

Psychological Bulletin

COGNITIVE DISSONANCE:

FIVE YEARS LATER ¹

NATALIA P. CHAPANIS AND ALPHONSE CHAPANIS

Johns Hopkins University

This article reviews critically the experimental evidence in support of cognitive dissonance theory as applied to complex social events. The criticisms which can be made of this literature fall into 2 main classes. 1st, the experimental manipulations are usually so complex and the crucial variables so confounded that no valid conclusions can be drawn from the data. 2nd, a number of fundamental methodological inadequacies in the analysis of results—as, e.g., rejection of cases and faulty statistical analysis of the data—vitiate the findings. As a result, one can only say that the evidence adduced for cognitive dissonance theory is inconclusive. Suggestions are offered for the methodological improvement of studies in this area. The review concludes with the thesis that the most attractive feature of cognitive dissonance theory, its simplicity, is in actual fact a self-defeating limitation.

Social psychologists have been trying for many years to predict the conditions under which attitudes and opinions are changed. In general their attempts have not been conspicuously successful. One of the first major breakthroughs in this area came when Leon Festinger (1957) published his book on *A Theory of*

Cognitive Dissonance. In this book the author presented a simple conceptual scheme by which he could predict with precision the outcomes of certain social situations. To support his theory, Festinger marshaled data from an impressive variety of field and experimental studies. In addition, he and other workers have since then conducted a number of studies designed to test specific derivations of the theory. What can we say about all this literature?

Cognitive dissonance theory has already been reviewed by Bruner (1957), Asch (1958), Osgood (1960), and Zajonc (1960). These writers, however, have been primarily concerned with a

¹ This article is based on material prepared by the senior author while she was employed at the Tavistock Institute of Human Relations, London, England. It appeared originally as Tavistock Document No. 626, "Cognitive Dissonance: A Dissenting Voice," dated June 1961. That work was part of a research program financed by S. H. Benson, Limited, and carried out by the Tavistock Institute. The extensive expansion and revision of the original document undertaken with the collaboration of the junior author was supported in part by Contract Nonr 248(55) between the Office of Naval Research and the Johns Hopkins University. We are pleased to acknowledge the assistance provided us by both organizations. The views expressed in this article are, however, those of the authors. Neither S. H. Benson, Limited, nor the Office of Naval Research is responsible for any of the statements contained in it.

We are greatly indebted to the staff of the Tavistock Institute, and particularly to Frederick E. Emery, for many helpful comments, guidance, and encouragement throughout the preparation of the original document. Grateful appreciation is also extended to our many American colleagues—too numerous to name individually—who read various versions of this article and offered helpful comments and encouragement.

critical evaluation of the conceptual system employed in dissonance theory. And, whatever they might think of the theory, most workers (except perhaps Asch) have been impressed by the scope, relevance, and ingenuity of the experimental evidence gathered in its support.

There is an engaging simplicity about Festinger's dissonance formulations. No matter how complex the social situation, Festinger assumes that it is possible to represent the meaning which the situation has for an individual by a series of elementary *cognitions*—statements that an individual might make describing his "knowledge, opinions or beliefs" (Festinger, 1957, p. 3). Moreover, a simple inventory of a group of related cognitions is sufficient to reveal whether or not they are consistent. The theory assumes further that people prefer consistency among their cognitions and that they will initiate change in order to preserve this consistency. So far these ideas are not new. They had been promulgated as early as 1946 by Heider with his concept of balance and imbalance. The magic of Festinger's theory, however, seems to lie in the ease with which imponderably complex social situations are reduced to simple statements, most often just two such statements. This having been done, a simple inspection for rational consistency is enough to predict whether or not change will occur. Such uncomplicated rationality seems especially welcome after having been told for years that our attitudes and resulting behavior are strongly dependent on motivational, emotional, affective, and perceptual processes (e.g., Krech & Crutchfield, 1948; Rosenberg, 1960).

Five years have now elapsed since the publication of Festinger's book, and this seems to be an appropriate time to pause for a close look at the evidence in support of the theory. For no matter how

appealing a theory might be, in the final analysis it is the evidence that counts. This paper, therefore, will be concerned with a review of experiments on cognitive dissonance in humans from two points of view. First we shall consider whether an experimenter really did what he said he did. Then later we shall consider whether the experimenter really got the results he said he did.

CONTROVERSIAL EXPERIMENTAL MANIPULATIONS

As we all know, good experimental work always involves manipulating conditions in such a way that we may ascribe changes we observe in our dependent variables to the manipulations we carried out on the independent variables. In actual practice we rarely define these manipulations in careful operational terms. When a pellet drops into a cup in front of a hungry rat we call it a reward, or reinforcement; when a wire transmits an electric shock to a person we call it punishment, or stress; and so on. Moreover, we do not, in general, quarrel with our fellow experimenter's interpretation of the situation. After all, he was there, he ought to know what it was about. However, when we deal with experiments on cognitive dissonance we have a very special problem on our hands.

Experimental Dissonance

Simply stated, cognitive dissonance theory is concerned with what happens when the cognitions of a person are discrepant. The basic premise is that discrepant cognitions create tension which the individual strives to reduce by making his cognitions more consistent. This tension is called cognitive dissonance, and the drive towards consistency, dissonance reduction. "When two or more cognitive elements are psychologically inconsistent, dissonance is created. Dis-

sonance is defined as psychological tension having drive characteristics" so that when dissonance arises the individual attempts to reduce it (Zimbardo, 1960, p. 86).

For our purposes at the moment the most important thing to note about the theory is that dissonance is an intervening variable whose antecedents are the private internal cognitions of a person. To test a theory like this, it is up to the experimenter to create various degrees of dissonance by introducing various discrepant cognitions within an individual. Whenever contradictory statements or syllogisms or opinions are used, there is not likely to be much controversy about the fact that they must lead to discrepant internal cognitions, and so, by definition, to dissonance. Indeed, studies on cognitive dissonance of this type have yielded results which are well-established, clear-cut, and consistent. But for the experiments under review here, the situation is rarely as simple as this. The Festinger group is primarily concerned with applying their dissonance formulation to predict complex social events. In order to do this experimentally, they use elaborate instructions and intricate relationships between experimenter and *S* to introduce discrepant cognitions and therefore to produce dissonance. Under such conditions, how can we be sure that the experimental situation has been successful in creating dissonance and dissonance alone?

In the face of such difficulties, it is always a good policy to ask the *S* himself about the situation, either directly or indirectly. It should be possible, for instance, to find out how the *Ss* perceived each of the experimental manipulations. One could also determine whether *Ss* perceived the situation as conflictful and, if so, to what extent. This kind of information is crucial to the theory of cognitive dissonance because all its pre-

dictions are based on the assumption that a state of differing, incompatible cognitions has been produced within the *S*. Unfortunately, evidence of this kind from the *Ss* themselves is not always available in the studies under review here. As a result, it is up to the reader to decide whether the experimental manipulations had the effect which the authors claim.

The other side of the coin, equally important, is that we must also assure ourselves that the experimental manipulations did not at the same time produce other internal states or cognitions within the *S* which could contaminate or even account for the findings. In fact, certain "nonobvious" derivations of some of these experiments may perhaps become a little more obvious when the experiments are reinterpreted to take other factors into consideration.

It is worthwhile spending a few moments on these nonobvious derivations. If we disregard the intermediate steps and simply consider the independent and dependent variables, it is possible to describe the essential aspects of some of these derivations by saying that they follow a *pain principle*. Reduced to essentials, some of Festinger's derivations say that the more rewarding a situation, the more negative is the effect; and contrariwise, the more painful a situation, the more positive is the effect. This is clearly illustrated by the following quotation from Festinger (1961): "Rats and people come to love the things for which they have suffered." However, if we carefully examine the kinds of experiments which are supposed to test these derivations, we find that, in general, the situation contains both painful and rewarding conditions, but that the manipulation is interpreted in terms of only one of these. It should be apparent that if a situation is both rewarding and painful, and the dependent variable

shows a positive effect, it is not legitimate to attribute it solely to the painful variable, or vice versa. To use a statistician's terminology, the variables are confounded.

Our most general criticism, then, is that some dissonance experiments have been designed in such a way that it is impossible to draw any definite conclusions from them.

Examples

The best way of illustrating these points is to describe an experimental procedure and then to analyze it from two points of view: Did the experimenter really produce the discrepant cognitions he said he did? Did the experimental manipulations produce other cognitions that could contaminate or account for his findings?

Relief or Dissonance? Let us take this experiment: College women volunteered to participate in a series of group discussions on the psychology of sex. They were seen individually by the experimenter before being allowed to join an "on-going group." Some of the girls were told they would have to pass an embarrassment test to see if they were tough enough to stand the group discussion. They were free to withdraw at this point, and one *S* did so. Girls in the severe embarrassment group had to read out loud in the presence of the male experimenter some vivid descriptions of sexual activity and a list of obscene sex words. Another group of girls—the mild embarrassment group—read some mild sexual material. All of these girls were told that they were successful in passing the embarrassment test. Each *S* then listened as a silent member to a simulated, supposedly on-going group discussion, which was actually a standard tape recording of a rather dull and banal discussion about the sexual behavior of animals. A control group listened only

to the simulated group discussion. All groups then made ratings about this discussion, its participants, and their own interest in future discussions. The ratings made by the severe embarrassment group were, on the average, somewhat more favorable than those made by the other two groups.

What was this experiment about? Was it to demonstrate the effect of feelings of relief when people discover that a task (the group discussion) is not as painfully embarrassing as the embarrassment test led them to believe? No. Was it to demonstrate the effect of success in a difficult test (passing the embarrassment test) on task evaluation? No. Was it to demonstrate the displacement of vicarious sexual pleasure from a discomfiting, but sexually arousing, situation to a more socially acceptable one? No. The experimenters called it "The Effect of Severity of Initiation on Liking for a Group" (Aronson & Mills, 1959); that is, the more painful the initiation, the more the *Ss* like the group. They predict the outcome for the severe embarrassment group in the following way: In successfully passing the embarrassment test these girls "held the cognition that they had undergone a painful experience" in order to join a group; the discussion, however, was so dull and uninteresting that they realized the unpleasant initiation procedure was not worth it. This produced dissonance since "negative cognitions about the discussion . . . were dissonant with the cognition that they had undergone a painful experience." One of the ways they could reduce this dissonance was by re-evaluating the group discussion as more interesting than it really was.

All this may be so, but in order to accept the authors' explanation we must be sure the girls really did hold these discrepant cognitions, and no others. We have to be sure, for instance, that

they felt no relief when they found the group discussion banal instead of embarrassing, that success in passing a difficult test (the embarrassment test) did not alter their evaluation of the task, that the sexual material did not evoke any vicarious pleasure or expectation of pleasure in the future, and that the group discussion was so dull that the girls would have regretted participating. There is no way of checking directly on the first three conditions, although other experimental evidence suggests that their effect is not negligible. However, to check on the fourth factor we have the data from the control group showing that the group discussion was, in fact, more interesting than not (it received an average rating of 10 on a 0-15 scale). It is, therefore, difficult to believe that the girls regretted participating. To sum up, since the design of this experiment does not exclude the possibility that pleasurable cognitions were introduced by the sequence of events, and since, in addition, the existence of "painful" cognitions was not demonstrated, we cannot accept the authors' interpretation without serious reservations.

It is interesting to speculate what would have happened if the girls had been "initiated" into the group by the use of a more generally accepted painful procedure, such as using electric shock. Somehow it seems doubtful that this group would appreciate the group discussion more than the control group, unless—and here is the crucial point—the conditions were so manipulated that Ss experienced a feeling of successful accomplishment in overcoming the painful obstacle. It seems to us that if there is anything to the relationship between severity of initiation and liking for the group, it lies in this feeling of successful accomplishment. The more severe the test, the stronger is the pleasurable feeling of suc-

cess in overcoming the obstacle. There is no need to postulate a drive due to dissonance if a *pleasure principle* can account for the results quite successfully.

The same feeling of successful accomplishment may, incidentally, be the relevant variable involved in some of the "effort" experiments done by the Festinger group (e.g., Cohen, 1959). It seems reasonable to expect that in such experiments the higher the degree of perceived effort, the greater the feeling of successful accomplishment in performing a task. Thus, *effort* would be confounded with *feeling of success*. Note, however, that success is pleasant, whereas effort is painful. Here is a situation which could be both rewarding and painful, but dissonance workers see it only as painful. (Two other effort experiments will be analyzed in greater detail later in this section.)

Reward or Incredulity? Let us look at another experiment, this time by Festinger and Carlsmith (1959): Out of several possibilities, Ss chose to take part in a 2-hour experiment falsely labeled as an experiment on "measures of performance." The Ss were tested individually and were given a "very boring" and repetitive task for about 1 hour. At the end of the hour each S was given a false explanation about the purpose of the experiment. He was told that it was an experiment to test the effect of expectation on task performance. Some Ss were then asked if they would mind acting in a deception for the next couple of minutes since the person regularly employed for this was away. The Ss of one group were hired for \$1.00 each, those of another group for \$20.00 each, to tell the next incoming S how enjoyable and interesting the experiment had been (ostensibly the expectation variable). Each S was also told he might be called on to do this again. Some Ss refused to be hired. A control group of

Ss was not asked to take part in any deception. Subsequently, all Ss (control and hired Ss) were seen by a neutral interviewer, supposedly as part of the psychology department's program of evaluating experiments. During the interview, Ss were asked to rate the experiment along four dimensions. The only significant difference between the three groups was in terms of enjoyment. The control group rated the experiment as just a little on the dull side, the \$1.00 group thought it was somewhat enjoyable, and the \$20.00 group was neutral. The mean ratings for the control and \$20.00 groups were not significantly different from each other nor from the neutral point.

What was this experiment about? The authors call it "Cognitive Consequences of Forced Compliance." They make the prediction that "the larger the reward given to the S" the smaller the dissonance and therefore "the smaller the subsequent opinion change," and "furthermore . . . the observed opinion change should be greatest when the pressure used to elicit the overt behavior is just sufficient to do it." As an aside we should point out that, inasmuch as these statements clearly refer to a maximum and so by inference to some sort of a curvilinear or nonmonotonic relationship, it would have been better if more reward categories had been used. In addition, two more control groups—a *deception-but-no-reward* group, and a *reward-but-no-deception* group—should have been included to separate out the effects of reward and deception. However, our primary concern at the moment is not with such technical matters of experimental design.

Let us examine instead the meaning of the descriptive term "forced compliance." According to Festinger (1957), it means "public compliance without private acceptance [p. 87]."

The reward Ss, it is true, complied publicly with the instructions in that they described a boring task as enjoyable to another S. Notice, however, that even the control group rated the task as only slightly boring. This suggests that the false explanation placed the task in a wider context and may have led to "private acceptance" of the whole situation by both control and reward Ss. We could also question the choice of the word "forced." Forced implies a lack of freedom, but it is extremely difficult to predict how an S perceives his freedom of choice even when this variable is experimentally manipulated (e.g., Brehm & Cohen, 1959a). All we can say is that the term forced compliance is not a good description of the events in this experiment.

What seems to be even more important, however, is that the experiment could be more appropriately entitled "The Effect of a Plausible and Implausible Reward on Task Evaluation." As far as we can tell, Ss were not asked to describe their reactions to the size of the reward. Nevertheless, \$20.00 is a lot of money for an undergraduate even when it represents a whole day's work. When it is offered for something that must be much less than 30 minutes work, it is difficult to imagine a student accepting the money without becoming wary and alert to possible tricks. In fact, more than 16% of the original Ss in the \$20.00 group had to be discarded because they voiced suspicions, or refused to be hired. Under such circumstances, it seems likely that those who were retained might have hedged or been evasive about their evaluation of the experiment. The mean rating for the \$20.00 group was $-.05$ on a scale that ranged from -5.00 (dislike) to 5.00 (like), that is, the mean rating was at the neutral point. As other workers (e.g., Edwards, 1946) have suggested,

a rating at the zero or neutral point may be ambiguous, ambivalent, or indifferent in meaning and may simply represent an evasion. The authors' data, unfortunately, do not permit us to determine whether individual Ss did in fact respond this way. In any event, if we assume that \$1.00 is a plausible, but \$20.00 an implausible, reward, then the results fall neatly into the pattern of all previous and more extensive experiments on the effect of credulity on pressures to conformity (Fisher & Lubin, 1958).

To sum up, the design of this experiment does not allow us (a) to check whether discrepant cognitions were in fact produced, and (b) to rule out alternative explanations.

Incidentally, the authors of other related studies (Brehm, 1960; Brehm & Lipsher, 1959; Cohen, Terry, & Jones, 1959) have difficulty in accounting for all of their results according to dissonance theory predictions. These difficulties disappear if we use a plausibility explanation. The argument would proceed along these lines: If an individual is subjected to many pressures towards change from a number of different sources, (a) each pressure will act on the individual, and (b) their effect will be cumulative. For instance, we may increase pressure on an individual by limiting his freedom of choice, by giving him acceptable rewards, by presenting him with statements that strongly support a position discrepant to his own, by increasing the size of the discrepancy, and so on. Each of these alone will produce a greater and greater opinion change until—and this is the critical part of the argument—the situation becomes implausible, at which point the S will ignore the pressures and show no change. It also seems reasonable to suppose that a combination of *any* of these factors will act cumulatively to produce the implausible effect.

We can express this situation in statistical terminology. For example, if we have a two-factor experiment with two levels of pressure towards opinion change in each variable, we would expect to find that the two main effects are significant. Moreover, we would predict that the interaction would also be significant primarily because the combination of both "high pressures" would be implausible and so produce the least opinion change. In general, this is the pattern of results obtained by the dissonance workers in experiments of this type.

Mealtime Troubles. Another example of an untested interpretation occurs in the Brehm (1959) experiment on the effect of a *fait accompli* in which boys were offered a prize if they ate a portion of a disliked vegetable. While eating it, some Ss were casually told that a letter would be sent to their parents informing them of their participation in the experiment and of the vegetable they ate. Those boys who indicated they had trouble about eating the vegetable at home (i.e., it was more often served than eaten) subsequently changed their rating of the disliked vegetable towards a more favorable one. What did the letter mean to these boys? According to the author it meant that "the Ss would have to eat more" of the vegetable at home. But this is a guess, not based on any evidence in the experiment. Furthermore, in an extension of the same procedure at the same school with equivalent Ss from the same classes, direct manipulation of the commitment to further eating "failed to produce an overall effect on liking [Brehm, 1960, p. 382]." Under the circumstances, we find it difficult to accept the author's contention that the *fait accompli* increased cognitive dissonance by increasing the commitment to eating. There is little doubt that mentioning the letter

changed the ratings, but only for boys who had mealtime troubles. The key to the problem most likely lies in the expectation these boys had about the effect of the letter on their parents and on themselves. However, the design of the experiment does not allow us to find out what this expectation was.

Confounded Effort. In a recent experiment, Aronson (1961) tried to separate the effects of secondary reinforcement from dissonance in a rewarding situation.

Reinforcement theory suggests that stimuli associated with reward gain in attractiveness; dissonance theory suggests that stimuli associated with "no reward" gain in attractiveness . . . if a person has expended effort in an attempt to attain the reward [p. 375].

Aronson argues that since the effect of secondary reinforcement is constant, nonrewarded objects should become more attractive as the effort to obtain them increases.

In order to test this hypothesis, Ss fished for cans to obtain a reward (\$.25) inside one third of the cans. The rewarded cans were of one color, the nonrewarded ones of another, but the Ss could not determine which they had snared until the cans had actually been pulled out. One group of Ss—the low-effort group—was told that their task was not tedious. They had the relatively easy task of fishing out a can with a magnet, a task which took them, on the average, only 14 seconds per can. Another group of Ss—the high-effort group—was told that their task was extremely tedious. They had the relatively difficult task of fishing out a can with a hook, which took them, on the average, 52 seconds per can. All Ss continued fishing for the reward money until 16 unrewarded cans had been pulled out. The Ss rated the relative attractiveness of the two colors before and after carrying out the task. The results show that

in the low-effort condition, the attractiveness of the color on the rewarded cans increased (a secondary reinforcement effect), but in the high-effort condition no change was observed. All of this was interpreted as substantiating the cognitive dissonance predictions.

Aronson explains the lack of change in the high-effort condition by saying that the effects of dissonance and secondary reinforcement are equal but opposite in direction, and so cancel each other. However, if we look more carefully at the experimental manipulation of effort, we see that the low-effort condition is actually a reward rate of \$.25 about every 42 seconds, and the high-effort condition is actually a reward rate of \$.25 about every 156 seconds. In other words, the low-effort group is, at the same time, a high-reward-rate group, and the high-effort group, a low-reward-rate group. The difference obtained between the two groups could then be simply the result of the difference in reward rates, and the lack of change in the high-effort group, the result of their low reward rate.

To summarize, Aronson tried to demonstrate the effect of effort in a rewarding situation. However, the design of the experiment confounds effort with reward rate. As a result, no unambiguous conclusions can be drawn as to the effect of effort.

There is yet another experiment on effort in which confounding occurs. Yaryan and Festinger (1961) tried to show the effect of "preparatory effort" on belief in a future event. The Ss volunteered to participate in an experiment labeled "Techniques of Study" which was supposed to investigate the techniques, hunches, and hypotheses that students use to study for exams. The Ss were told that only half of them would take part in the complete experiment which involved taking an IQ test. All Ss

were given an information sheet on which there were definitions essential to this supposed IQ test. In the high-effort condition, Ss were told to study the sheet and memorize the definitions. In the low-effort condition, Ss were told to glance over the definitions briefly. The latter were also told that they would have access to the sheet later if they were to take the IQ test. Each S was then asked to express his estimate of the probability that he would take the IQ test. The results show that Ss in the high-effort group thought it was more probable that they would take the test.

The authors (Yaryan & Festinger, 1961) explain the results in the following way: Exerting "a great deal of effort" is inconsistent with the cognition that one may not take the test, so Ss in the high-effort group should believe "more strongly in the likelihood of the occurrence of the event [p. 606]." This might very well be the case, but in this experiment the variable of effort is confounded with the presence of other predictors for the event. All Ss had been told that this was an experiment on the techniques of study, but the only group which *did* any studying was the high-effort group. In addition, the studying that was done was highly relevant for the IQ test. Under the circumstances, it does not seem at all surprising that Ss in this group took these additional cues to mean that they were assigned to the complete experiment and to the IQ test. As it stands now, the Yaryan and Festinger experiment does not separate the effect of effort from that of additional cues.

Reliability Is Not Validity. As we have seen, most cognitive dissonance formulations are concerned with what happens after a person makes a decision. One of the earliest experiments designed specifically to investigate this problem was the gambling experiment

described by Festinger (1957, p. 164) and successfully replicated by Cohen, Brehm, and Latané (1959) with minor variations in procedure. The agreement between these two studies has done much to enhance the belief in the validity of cognitive dissonance formulations (e.g., Riecken, 1960, p. 489).

The experimental procedure in these two studies was relatively simple. Each S played a card game with the experimenter for variable money stakes. Before beginning the game each S was informed of the rules of the game, and on the basis of this information chose one of two sides on which to play. He was told that he could change sides once during the game, but that it would cost him money. He was led to expect that he would play 30 games. At the end of 12 games, play was interrupted and S was given a probability graph to study. The graph, a different one for each side, gave the (false) information that the chosen side was the losing one. The dependent variables were the time spent in studying the graph and the number of people who changed sides. Results were analyzed in terms of a weighted average of the amount of money won or lost. The pattern of results obtained is very complex and would require at least a fourth-order parabola to describe it. Nevertheless, the various ups-and-downs were interpreted as supporting the dissonance theory predictions for postdecision, information seeking processes.

Two things strike us about the dissonance theory interpretation of this experiment. First, Festinger is not consistent in his dissonance formulations. Let us look at the way in which the results are interpreted. The money winners spent a moderate amount of time studying the graph. Festinger considers this the result of dissonance produced by the information in the graph which purported to show that these winners

were actually on the losing side. If we accept this line of reasoning, it should follow that the losers would have no such dissonance (the graph confirmed their losses), and would therefore spend little time on the graph. This was not so. Festinger (1957), however, has three other dissonance explanations to account for the complex behavior of the losers. He argues, first of all, that loss of money is itself dissonance producing and the bigger the loss, the bigger the dissonance. The small losers spent as much time as they did in the "hope that the graph would tell them they were actually on the correct side [p. 171]." If this explanation is correct, the bigger losers should have spent an even longer time searching the graph—but they did not. Festinger explains this away by saying that the bigger losers would avoid the graph "if the graph were perceived as yielding information which would probably increase the dissonance which already existed [p. 172]." If this explanation is correct, the biggest losers should have spent the least amount of time on the graph—but they did not. To explain this behavior Festinger postulates yet another hypothesis. For the biggest losers "the easiest way to eliminate the dissonance would be to increase it temporarily to a point when it was greater than the resistance to change of the behavior," that is, they would study the graph, then switch sides. If this explanation is true, then we would expect that all of the biggest losers, and only the biggest losers, would switch sides—but this was not so. It should be noted that these four dissonance hypotheses are inconsistent with each other, since they predict effects in different directions. Moreover, there is no *a priori* way of determining the degree to which each particular hypothesis applies to the groups. This whole matter can be summarized in another way: If

the pattern of results had been exactly the reverse, these same explanations would apply just as well.

This brings us to the second point. The most important criticism of this gambling experiment is that it is not so much an experiment on the dissonance-reducing effects of information in post-decision processes, as it is an experiment on "information seeking in predecision processes" (suggested by F. E. Emery). The Ss had been told that they could change sides and they were actually given an opportunity to do so when they were handed the graph. Festinger (1957) and Cohen et al. (1959) reported that many Ss, both winners and losers, announced their decision to switch sides at this time. What was not reported, however, was the number of Ss who looked at the graph in order to reach a decision whether or not it would be more profitable to change sides. In other words, Ss looked at the graph not to reduce dissonance, but to look for information to help them decide whether they should change sides. With this interpretation, the pattern of results becomes more obvious and reasonable. For example, in one of the two conditions of the Cohen et al. replication those Ss who neither won nor lost any money spent the most time looking at the graph. This finding is entirely inexplicable in dissonance terms, even with an imaginative use of all four of Festinger's hypotheses. However, if we consider this experiment to be concerned with predecision processes, then we see that these Ss had gained the least amount of information from the actual play of the game and were, therefore, trying to extract as much information as possible from the graph before reaching a decision.

Taking all of these factors into consideration, we are forced to conclude that Festinger's interpretations, however in-

genious, are unnecessarily elaborate and unjustified. Moreover, the successful replication of the experiment suggests—not that the cognitive dissonance formulations are valid—but only that the results of experiments of this type are reproducible.

Subsequent experiments on selected aspects of information seeking in post-decision processes have done nothing to clarify the situation. Adams (1961) and Maccoby, Maccoby, Romney, and Adams (1961) showed that people tend to seek information which agrees with their viewpoint, but Rosen (1961) obtained results which show that people tend to seek information which disagrees with their viewpoint. It seems more likely to us that, in general, people will often seek new information, whether it be consonant or contrary. Indeed this is apparently the kind of result that Feather (1962) obtained. To be sure, when prejudice or some other highly motivated state is involved, people are selective in their perceptions and avoid contrary material. But under these conditions it is the motive itself, and not dissonance, that seems to be crucial.

Interpretation of Manipulations

Perhaps the illustrations cited above will suffice to show that the experiments adduced to support the theory of cognitive dissonance involve highly complex

manipulations. The effects of these manipulations are open to alternative explanations which have generally not been dealt with adequately by the authors. We can diagram this in the following way: Let us suppose that the complex experimental manipulations produce the Cognitions 1, 2, 3, 4 . . . n in the S , as illustrated on the left side of Figure 1. Two of these cognitions (Cognitions 1 and 2) are chosen by the experimenter as being the relevant discrepant cognitions producing dissonance. Any observed change in the dependent variable is then attributed to that dissonance. But, as we see by examining the whole of Figure 1, this may not necessarily be the case. Any one cognition, or any combination of cognitions, could have been responsible for the change in the dependent variable. There is no way of ascertaining which, because the effects of all these cognitions have been confounded.

It is possible to design experiments so that these effects are not confounded. As a step in this direction we recommend first of all that the experimental manipulations be simplified. It is difficult to agree about differences in cognitions when the instructions, task, and procedure differ in many ways for control and experimental groups. Our second recommendation is that additional control groups be included in the design

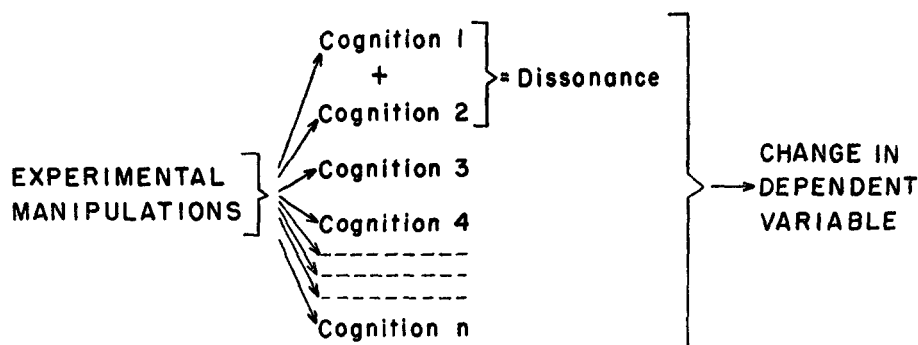


FIG. 1. The type of confounding frequently found in experiments on cognitive dissonance.

of these experiments to deal with the irreducible differences in experimental manipulation. Our third recommendation is that a little more attention be given to discovering the possible cognitions that an *S* might have about the situation, particularly those which might be contrary to dissonance theory. Only under such carefully controlled conditions can we begin to talk about unequivocal evidence in support of cognitive dissonance theory.

CONTROVERSIAL TREATMENTS OF THE DATA

So much for experimental manipulations. Now to see if the experimenter really got the results he says he did. Our most serious criticisms of the experiments cited in support of dissonance theory fall under the heading of methodological inadequacies in the analyses of results. Of these inadequacies the most important is the rejection of cases, not only because it is so fundamental a flaw, but also because the supporters of dissonance theory so often do it.

Rejection of Cases

If an experimenter is interested in the performance of only a certain group of *Ss*, it is legitimate for him to select these *Ss* before beginning the experiment, or sometimes even after the experiment, before the results are analyzed. However, when *Ss* are selectively discarded after the data of an experiment have been collected, tabulated, and sometimes even analyzed, it leaves the reader with a feeling of uneasiness. The uneasy feeling grows if the *Ss* are discarded because their results are said to be "unreliable," or if the experimenter gives inconsistent reasons—or no reasons at all—for their rejection. But let us look at the experiments themselves.

Unreliable Ss. Brehm and Cohen (1959b) asked sixth-grade children to

indicate how much they liked several different toys before and after they had chosen one for themselves. The choice of the gift and the postchoice rating were made a week after the prechoice rating. The authors hypothesized that there would be an increase in the evaluation of the chosen article, and a decrease in the evaluation of the nonchosen article, the greater the dissimilarity among the toys, and the greater the number of alternatives from which to choose. In general these predictions were upheld. But of the original sample of 203 children *only 72* were used in the analysis. In the authors' (Brehm & Cohen, 1959b) words, the reasons for the reduction were as follows:

First, the choice alternatives for each *S* had to be liked, but not so much that an increase in liking would be impossible. Second, one alternative had to be initially more liked than any other so that its choice could be expected. Increased ratings of the chosen item are thus not likely to be simply a result of normal (and random) changes in actual liking from the first questionnaire to the second. *In addition, Ss who failed to choose the alternative initially marked as most liked, were excluded because they gave unreliable or invalid ratings.* Finally, in order to ensure that initially less liked alternatives were seriously considered as possible, initial ratings of these alternatives could not be much lower than the most liked alternative [p. 375; italics added].

Note first that the exact limits of all these requirements were determined only after inspection of the data, despite the fact that each *S* had been given a prearranged choice based on his initial ratings. However, let us look more carefully at the italicized item—the exclusion of *Ss* because of unreliability. If *Ss* give unreliable results, it is usual to assume that the measuring instrument itself is unreliable; indeed, the authors themselves admit this when they mention "the low reliability of our measure of liking." However, discarding selected *Ss* does nothing to improve the reliability of an instrument.

Discarding Ss who did not choose the alternative initially marked as most liked may in fact falsely reduce the computed error variance, change the mean values, and so enhance the possibility of obtaining a significant difference in rating. To illustrate, the upper half of Figure 2 shows the ratings for two toys, X and Y, which satisfy the conditions specified by Brehm and Cohen: they are both liked, one is liked more than the other, but the difference is not great. Now let us assume that the ratings are subject to errors of measurement and that they vary randomly from time to time. Let us further assume that the expressed rating, the liking, at any one moment in time, is perfectly correlated with choice.

The situation a week later is shown in the bottom half of Figure 2. The ratings for X are now distributed as X' , the ratings for Y, as Y' . Now the Ss are asked to make a choice. Let us assume the null hypothesis, that is, the actual process of choosing a gift does not alter the liking or rating of a toy. Since the choice and the postchoice rating occur so close together in time we can also assume that no random change occurs from just before the choice to just after

it. If, as Brehm and Cohen did, we discard all those Ss who chose Y rather than X, this means that we eliminate from the shaded area in the bottom half of Figure 2 all those cases in which $Y' > X'$. Such a process can only reduce the variance of both distributions of differences in ratings (that is, $X' - X$ and $Y' - Y$), and automatically increase the mean difference between them. Moreover, this selection procedure will automatically produce precisely the effect which the authors predicted, namely, that the mean rating for the chosen toy increases, while the rating for the non-chosen toy decreases. Furthermore, these effects will be greatest in the group from which most such Ss were discarded. Most of these discards came from the "four alternatives condition,"² and the change in rating is actually greatest for this condition.

In a footnote (p. 376) the authors (Brehm & Cohen, 1959b) state they carried out similar computations on their "unselected population," that is, on the entire sample of 203 Ss. Although the sensitivity of the statistical test is now nearly twice as great (because of the increase in N from 72 to 203), they find

² J. Brehm, personal communication, 1961.

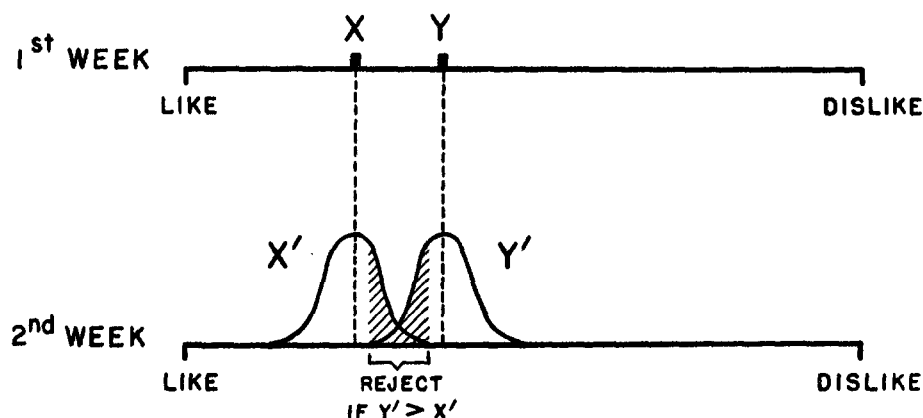


FIG. 2. This illustration shows how the rejection of Ss from the shaded area may have introduced a statistical artifact into the experiment by Brehm and Cohen (1959b).

no effect due to the number of alternatives (one of the two predictions made). Tests of the other prediction "yield support" for the "dissimilarity hypothesis." It is not clear, however, whether the authors mean by this a statistically significant difference, or simply a nonsignificant trend. To sum up, it seems reasonable to conclude that the significant results obtained in this and similar experiments may very well be statistical artifacts.

Contradictory Reasons. Sometimes it is difficult to reconcile the reasons given for the rejection of cases with other statements by the same author. For example, in the experiment on "Attitude Change and Justification for Compliance" (Cohen, Brehm, & Fleming, 1958), the initial analysis showed no significant difference between the two justification groups. The authors then eliminated more than half of the Ss (47 out of 92), carried out a second analysis on the remainder, and concluded that "the difference in amount of change is significant by one-tailed *t* test at .07 level." Relatively more Ss whose opinion did not change were eliminated from the low-justification condition (35 out of 63). Not surprisingly, the new mean for the low-justification condition turned out to be greater than for the high-justification condition.

Part of the reasoning for this selection of cases was as follows (Cohen et al., 1958): "since extremity of position inhibits attitude change . . . it seems reasonable to eliminate the extremes [p. 277]." A year later, however, Cohen (1959) made this statement: "If the individual . . . engages in some behavior with regard to the contrary communication . . . then the greater the discrepancy [extremity of position], the greater the opinion change [p. 387]."

In their original article, Cohen et al. state that their results should be in-

terpreted cautiously, but, unfortunately, they do not follow their own advice. Whenever these authors refer to their findings in later articles (e.g., Brehm, 1960; Cohen, 1960), they quote their results as substantiating cognitive dissonance theory without any of these cautionary reminders.

What Is Going On? An example of sample reduction for obscure reasons occurs in an experiment on the readership of "own car" and "other car" advertisements by new and old car owners (Ehrlich, Guttman, Schonbach, & Mills, 1957). A group of 65 new car owners was randomly chosen from a list of recent auto registrants. The car advertisements read by this group were compared with those read by a group of 60 old car owners chosen from a telephone directory. The raw data for these analyses were the percentages of car advertisements noticed and read in a selection of magazines and newspapers which the owners had previously indicated they read regularly. The cognitive dissonance theory predictions were that new car owners would most often read advertisements about their own make of car and avoid reading those of competing makes. In general, these predictions were upheld for the data presented.

The principal difficulty with this experiment is that cases were successively rejected in various stages of the analysis so that when one finally arrives at the critical statistical test it is virtually impossible to determine what the remaining data mean. Let us see if we can trace the authors' steps in this process. The authors first present a table showing the mean percentage of advertisements noticed and mean percentage of advertisements read of those noticed for each of the categories "own," "considered," and "other car." They (Ehrlich et al., 1957) state in a note accompanying the table that:

The *N*'s are reduced because in some cases no advertisements of a particular kind appeared in the issues shown or none of those which appeared were noticed. They are further reduced because not all respondents named cars as "seriously considered" [p. 99].

The first and third reasons impose a limit on the number of *S*s whose results could be used. The largest reduction due to these two limitations was in the category "considered car" for old car owners, where the *N* of 60 was reduced to 31.

The second reason given in the quotation above means that an owner who did not notice any advertisement in a particular category was discarded from the table of "advertisements read" for that category. For example, if an owner noticed (or noticed and read) an ad about another car but did not notice any advertisements about his own car, he was included under the category of "other car," but excluded from the category of "own car" in computing the mean percentage of "advertisements read of those noticed." Up to one third of the remaining cases were eliminated from the various categories for this reason.

The next point at which still more cases are rejected is in the computation of several sign tests of significance. We are told that the *N*s are reduced because not all comparisons were possible. What this means is that significance tests were computed only on those owners who *noticed at least one advertisement in each of the pairs of categories compared*. Finally, those owners who read an equal percentage of advertisements in each of the two categories were also discarded.

Taking all of the above factors into account we find that as much as 82% of the original sample was discarded in certain categories!

The sign tests mentioned above were used only for making certain pairs of comparisons. For overall tests of their hypotheses the authors resorted to chi

square and give terminal chi square values, with their associated probabilities, alongside the tables for the sign tests. The article itself does not say upon what *N*s, or what groupings, the chi squares were computed, but correspondence³ reveals that the chi square tests were made on the same *S*s as were used in computing the sign tests.

At best this entire situation may be described as unclear. In the first place, it is difficult to know how to interpret significance tests based on such highly selected data. Furthermore, in computing chi squares for the same *S*s as were used in the sign tests, it appears that the authors discarded some data (the ties) which should properly have been included. If we have been able to thread our way correctly through the authors' manipulations of the data we find that the chi squares, computed for all the relevant data, are less significant than reported by the authors, and, in two of three cases, change a nominally significant value to a nonsignificant one. In any event, there can be no doubt that the authors' (Ehrlich et al., 1957) summary statement "It was found that new car owners read advertisements of their own car more often than . . ." needs considerable qualification. With so much selection of *S*s and with such intricate manipulations of the data, some of it never fully explained, one can hardly describe the results as *public*, or the findings as necessarily significant.

Manipulation Not Successful. Still another type of rejection we find in these studies is the elimination of entire groups of *S*s. If one variant of the manipulation fails to show an effect, it is not legitimate to discard all the *S*s in that group from the analysis. The analysis should properly be carried out on all the data and the interpretations should be based on the complete analy-

³ J. Mills, personal communication, 1962.

sis. Brehm (1960), for example, used reports on the vitamin and mineral content of vegetables to try to influence the attitudes of Ss after they had eaten a disliked vegetable. One group of Ss received the vitamin report, the other group the mineral report. Since the mineral report "failed to affect" the dependent variable "the results for these subjects [were] omitted from this report [Brehm, 1960, footnote, p. 380]." One consequence of rejecting an entire group is that we do not know if there is a significant interaction between type of report and the other variables. Until this is established, it is misleading to consider a segment of the results as significant. In addition, the author nowhere states that his findings are specific to one type of report only. His summary is in terms of "communications about food value."

Reallocation Instead of Rejection. An interesting variant of the rejection of Ss occurs in the Raven and Fishbein (1961) study on the effect of "Acceptance of Punishment and Change in Belief." Groups of Ss were run under two conditions, "shock" and "no-shock." There were 13 females and 13 males in each of these two conditions. The results show that there was no overall difference between the shock and no-shock groups. However, when the results were tabulated separately for the two sexes, it appeared that the female Ss in the shock group changed in the predicted direction, but that the male Ss did not. Here is how the authors (Raven & Fishbein, 1961) dealt with the situation:

Overall analysis of variance and interaction was not significant. Assuming that male shock Ss were part of a common population with the non-shock subjects, with respect to dissonance, an analysis of variance was conducted which showed the female shock subjects to be significantly different from the others [p. 415].

In other words the authors disposed of

the Ss who did not conform to their prediction, not by rejecting them, but by reallocating them to another group, the no-shock group. If females in the shock group really are significantly different from all the others, this should show up in a significant interaction. It does not.

Rejection of cases is poor procedure, but reallocation of Ss from experimental to control group, across the independent variable, violates the whole concept of controlled experimentation.

Danger of Rejecting Ss. A theme common to many of these rejections is that the unselected sample "does not permit of an adequate test" of the dissonance hypothesis. We are told:

In a social influence situation there are a number of potential channels of dissonance reduction, such as changing one's own opinion, changing the opinion of the communicator, making the communicator noncomparable to oneself, seeking further support for one's position, dissociating the source from the content of the communication, and distorting the meaning of the communication [dissonance theory position ably summarized by Zimbardo, 1960, p. 86].

Such a theoretical formulation is indeed all-encompassing and it provides a rationale which certain other dissonance theory workers have used for rejecting cases. The reasoning goes like this: If some Ss do not follow the specific predictions in a particular experiment (for instance, if they fail to show any opinion change) then those Ss are probably reducing their dissonance through some other channel or else they had little dissonance to begin with. If either of these conditions holds it is legitimate to exclude these Ss from the analysis since they could not possibly be used to test the particular hypothesis in the experiment. An inspection of results is considered sufficient to determine whether Ss are, or are not, to be excluded. Unfortunately, this line of reasoning contains one fundamental flaw: *it does not*

allow the possibility that the null hypothesis may be correct. The experimenter, in effect, is asserting that his dissonance prediction is correct and that Ss who do not conform to the prediction should be excluded from the analysis. This is a foolproof method of guaranteeing positive results.

Some people may feel that no matter how questionable the selection procedure, it must still mean something if it leads to significant results. This point of view, however, cannot be reconciled with the following fact of life: it is always possible to obtain a significant difference between two columns of figures in a table of random numbers provided we use the appropriate scheme for rejecting certain of those numbers. For all we know, selecting Ss so as "to permit an adequate test of the hypothesis" may have had precisely this effect. A significance test on selected Ss may therefore be completely worthless.

We strongly recommend that Ss not be discarded from the sample *after* data collection and inspection of the results. Nor is it methodologically sound to reject Ss whose results do not conform to the prediction on the grounds that they have no dissonance, or that they must be reducing it some other way. If there are any theoretical grounds for suspecting that some Ss will not show the predicted dissonance-reduction effect, the characteristics of such Ss, or the conditions, should be specifiable in advance. It should then be possible to do an analysis on all Ss by dividing them into two groups, those predicted to show dissonance reduction, and those predicted not to show it. If such a thing as dissonance reduction exists, it is theoretically and practically important to know the precise conditions under which it does and does not occur.

A summary of experiments in which Ss are rejected is given in Table 1.

TABLE 1
LIST OF EXPERIMENTS FROM WHICH Ss WERE DISCARDED AFTER DATA COLLECTION

Experiment	Total <i>N</i>	Discarded (%)	Reasons given
Brehm (1956)	225	35	To permit adequate test of hypothesis 1. Unreliable Ss 2. Conditions not fulfilled
Brehm (1960)	85 ^a	38 ^a	One manipulated condition not significant
Brehm & Cohen (1959b)	203	65	To permit adequate test of hypothesis 1. Ceiling effect for high scorers 2. Adequate separation of choice points for dissonance to occur 3. Unreliable Ss
Brehm & Lipsher (1959)	114	10-14	None
Cohen, Brehm, & Fleming (1958)	92	51	To permit adequate test of hypothesis 1. Extremity of attitude inhibits attitude change
Ehrlich et al. (1957)	125	17-82	1. Material missing 2. Advertisements not noticed 3. Not all comparisons possible 4. Ties
Mills (1958)	643	30	To permit adequate test of hypothesis 1. Ceiling effect for high scorers 2. Honest improvers have no dissonance

^a Estimated.

Refusals

The previous section has been concerned with sampling bias due to the deliberate rejection of cases by the experimenter. There is another type of sampling bias, equally important but much more subtle, that occurs when Ss reject themselves from the study by refusing to participate.

In a recent review of cognitive dissonance experiments, Cohen (1960) concluded with what he considered was a "depressing" and "Orwellian" statement:

It could be said that when the individual feels that he has most freedom of choice, when his volition and responsibility are most engaged, he is then most vulnerable to the effects of persuasive communications and to all sorts of controlled inducements from the world at large [p. 318].

This statement follows hard on the heels of "the more negative the person is toward a communication or communicator, the more he can be expected to change his attitudes in the direction of the communication or communicator." These are indeed sweeping generalizations, particularly since they are based on the results of experiments in which from 4% (Cohen, Terry, & Jones, 1959) to as many as 46% (Rabbie, Brehm, & Cohen, 1959) of the total number of Ss refused to participate. Moreover, there is evidence in these studies that the Ss who refused to participate were actually those who had both the greatest freedom of choice and the strongest (most negative) views on the attitude in question. What actually appears to have happened is that those Ss with the strongest (most negative) views were so *invulnerable* to the effects of persuasive communications that they exercised their freedom of choice by walking out on the experimenter or refusing to comply in other ways. To take the results

of the remaining more vulnerable Ss and extrapolate from them to the population in general seems unjustified.

Inadequate Design and Analysis

It is rare to find in this area a study that has been adequately designed and analyzed. In fact, it is almost as though dissonance theorists have a bias against neat, factorial designs with adequate Ns, capable of thorough analysis either parametrically or nonparametrically. The majority of their experiments are some variant of the 2×2 factorial with unequal, nonproportional, and generally small Ns in each cell. These restrictions make it impossible for the authors to carry out ordinary analyses of variance. Instead we find them making use of a hodgepodge of *t* tests and a statistic which they refer to as an "interaction *t*" (Walker & Lev, 1953, pp. 159-160).

Making a number of ordinary *t* tests on the same set of data, without a prior overall test of the null hypothesis, can be misleading. The principal difficulty is that in making such multiple comparisons the experimenter is allowing himself a number of opportunities to find an event (significance) which normally occurs infrequently. As a result, the usual *t* tables underestimate the true probabilities, that is, the probabilities obtained suggest a level of significance which is higher than warranted. Another way of saying it is that if, out of several subgroups, one finds one or two *t*'s significant, he is, in effect, capitalizing on chance (e.g., Sakoda, Cohen, & Beall, 1954). A further complication arises if the interaction is significant, since this introduces the usual difficulties about interpreting the main effects (e.g., Lindquist, 1953, p. 209). Some of the special statistical problems involved in the "postmortem" testing of comparisons were, of course, being discussed in the

psychological literature well before dissonance theory appeared on the scene (e.g., McHugh & Ellis, 1955); but for an excellent discussion of the basic issues involved in making multiple comparisons, see the article by Ryan (1959). None of these problems is ever faced squarely by the writers in this field. As a result, the authors sometimes reach conclusions that are not really warranted.

Examples. We can illustrate these remarks by referring to an analysis carried out by Brehm (1960) on two treatment variables, commitment and communication. There are three levels of commitment—control, low-eating, and high-eating—and, in addition, two types of communication—support and no support. Since the *N*s for these six groups are different (they vary from 7 to 11), it is not possible to carry out an ordinary analysis of variance. With such data at least 15 *t* tests and 3 interaction *t*'s are possible. Brehm gives the results of 7 such *t* tests (4 are nominally significant) and 2 such interaction *t*'s (both nominally significant). How do we interpret the results? Frankly, it is impossible. Taken at its face value, the analysis is not only useless, it is misleading.

An allied set of criticisms can be leveled at the analysis carried out by Brehm and Cohen in their study of the effects of choice and chance in cognitive dissonance (1959a). The design involved two types of relative deprivation, high and low. Five sections of an introductory psychology course were used as *S*s. The low- and high-deprivation conditions were experimentally manipulated and perceived as such by the *S*s. The low- and high-choice conditions were, however, determined separately for each section on the basis of their medians on the perceived-choice rating scale. Separate interaction *t*'s were calculated for each of the five sections. The *N*s in each

cell were very small, ranging between 3 and 10 with an average of about 7. The probability values for these 5 interaction *t*'s showed that one was significant, two tended to significance, and two were nonsignificant (one was actually a reversal). Here again the authors' failure to compute and report the results of an overall test make it exceedingly difficult for readers to interpret their findings. Moreover, there seems to be little justification for using a different value for the cutoff point between high- and low-choice for each section. In fact, such a procedure might in itself lead to statistically significant median differences between the sections. There may indeed be a significant interaction between choice and deprivation, but the evidence for it is, at best, questionable.

In dissonance experiments there is often a marked change between the pre- and posttest measures for both control and experimental groups. This is in itself an interesting phenomenon and should be thoroughly evaluated. An analysis should be complete—large main effects should not be ignored just because dissonance theory predicts only an interaction, or vice versa.

It is not impossible to apply a rigorous methodology to this area. Dissonance theorists would have done well to emulate the example set by Kelman as far back as 1953 (a study which, incidentally, anticipates and predates most of the areas of interest for cognitive dissonance workers). All of the problems that beset research in this area, such as unequal *N*s, class differences, and so on, were handled expertly by Kelman. More recently, such eclectic workers as Rosenbaum and Franc (1960) and McGuire (1960) have also been working in this area and have been using rigorous and comprehensive methods of analysis. In short, there appears to be no reason why

methodology in this area cannot be sharpened.

A summary of experiments in which the analyses and statistical interpretations are doubtful is given in Table 2.

Straining for Significance

The final feature of the analyses that is apt to be misleading is the fact that authors tend to present results as significant and as supporting the dissonance theory prediction when the probabilities are greater than the usually accepted value of .05. Probability values between .06 and .15 (once even .50!) do not constitute striking support for any theory, particularly if it is preceded by a selection of Ss and poor analysis. It is also extremely disconcerting to find these statistically nonsignificant trends quoted authoritatively in subsequent reports and later reviews as substantiating the theory, without any qualifying statements.

OVERALL EVALUATION

Having now reviewed much of the experimental work supporting cognitive dissonance theory, we conclude that, as a body of literature, it is downright disappointing. Too many studies have failed to stand up to close scrutiny. Yet it is also obvious that the dissonance framework has a seductive allure for many social scientists, an allure not possessed by the rather similar, but symbolically more complex, interpretations by Heider (1958), Osgood and Tannenbaum (1955), or Newcomb (1953).

Paradox of Simplicity

The magical appeal of Festinger's theory arises from its extreme simplicity both in formulation and in application. But in our review we have seen that this simplicity was generally deceptive; in point of fact it often concealed

TABLE 2

SUMMARY OF SOME EXPERIMENTS WITH INADEQUATE DESIGN AND ANALYSIS

Study	Criticism of design and analysis
Allyn & Festinger (1961)	No control group (repeat attitude test, no talk); interaction significance not presented
Aronson & Mills (1959)	Overall significance not presented
Brehm (1956)	Maximum $N = 225$, but regression equation based on $N = 557$ and $N = 534$
Brehm (1960)	Overall significance not presented
Brehm & Cohen (1959a)	Overall significance not presented
Cohen (1959)	No control group (repeat attitude test, no counterinformation); groups not equated on initial attitude
Cohen, Terry, & Jones (1959)	No control group (repeat attitude test, no new information); groups not equated on initial attitude
Ehrlich et al. (1957)	No control group (predecision car ad reading)
Festinger & Carlsmith (1959)	Overall significance not presented
Mills (1958)	Overall significance not presented
Mills, Aronson, & Robinson (1959)	No control group (preferences, but no decision); overall significance not presented
Rosen (1961)	No control group (preferences, but no decision)

a large number of confounded variables. Clearly much can be done to untangle this confounding of variables by careful experimental design. Nonetheless, there may still remain another problem more fundamental than this. In general, a cognitive dissonance interpretation of a social situation means that the relevant social factors can be condensed into two simple statements. To be sure, Festinger does not say formally that a dissonance theory interpretation works only for two discrepant statements; but it is precisely because in practice he does so limit it that the theory has had so much acceptance. Which brings us now to the crux of the matter: *is it really possible to reduce the essentials of a complex social situation to just two phrases?* Reluctantly we must say "No." To condense most complex social situations into two, and only two, simple dissonant statements represents so great a level of abstraction that the model no longer bears any reasonable resemblance to reality. Indeed the experimenter is left thereby with such emasculated predictors that he must perforce resort to a multiplicity of ad hoc hypotheses to account for unexpected findings. We see then that the most attractive feature of cognitive dissonance theory, its simplicity, is in actual fact a self-defeating limitation.

In conclusion, all of the considerations detailed above lead us to concur with Asch's (1958) evaluation of the evidence for cognitive dissonance theory, and return once more a verdict of NOT PROVEN.

REFERENCES

- ADAMS, J. S. Reduction of cognitive dissonance by seeking consonant information. *J. abnorm. soc. Psychol.*, 1961, 62, 74-78.
- ALLYN, JANE, & FESTINGER, L. The effectiveness of unanticipated persuasive communications. *J. abnorm. soc. Psychol.*, 1961, 62, 35-40.
- ARONSON, E. The effect of effort on the attractiveness of rewarded and unrewarded stimuli. *J. abnorm. soc. Psychol.*, 1961, 63, 375-380.
- ARONSON, E., & MILLS, J. The effect of severity of initiation on liking for a group. *J. abnorm. soc. Psychol.*, 1959, 59, 177-181.
- ASCH, S. E. Review of L. Festinger, *A theory of cognitive dissonance*. *Contemp. Psychol.*, 1958, 3, 194-195.
- BREHM, J. W. Postdecision changes in the desirability of alternatives. *J. abnorm. soc. Psychol.*, 1956, 52, 384-389.
- BREHM, J. W. Increasing cognitive dissonance by a *fait accompli*. *J. abnorm. soc. Psychol.*, 1959, 58, 379-382.
- BREHM, J. W. Attitudinal consequences of commitment to unpleasant behavior. *J. abnorm. soc. Psychol.*, 1960, 60, 379-383.
- BREHM, J. W., & COHEN, A. R. Choice and chance relative deprivation as determinants of cognitive dissonance. *J. abnorm. soc. Psychol.*, 1959, 58, 383-387. (a)
- BREHM, J. W., & COHEN, A. R. Re-evaluation of choice alternatives as a function of their number and qualitative similarity. *J. abnorm. soc. Psychol.*, 1959, 58, 373-378. (b)
- BREHM, J. W., & LIPSHER, D. Communicator-communicatee discrepancy and perceived communicator trustworthiness. *J. Pers.*, 1959, 27, 352-361.
- BRUNER, J. Discussion of Leon Festinger: The relation between behavior and cognition. In J. S. Bruner, E. Brunswik, L. Festinger, F. Heider, K. F. Muenzinger, C. E. Osgood, & D. Rapaport, *Contemporary approaches to cognition: A symposium held at the University of Colorado*. Cambridge: Harvard Univ. Press, 1957. Pp. 151-156.
- COHEN, A. R. Communication discrepancy and attitude change: A dissonance theory approach. *J. Pers.*, 1959, 27, 386-396.
- COHEN, A. R. Attitudinal consequences of induced discrepancies between cognitions and behavior. *Publ. Opin. Quart.*, 1960, 24, 297-318.
- COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *J. abnorm. soc. Psychol.*, 1958, 56, 276-278.
- COHEN, A. R., BREHM, J. W., & LATANÉ, B. Choice of strategy and voluntary exposure to information under public and private conditions. *J. Pers.*, 1959, 27, 63-73.
- COHEN, A. R., TERRY, H. I., & JONES, C. B. Attitudinal effects of choice in exposure to counter-propaganda. *J. abnorm. soc. Psychol.*, 1959, 58, 388-391.
- EDWARDS, A. L. A critique of "neutral" items in attitude scales constructed by the method

- of equal appearing intervals. *Psychol. Rev.*, 1946, 53, 159-169.
- EHRLICH, D., GUTTMAN, I., SCHONBACH, P., & MILLS, J. Postdecision exposure to relevant information. *J. abnorm. soc. Psychol.*, 1957, 54, 98-102.
- FEATHER, N. T. Cigarette smoking and lung cancer: A study of cognitive dissonance. *Aust. J. Psychol.*, 1962, 14, 55-64.
- FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.
- FESTINGER, L. The psychological effects of insufficient rewards. *Amer. Psychologist*, 1961, 16, 1-11.
- FESTINGER, L., & CARLSMITH, J. M. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, 58, 203-210.
- FISHER, S., & LUBIN, A. Distance as a determinant of influence in a two-person serial interaction situation. *J. abnorm. soc. Psychol.*, 1958, 56, 230-238.
- HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, 21, 107-112.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- KELMAN, H. C. Attitude change as a function of response restriction. *Hum. Relat.*, 1953, 6, 185-214.
- KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- MACCOBY, ELEANOR E., MACCOBY, N., ROMNEY, A. K., & ADAMS, J. S. Social reinforcement in attitude change. *J. abnorm. soc. Psychol.*, 1961, 63, 109-115.
- MCGUIRE, W. J. Cognitive consistency and attitude change. *J. abnorm. soc. Psychol.*, 1960, 60, 345-353.
- MC HUGH, R. B., & ELLIS, D. S. The "post-mortem" testing of experimental comparisons. *Psychol. Bull.*, 1955, 52, 425-428.
- MILLS, J. Changes in moral attitudes following temptation. *J. Pers.*, 1958, 26, 517-531.
- MILLS, J., ARONSON, E., & ROBINSON, H. Selectivity in exposure to information. *J. abnorm. soc. Psychol.*, 1959, 59, 250-253.
- NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, 60, 393-404.
- OSGOOD, C. E. Cognitive dynamics in the conduct of human affairs. *Publ. Opin. Quart.*, 1960, 24, 341-365.
- OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, 62, 42-55.
- RABBIE, J. M., BREHM, J. W., & COHEN, A. R. Verbalization and reactions to cognitive dissonance. *J. Pers.*, 1959, 27, 407-417.
- RAVEN, B. H., & FISHEIN, M. Acceptance of punishment and change in belief. *J. abnorm. soc. Psychol.*, 1961, 63, 411-416.
- RIECKEN, H. W. Social psychology. *Annu. Rev. Psychol.*, 1960, 11, 479-510.
- ROSEN, S. Postdecision affinity for incompatible information. *J. abnorm. soc. Psychol.*, 1961, 63, 188-190.
- ROSENBAUM, M. E., & FRANC, D. E. Opinion change as a function of external commitment and amount of discrepancy from the opinion of another. *J. abnorm. soc. Psychol.*, 1960, 61, 15-20.
- ROSENBERG, M. J. An analysis of affective-cognitive consistency. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm, *Attitude organization and change*. New Haven: Yale Univ. Press, 1960. Pp. 15-64.
- RYAN, T. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, 56, 26-47.
- SAKODA, J. M., COHEN, B. H., & BEALL, G. Test of significance for a series of statistical tests. *Psychol. Bull.*, 1954, 51, 172-175.
- WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.
- YARYAN, RUBY B., & FESTINGER, L. Preparatory action and belief in the probable occurrence of future events. *J. abnorm. soc. Psychol.*, 1961, 63, 603-606.
- ZAJONC, R. B. The concepts of balance, congruity, and dissonance. *Publ. Opin. Quart.*, 1960, 24, 280-296.
- ZIMBARDO, P. G. Involvement and communication discrepancy as determinants of opinion conformity. *J. abnorm. soc. Psychol.*, 1960, 60, 86-94.

(Received September 4, 1962)