal of Personality and Social Psychology,* 1969, **11,** 293–300.

Berkowitz, L., & Turner, C. Perceived anger level, instigating agent, and aggression. In H. London & R. E. Nisbett (Eds.), *Cognitive alteration of feeling states.* Chicago: Aldine, 1972.

Bowers, K. S. An attributional analysis of operant conditioning: The problem of behavioral persistence. Unpublished manuscript, University of Waterloo, 1971.

Brehm, J. W., & Cohen, A. R. Re-evaluation of choice alternatives as a function of their number and qualitative similarity. *Journal of Abnormal and Social Psychology,* 1959, **58,** 373–378.

Brehm, J. W., & Cohen, A. R. *Explorations in cognitive dissonance.* New York: Wiley, 1962.

Brock, T. C., & Grant, L. D. Dissonance, awareness, and motivation. *Journal of Abnormal and Social Psychology,* 1963, **67,** 53–60.

Brown, J. S. A behavioral analysis of masochism. *Journal of Experimental Research in Personality,* 1965, **5,** 65–70.

Carlsmith, J. M., Collins, B. E., & Helmreich, R. L. Studies in forced compliance: I. The effect of pressure for compliance on attitude change produced by face-to-face role playing and anonymous essay writing. *Journal of Personality and Social Psychology,* 1966, **4,** 1–13.

Carlsmith, J. M., Ebbesen, E. B., Lepper, M. R., Zanna, M. P., Joncas, A. J., & Abelson, R. P. Dissonance reduction following forced attention to the dissonance. *Proceedings of the American Psychological Association,* 1969, 321–322.

Chapanis, N. P., & Chapanis, A. Cognitive dissonance: Five years later. *Psychological Bulletin,* 1964, **61,** 1–22.

Chappell, V. C. (Ed.) *The philosophy of mind.* Englewood Cliffs, N. J.: Prentice-Hall, 1962.

Corah, N. L., & Boffa, J. Perceived control, self-observation, and response to aversive stimulation. *Journal of Personality and Social Psychology,* 1970, **16,** 1–4.

Davison, G. C., & Valins, S. Maintenance of self-attributed and drug-attributed behavior change. *Journal of Personality and Social Psychology,* 1969, **11,** 25–33.

deCharms, R. *Personal causation: The internal affective determinants of behavior.* New York: Academic Press, 1968.

Deci, E. L. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology,* 1971, **18,** 105–115.

Deci, E. L. Intrinsic motivation, extrinsic reinforcement, and inequity. *Journal of Personality and Social Psychology,* 1972, in press.

Deci, E. L., & Cascio, W. F. Changes in intrinsic motivation as a function of negative feedback and threats. Unpublished manuscript, University of Rochester, 1971.

Elms, A. C. Role playing, incentive, and dissonance. *Psychological Bulletin,* 1967, **68,** 132–148.

Festinger, L. A theory of social comparison processes. *Human Relations,* 1954, **7,** 117–140.

Festinger, L. *A theory of cognitive dissonance.* Stanford: Stanford University Press, 1957.

Festinger, L. *Conflict, decision and dissonance.* Stanford: Stanford University Press, 1964.

Festinger, L., & Carlsmith, J. M. Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology,* 1959, **58,** 203–210.

Frank, J. *Persuasion and healing.* Baltimore: Johns Hopkins Press, 1961.

Freedman, J. L. Long-term behavioral effects of cognitive dissonance. *Journal of Experimental Social Psychology,* 1965, **1,** 103–120.

Freedman, J. L., & Fraser, S. C. Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology,* 1966, **4,** 195–202.

Goldman, R., Jaffa, M., & Schachter, S. Yom Kippur, Air France, dormitory food, and

the eating behavior of obese and normal persons. *Journal of Personality and Social Psychology*, 1968, **10**, 117–123.

Grinker, J. Cognitive control of classical eyelid conditioning. In P. G. Zimbardo (Ed.), *The cognitive control of motivation*. Glenview, Ill.: Scott, Foresman, 1969.

Harris, V. A., & Tamler, H. The effects of attitude reinstatement in bystander replications. Unpublished manuscript, State University of New York at Buffalo, 1971. (a)

Harris, V. A., & Tamler, H. Reinstatement of initial attitude and forced-compliance attitude change. *Journal of Social Psychology*, 1971, **84**, 127–134. (b)

Harvey, J., & Mills, J. Effect of an opportunity to revoke a counterattitudinal action upon attitude change. *Journal of Personality and Social Psychology*, 1971, **18**, 201–209.

Heider, F. Attitudes and cognitive organizations. *Journal of Psychology*, 1946, **21**, 107–112.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Heider, F. The gestalt theory of motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Vol. 8. Lincoln: University of Nebraska Press, 1960. Pp. 145–172.

Jones, E. E., & Davis, K. E. From acts to dispositions. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965. Pp. 219–266.

Jones, E. E., & Harris, V. A. The attribution of attitudes. *Journal of Experimental Social Psychology*, 1967, **3**, 1–24.

Jones, E. E., & Nisbett, R. E. The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. New York: General Learning Press, 1972.

Jones, R. A., Linder, D. E., Kiesler, C. A., Zanna, M., & Brehm, J. W. Internal states or external stimuli: Observers' attitude judgments and the dissonance theory—self-persuasion controversy. *Journal of Experimental Social Psychology*, 1968, **4**, 247–269.

Jones, R. G. Forced compliance dissonance predictions: obvious, non-obvious, or non-sense? Paper presented at the meeting of the American Psychological Association, New York, September 1966.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation*. Vol. 15. Lincoln: University of Nebraska Press, 1967. Pp. 192–238.

Kiesler, C. A., Nisbett, R. E., & Zanna, M. P. On inferring one's beliefs from one's behavior. *Journal of Personality and Social Psychology*, 1969, **11**, 321–327.

Klemp, G. O., & Leventhal, H. Self-persuasion and fear reduction from escape behavior. In H. London & R. E. Nisbett (Eds.), *Cognitive alteration of feeling states*. Chicago: Aldine, 1972.

Koenig, K. P., & Henriksen, K. Cognitive manipulation of GSR extinction: Analogues for conditioning therapies. In H. London & R. E. Nisbett (Eds.), *Cognitive alteration of feeling states*. Chicago: Aldine, 1972.

Lepper, M. R. Dissonance, self perception, and honesty in children. Unpublished manuscript, Stanford University, 1971.

Lepper, M. R., Greene, D., & Nisbett, R. E. Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. Unpublished manuscript, Stanford University, 1971.

Lepper, M. R., Zanna, M. P., & Abelson, R. P. Cognitive irreversibility in a dissonance reduction situation. *Journal of Personality and Social Psychology*, 1970, **16**, 191–198.

Leventhal, H. Findings and theory in the study of fear communications. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 5. New York: Academic Press, 1970. Pp. 120–186.

Linder, D. E., & Jones, R. A. Discriminative stimuli as determinants of consonance and dissonance. *Journal of Experimental Social Psychology*, 1969, **5**, 467–482.

Loftis, J., & Ross, L. Facilitation of GSR extinction through misattribution. Unpublished manuscript, Stanford University, 1971.

London, H., & Monello, L. Cognitive manipulation of boredom. In H. London & R. E. Nisbett (Eds.), *Cognitive alteration of feeling states*. Chicago: Aldine, 1972.

Maslach, C. The "truth" about false confessions. *Journal of Personality and Social Psychology*, 1971, **20**, 141–146.

McGuire, W. J. The current status of cognitive consistency theories. In S. Feldman (Ed.), *Cognitive consistency: Motivational antecedents and behavioral consequents*. New York: Academic Press, 1966. Pp. 1–46.

Melzack, R. The perception of pain. *Scientific American*, 1961, **204**, 41–49.

Miller, N., & Dollard, J. *Social learning and imitation*. New Haven: Yale University Press, 1941.

Mills, J. Comment on Bem's "Self-perception: An alternative interpretation of cognitive dissonance phenomena." *Psychological Review*, 1967, **74**, 535.

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

Nisbett, R. E., & Schachter, S. Cognitive manipulation of pain. *Journal of Experimental Social Psychology*, 1966, **2**, 227–236.

Nisbett, R. E., & Storms, M. D. Cognitive and social determinants of food intake. In H. London & R. E. Nisbett (Eds.), *Cognitive alteration of feeling states*. Chicago: Aldine, 1972.

Nisbett, R. E., & Valins, S. Perceiving the causes of one's own behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. New York: General Learning Press, 1972.

Ostfeld, B., & Katz, P. A. The effect of threat severity in children of varying socioeconomic levels. *Developmental Psychology*, 1969, **1**, 205–210.

Pepitone, A., McCauley, C., & Hammond, P. Change in attractiveness of forbidden toys as a function of severity of threat. *Journal of Experimental Social Psychology*, 1967, **3**, 221–229.

Piliavin, J. A., Piliavin, I. M., Loewenton, E. P., McCauley, C., & Hammond, P. On observers' reproductions of dissonance effects: The right answers for the wrong reasons? *Journal of Personality and Social Psychology*, 1969, **13**, 98–106.

Ross, L., Bierbrauer, G. A., & Polly, S. The attribution of success and failure in student-teacher interaction. Unpublished manuscript, Stanford University, 1971.

Ross, L., Rodin, J., & Zimbardo, P. G. Toward an attribution therapy: The reduction of fear through induced cognitive-emotional misattribution. *Journal of Personality and Social Psychology*, 1969, **4**, 279–288.

Ryle, G. *The concept of mind*. London: Hutchinson, 1949.

Schachter, S. *The psychology of affiliation*. Stanford: Stanford University Press, 1959.

Schachter, S. The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 1. New York: Academic Press, 1964. Pp. 49–80.

Schachter, S., & Singer, J. E. Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 1962, **69**, 379–399.

Schachter, S., & Wheeler, L. Epinephrine, chlorpromazine and amusement. *Journal of Abnormal and Social Psychology*, 1962, **65**, 121–128.

Skinner, B. F. The operational analysis of psychological terms. *Psychological Review*, 1945, **52**, 270–277, 291–294.

Skinner, B. F. *Science and human behavior*. New York: Macmillan, 1953.

Skinner, B. F. *Verbal behavior*. New York: Appleton, 1957.

Snyder, M., & Ebbesen, E. B. Dissonance awareness: A test of dissonance theory versus self-perception theory. In press.

Steiner, I. D. Perceived freedom. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 5. New York: Academic Press, 1970. Pp. 187–248.

Storms, M. D. Video tape and the attribution process: Changing actors' and observers' points of view. Unpublished doctoral dissertation, Yale University, 1971.

Storms, M. D., & Nisbett, R. E. Insomnia and the attribution process. *Journal of Personality and Social Psychology*, 1970, **2**, 319–328.

Tannenbaum, P. H. The congruity principle: Retrospective reflections and recent research. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968. Pp. 52–72.

Turner, E. A., & Wright, J. Effects of severity of threat and perceived availability on the attractiveness of objects. *Journal of Personality and Social Psychology*, 1965, **2**, 128–132.

Valins, S. Cognitive effects of false heart-rate feedback. *Journal of Personality and Social Psychology*, 1966, **4**, 400–408.

Valins, S., & Ray, A. A. Effects of cognitive desensitization on avoidance behavior. *Journal of Personality and Social Psychology*, 1967, **7**, 345–350.

Weick, K. E. Dissonance and task enhancement: A problem for compensation theory? *Organizational Behavior and Human Performance*, 1967, **2**, 175–216.

Wolosin, R. J. *Self- and social perception and the attribution of internal states*. (Doctoral dissertation, University of Michigan) Ann Arbor, Mich.: University Microfilms, 1969. No. 69–12,276.

Wolosin, R. J. Attribution of freedom to the self and others. Paper presented at the meeting of the Midwestern Psychological Association, Detroit, May 1971.

Wolosin, R. J., & Denner, B. Three studies of the attribution of freedom to the self and to others. Unpublished manuscript, Indiana University, 1970.

Zanna, M. P. Inference of belief from rejection of an alternative action. Unpublished manuscript, Princeton University, 1970.

Zimbardo, P. G. *The cognitive control of motivation: The consequences of choice and dissonance*. Glenview, Ill.: Scott, Foresman, 1969.

Zimbardo, P. G., Cohen, A., Weisenberg, M., Dworkin, L., & Firestone, I. The control of experimental pain. In P. G. Zimbardo (Ed.), *The cognitive control of motivation*. Glenview, Ill.: Scott, Foresman, 1969. Pp. 100–125.