

EFFECTIVENESS OF DEBRIEFING FOLLOWING DECEPTION EXPERIMENTS

ELAINE WALSTER

University of Rochester

ELLEN BERSCHIED, DARCY ABRAHAM, AND VERA ARONSON¹

University of Minnesota

This study was designed to test the notion that deception experiments might be especially disturbing to a person already concerned about the area involved in the deception. Specifically, it was hypothesized that if a person who was extremely concerned about his adequacy in social situations was given false information about his adequacy as part of an experiment, it might not be easy to debrief him at the end of the experiment. In this experiment, degree of concern about one's social skills was varied in 2 ways: concern was manipulated; high-concern and low-concern Ss were selected on the basis of a personality test. Regardless of the way in which concern was varied, there is no evidence that high-concern Ss are more difficult to debrief than low-concern Ss. Unfortunately, however, there is evidence that (a) regardless of degree of concern, debriefing may not be effective immediately, and (b) debriefing may be ineffective, even after a longer time delay, for certain personality types.

In order to investigate many social psychological questions, it is often necessary to deceive subjects within the experimental context. Deception generally serves two purposes: First, in many experiments the experimenter's hypothesis would be all too apparent to the subjects if a false explanation of the experimenter's purposes were not provided. If the purpose of the experiment were clear to subjects, they might try to assist or thwart what they believe to be the experimenter's aims (Orne, 1965). A false explanation eliminates this problem, for even if subjects try to help or hurt the experimenter, they will in fact be responding to an irrelevant hypothesis. Second, deception might be required in order to manipulate the independent variable. Such deception permits the laboratory study of variables which would otherwise not be accessible to investigation.

In spite of, or perhaps because of, the regularity with which deception is used in experimentation, several reservations have been

expressed about the use of this experimental technique (e.g., Kelman, 1965; Orne, 1965; Silverman, 1965). Critics of deception experiments seem to be primarily concerned with the possibility that deception might cause the subject permanent harm. In partial response to this criticism, defenders of the deception technique have pointed out that subjects in deception experiments are almost always thoroughly debriefed at the end of the experiment. In the debriefing, the subjects are usually told the nature of the deception, the true purpose of the experiment, and the reasons why deception was necessary. Kelman (1965) suggested that "in general, the principle that a subject ought not to leave the laboratory with greater anxiety or lower self-esteem than he came with is a good one to follow." Most experimenters seem to operate under the conviction that a thorough debriefing will accomplish the goal of returning the subject to his preexperimental state.

Yet the fact that experimenters who employ deception usually take a great deal of time and care to debrief their subjects has not entirely allayed fears concerning the use of deception. The obvious question which remains is how can the researcher be certain that his debriefing has been successful? How can he be sure that when the subject walks out the door, he does not do so with lower

¹ The authors would like to express great appreciation to Bill Walster and Raymond O. Collier (who analyzed our data), and to Susan Boye, Shirley Damrosch, David Landy, and Darwyn Linder, who helped to conduct the studies reported in this paper and wish to acknowledge the many helpful comments and suggestions made by Dr. John M. Daryl. This study was supported in part by National Science Foundation Grant GS 1056.

self-esteem than when he came in? In spite of the fact that it is of crucial importance to experimenters that debriefing remove the effects of the deception, there is no experimental evidence that debriefing, in fact, does so.

Deception experiments differ so greatly from one another in the nature and degree of deception used that even the harshest critic of this technique would be hard pressed to state unequivocally that all deception has potentially harmful effects. There are, however, two frequently mentioned dangers of deception experiments, to which some experiments are more liable than others. First, some critics have voiced their concern that lying to people may lead them to lose faith in their fellow human beings. Because scientists are ordinarily highly respected, the discovery that a scientist will lie might upset subjects even more than lies told by others. Secondly, and perhaps more importantly, it has been pointed out that some deception manipulations are emotionally disturbing to a subject, and that some disturbances might not be entirely amendable by debriefing. Consider, for example, an experiment dealing with the differences between task performance of high self-esteem and low self-esteem people. High self-esteem might be manipulated by telling subjects that they are virile, creative, and personable individuals. Low self-esteem might be manipulated by telling subjects they have negative qualities. It is evident that for the duration of the experiment the subjects in the low self-esteem condition are likely to be upset. But what about *after* they are debriefed? It has been suggested by Kelman that if one is a normal person, it is probable that such a shock to one's self-concept can be overcome and the assurances of the experimenter in the debriefing session can be readily accepted. However, Kelman points out that there are undoubtedly instances in which simply telling the subject that the results of his virility-creativity-personableness test were falsified will not erase the effects of the deception.

Under what conditions might we expect debriefing to fail? Suppose that a subject is told that his test results indicate that he is not very creative. In fact, the experimenter reports, few people tested have ever scored so

low on creativity. Further suppose that this subject happened to be a budding poet who picked up his mail on his way to the experimental session and discovered the fourteenth publisher's rejection slip for his first serious effort. While pursuing his way to the experiment, the subject might quite naturally wonder if the series of rejections should be attributed to his lack of talent or to the possibility that the uncultured masses are not clamoring for sonnets about the Crimean War.

It seems quite possible that the experimenter's authoritative evaluation of the subject's creative talents would initiate in this particular subject some independent thinking during the course of the experiment. It is quite likely, for example, that this subject would try to reach some consistency between the content of the experimenter's report and his own original ideas about his level of creativity. To do this, he might well reexamine his self-concept and selectively recall past incidents, most notably the 14 rejection slips, which would agree with the experimenter's information. Memories of criticisms from friends and family, the recollection of some low grades in English composition, would, when interpreted in the light of the test results, strengthen his belief in his supposed low creativity. Consequently, at some point in the course of the experiment, the subject might decide that since his own cognitions augment the experimental evidence, what the experimenter said was true. He might even come to the conclusion that he himself had been imperceptive not to realize his lack of creativity before.

At the completion of the experiment, of course, the subject-poet would be informed that the negative evaluation he received was chosen at random, and it would be explained that it was just as likely that he could have received a neutral or favorable evaluation of his creative talents. As previously mentioned, there is ordinarily little reason for the subject not to accept completely the notion that he has been deceived, that he is not the unimaginative dullard he thought he was, and be none the worse for wear. It even seems quite probable that the subject might believe the debriefing message in its entirety, that is, that the creativity test was not genuine. In

this case, however, his own supporting and freshly organized cognitions might remain. It is still true that he *has* gotten 14 consecutive rejection slips and that he *did* receive those low grades. Consequently, it is possible that, though the specific anxieties produced by the deception might be completely removed by the debriefing, his general opinion about himself might well be lowered, and his life turned upon a new course, because of the extra thinking the manipulation initiated.

When the deception happens to strike an area of deep concern and worry to the individual, when it is likely to initiate a train of thought which would not be altered by the revelation of the deception, it is possible that the damage done to a subject by the deception might be irreversible.

The authors will report in this paper an experiment which tests the hypothesis that it will be more difficult to successfully debrief (i.e., return to his preexperimental state) a subject who has received false information on some aspect of himself about which he is currently concerned, than it will be to debrief a subject who has received information which is irrelevant to his current concerns. While the preceding discussion focused on the possible residual effects of receiving negative information, a parallel result could be expected to occur when the subject has received positive information. That is, it might also be more difficult to debrief someone who receives positive information in an area of current concern. This experiment, then, was designed to test for the existence of both positive and negative residual effects after debriefing.

METHOD

Overview

Three steps were required to test our hypothesis. It was necessary that: (a) some subjects be very concerned about their characteristics in a given area while other subjects be unconcerned. Subjects either could be selected because they did or did not manifest this concern, or concern could be experimentally manipulated. We chose both to measure preexperimental concern and to manipulate concern experimentally; (b) subjects participate in a deception experiment in which false information relevant to the area of concern is provided; (c) subjects have time to think about the false information and an opportunity to use it to organize events in their past life. We, of course, expect subjects to be more inclined to relate the false material to previous life

experiences when they are deeply concerned about the area than when they are not; (d) subjects be thoroughly debriefed and, following the debriefing, success or failure be measured.

Procedure

The 80 subjects who participated in this study were freshmen and sophomore women enrolled at the University of Minnesota.² Subjects were recruited from an introductory psychology course and from the university library. All subjects had taken the MMPI as part of the freshman testing program.

In an initial contact, subjects agreed to participate in two separate experiments. They were told that the first experiment would have an hour delay between the first and second part. Thus, a second experiment had been scheduled during this hour for their convenience. They were told they could either participate in this experiment or not as they chose. In fact, all subjects chose to participate in both experiments. In reality, of course, the two experiments were both parts of the same experiment.

The purpose of Experiment 1 was twofold: (a) to *measure* the subject's preexperimental concern about the kind of social impression she makes, and (b) to randomly assign the subject to an experimental group and *manipulate* her concern about the kind of social impression she makes. We wanted half of our subjects to be highly concerned and curious about their social abilities and the other half to be little concerned about these abilities.

Subjects arrived at Experiment 1 in groups of four, and each subject's concern about her social abilities was measured as soon as all had arrived at the experimental room. Each subject completed the Fear of Rejection Scale. According to Rosenfeld (1964), this scale measures one's concern or fears about social rejection.

The remainder of Experiment 1 was designed to manipulate subjects' concern about their social skills. This was done in the following way: Experimenter 1 told the subjects that the purpose of Experiment 1 was to find out "something about how people make first impressions and how they react to a first-impression situation." She explained to them that two college seniors would interview each subject for about 10 minutes, asking her several questions. Following the interview, the seniors would evaluate the subject. The seniors were described as outstanding students, both academically and socially, who had been doing such interviews for

²Sixteen of these subjects were discarded from our analysis: 1 subject had recently received an MMPI report in counseling, 1 subject could not understand the procedure, 2 subjects failed to answer all questions, 1 subject was very suspicious that we were deceiving him, and 11 subjects were slightly suspicious of the procedure for reasons they could not specify. We discarded any subjects who were in the least suspicious. We did, however, reanalyze the data including all subjects, and it is clear that including the data of the suspicious subjects would not alter our conclusions in any way.

about a year. The subjects were then introduced to our two interviewers, a very attractive male and female.

Subjects were interviewed one at a time, while the other subjects continued filling out forms in an adjoining room. The interviewers stood behind a one-way mirror, so that the subject could not see their facial reactions. It was the confederates' job, by the tone of their voices and their comments, to make each subject very unsure about how well she had done. In pretests the authors selected several interview questions which were difficult for subjects to answer (e.g., "How do you feel towards people who are better looking than yourself?" or "What is your favorite joke?"). The two confederates by their manner approved some answers and disapproved of other answers.

At the completion of each subject's interview, the experimenter checked to see whether the subject had been randomly assigned to a group that should have *high concern* with how they had done in the interview situation or to a group that should have *low concern* with their performance. If the subject had been assigned to a *high-concern* group, the experimenter made several comments designed to impress the subject with the fact that her performance in the ambiguous interview situation was extremely important. To point out how important interviews are in daily life, the experimenter cited statistics concerning the importance of first impressions. She explained that since automation could take over many routine jobs but could replace only a few jobs requiring personal contacts, the ability to create a good first impression would be even more important in the future. Finally, the experimenter attempted to refute any arguments that the subject might come up with to challenge her claim that the first impression she made in the interview situation was important. She noted that some subjects had felt that because they had not been warned in advance that the interview was important, or because the interview was too short, or because they were nervous, the impression they made was not a valid one. Experimenter 1 explained that since these factors were present in real life interviews, they actually increased the validity of the evaluation. If the subject had been assigned to the low-concern condition, Experimenter 1 made totally different comments after her interview. She explained that although her performance would be of interest, one could not tell too much about a person on the basis of just one interview. She said such a situation was not a very good predictor of how one did in real life and commented that at present she was just pretesting the procedure. Finally, she presented all the reasons why a person might argue that her interview performance was not a valid indicator of her personality, but this time she agreed that these objections were valid ones.

When all four subjects had completed their interviews, Experimenter 1 explained that during the next 20 minutes the interviewers would be busy evaluating the subject's performance. She told the subjects that they could leave for the second ex-

periment, but that those who had given an unusually good or an unusually poor performance would be called back to Experiment 1 for more extensive interviewing after Experiment 2 had been completed. The remaining subjects would come back simply to complete a brief questionnaire. She said that she would send a note to them sometime during the second experiment, telling them whether to plan on a long interview or a short one.

The purpose of Experiment 2 was to lead one-half of the unconcerned and one-half of the concerned subjects to believe for the duration of Experiment 2 that they possessed good social skills, and the remainder of the subjects to believe that they possessed poor social skills. To impart these beliefs to subjects, Experimenter 2 told them, when they arrived at Experiment 2, that she was interested in several aspects of sociability or "the ability of a person to enter into warm, long-lasting and meaningful relationships with other people." She said she already had secured MMPI sociability scores for all of them. After briefly discussing the omniscience of the MMPI's sociability scale, the experimenter told the subjects that many researchers were now engaged in developing a shorter test than the MMPI to measure sociability, and also a test more easily utilized with youthful test takers. The experimenter explained that today they would take a newly developed test, the "Social Aptitude-Achievement Test." The test booklet which the experimenter then proceeded to pass out was designed simply to appear to be a valid measure of sociability. It presented several social situations in which one might find oneself and asked the subject how she would respond to each situation. After approximately $\frac{1}{2}$ hour the experimenter collected the tests, and announced she would score them immediately so she could see to what extent the results agreed with those of the MMPI. Before leaving the room to score the test, Experimenter 2 gave each subject what was purported to be her MMPI report to read, while she was out scoring the tests "in case she was interested." Of course, all subjects were interested.

The MMPI report which the subject received was false. Only two alternate forms existed. By random assignment, one-half of the subjects were to receive a positive social assessment in their report. Their reports specified that the subject had scored near the ninetieth percentile on the "Sociability Scale" of the MMPI and contained a two-page personality assessment, supposedly written by a clinician at the University of Minnesota. This assessment indicated that they were "natural and balanced" in their relations with others, that they played the ascendant role in relationships with others without being overbearing, but that possibly they were not aware of the "potential richness of their skills" in the social area. Subjects assigned to the negative condition received profiles indicating that they had scored near the eighteenth percentile in sociability. The written portion of their profiles claimed that they were reluctant to become involved with others, that they could not express their real feelings to others,

and that they lacked self-insight to the point of being largely unaware of these shortcomings.

The subjects were then given time to read and to think about these clinical reviews. It was the authors' expectation that those subjects who were very concerned about how they had performed in the interview in Experiment 1 would use the MMPI personality information in an attempt to resolve the ambiguity they felt concerning their interview performance. It was thought that unconcerned subjects would be less interested in their performance (which they had been told was not a good indicator of their real personality), and thus less concerned about resolving any ambiguity they felt.

To remind subjects of Experiment 1, and to reinforce the ambiguity of evidence concerning their interview performance, Experimenter 1 walked in the room while Experimenter 2 was out of the room scoring the sociability tests, and handed a message to each subject. Presumably, this note was to tell subjects whether or not they had given an unusual performance in Experiment 1 (and thus had to return for a long interview at the end of Experiment 2) or had done about average. Actually, each subject received a note saying that her performance had been unusual. Whether it was unusually good or unusually poor was not specified. Experimenter 1 then supervised subjects until Experimenter 2 returned.

When Experimenter 2 returned she handed each subject her test scores on the Social Aptitude-Achievement Test. In all cases the scores agreed with those provided in the false MMPI report. A subject assigned to the positive condition was told that she had received a percentile score of 94. A subject in the negative condition was told she had received a score of only 20.

Once subjects had had a chance to consider the deceptive material we had provided, we were ready to debrief them. However, before we could do so, a bit more deception, designed to increase the plausibility of the situation, was necessary. We wanted Experiment 2 to seem complete in and of itself to the subject. We certainly did not want any subjects to connect Experiment 2 with Experiment 1. If we had simply explained to subjects at this point that we had been lying to them about their personality reports, many of them would have wondered what we had accomplished by lying to them. Consequently, the next step was designed to provide a rationale for Experiment 2 and to make it seem complete in and of itself by the time of debriefing.

Thus, Experimenter 2 asked the subjects to read and evaluate two case reports. In these case reports, a great deal of personality information was provided about two girls. Interspersed with the information was the fact that one of the girls was very sociable and outgoing, while the other was a very introverted person, lacking in social skills. Subjects were asked to comment upon their own reactions to the girls. Specifically, they were asked to indicate on a series of 20-point scales how much they liked the girls, to evaluate various personality characteristics of the girls, and to indicate how similar each of

the girls seemed to themselves. After the subjects had finished rating the two girls, the stage was set for the debriefing. Experimenter 2 explained that the MMPI and Social Aptitude-Achievement Test scores she had given the subjects had been entirely false. She explained that she had never seen the subject's MMPI test, and that in fact there was no such thing as an MMPI sociability scale nor a Social Aptitude-Achievement Test. Subjects were shown both the high sociability and the low sociability MMPI reports, and it was pointed out that which report the subject received had been determined by chance. This explanation took a minimum of 20 minutes. Experimenter 2 then briefly explained to subjects why they had received a false personality report. She told the subjects that the experiment was actually concerned with how high or low self-esteem individuals felt about other people with the same traits. She explained that by raising the self-esteem of some subjects and lowering the self-esteem of others and then asking them to rate similar or dissimilar stimulus girls, she could answer this question in a clear-cut way.

Any questions asked by the subjects were answered and the whole procedure was discussed with subjects until Experimenter 2 felt confident that subjects understood that the false personality reports they had received were not their own.

Measuring of Postdebriefing Effects

The relative effectiveness or ineffectiveness of our debriefing procedures was measured in the following ways:

1. Immediately after the false debriefing, Experimenter 2 explained that a knowledge of what the subjects' personalities were *really* like would be valuable to her in analyzing the data. Thus, subjects were asked to fill out a questionnaire which they believed to be anonymous. They were asked to describe how plausible the personality information they had received had been, and, more importantly, to rate themselves on several personality traits. The traits on which subjects rated themselves were: their "ability to make deep, close, lasting, mature friendships," their "popularity," their "sensitivity to the feelings of others," and their "naturalness and spontaneity with others." Subjects utilized 4-point rating scales. The higher the score, the more "sociable" the subject indicated she was. The subject's scores on each trait were summed and the total score gave us a *Sociability Index*.

2. After being released from Experiment 2, all subjects reported back to Experiment 1. Since all had been told that they had done "unusually" in Experiment 1, they all believed that they would be required to answer a great number of questions for Experimenter 1. Undoubtedly, most subjects assumed that the other subjects were "average," and were returning to Experiment 1 simply to answer a few questions. All subjects were given a questionnaire to complete. The first question asked, "How well do you think you scored on your interview in Experiment 1?" Question 2 asked, "How well do you generally do in first impression situations?"

Possible answers to these questions ranged from 0 ("I guess I did extremely poorly") to 18 ("I guess I did extremely well"). Answers on these two questions were combined to form the *Interview Performance Index*, a measure of the extent to which a subject felt that she had done well in the initial interview.

The final question on this questionnaire was essentially a manipulation check. It asked, "Is this [the interview] a good example of what you are like in real life?" Subjects could answer either "Yes" or "No." It will be recalled that subjects in the high-concern condition had been told that their interview performance was important since it was representative of them, while low-concern subjects had been told that their performance was unimportant and unrepresentative of them.

Subjects were debriefed at great length after all of these measures were collected. Since subjects viewed this experiment as an effort to find out the extent to which deception might or might not be harmful to them, virtually all of them indicated that they were happy to have participated.

RESULTS AND DISCUSSION

All of the F tests in this paper are $2 \times 2 \times 2$ analyses involving selected concern, manipulated concern, and type of personality report provided to subjects.

Manipulation Checks

In Experiment 1, we tried to make half of our subjects concerned with the quality of their interview performance, and half unconcerned with their performance. This manipulation appears to have been effective. On their final questionnaire, subjects were asked to state whether or not the interview situation was a good example of what they were like in real life. More of the manipulated high-concern than low-concern subjects felt that the interview was representative ($F = 3.40$, $df = 1/56$, $p = .07$).³

During Experiment 2, some subjects were led to believe that they had received high scores on two tests of sociability, and others were led to believe that they had received low sociability scores. Did our subjects accept this information? The evidence indicates they did. It will be recalled that as part of Experi-

³The reader might feel that a multiple-contingency test is a more appropriate method of analysis (Sutcliffe, 1957). Such an analysis gives us almost identical results to those reported above ($\chi^2 = 3.33$, $df = 1$, $.10 > p > .05$). Recent research would indicate that with a reasonable sample size there is no appreciable difference between a chi-square or an F test on discrete data (Cochran, 1950).

ment 2, subjects were asked how similar they were to each of the girls described in the two case reports. Girls in the high-sociability condition indicated that they were much more similar to the high-sociability stimulus girl ($F = 31.04$, $df = 1/56$, $p < .001$), and that they were much less similar to the low-sociability stimulus girl ($F = 80.13$, $df = 1/56$, $p < .001$) than did girls in the low-sociability condition.

Results

It was hypothesized that it would be more difficult to debrief the high-concern subjects than low-concern subjects. Thus, we expected high-concern subjects who were told they possessed good skills in Experiment 2 to overestimate their performance, even after debriefing, to a greater extent than would those low-concern subjects who were also told they possessed good social skills. Similarly, we expected high-concern subjects who were told they possessed poor social skills to underestimate their social skills, even after debriefing, to a greater extent than low-concern subjects.

In this experiment, subjects were classified as high concern in two ways: (a) *Selected concern*: concerned and unconcerned subjects were selected on the basis of their scores on Rosenfeld's Fear of Rejection test. Those subjects who scored from 35-65 on this test were classified as high in their concern about social rejection. Those subjects who scored 0-34 were classified as low in their concern about social rejection; (b) *manipulated concern*: concern was manipulated in Experiment 1 as has been previously described.

How much social skill a subject believed he possessed after debriefing was measured by his self-ratings on the Sociability Index.

From Table 1 it is clear that regardless of whether we deal with manipulated concern or selected concern, and regardless of whether we consider self-estimates on the sociability index or on the interview performance index, high-concern subjects do *not* seem to be more difficult to debrief than low-concern subjects. Selected concern and type of deception do not interact in affecting subjects' estimates of their sociability ($F = 1.52$, $df = 1/56$) or of their interview performance ($F = 5.76$, $df = 1/56$) as we predicted they would. Though the last

TABLE 1
SUBJECTS' ESTIMATES OF THEIR SOCIABILITY AND INTERVIEW PERFORMANCE AFTER
DEBRIEFING IN VARIOUS CONDITIONS

| Degree of concern | Deceptive personality information given | N | Sociability index ^a | Interview performance index ^b |
|---------------------|---|----|--------------------------------|--|
| Selected concern | Told sociable | 16 | 9.86 | 20.54 |
| | | 20 | 8.70 | 17.62 |
| High concern | Told sociable | 16 | 10.19 | 20.64 |
| | Told unsociable | 12 | 9.83 | 23.20 |
| Low concern | Told sociable | 16 | 9.93 | 20.70 |
| | Told unsociable | 16 | 8.93 | 19.78 |
| Manipulated concern | Told sociable | 16 | 10.12 | 20.49 |
| | Told unsociable | 16 | 9.31 | 19.64 |

^a A high score indicates the subject reported high ability to make friends, popularity, sensitivity, and spontaneity.

^b A high score indicates the subject estimated that she did well in the initial interview and that she generally makes a good first impression.

interaction F is significant, the mean estimates are clearly not in the predicted direction. We will discuss this interaction later in the paper. When we consider manipulated concern, the results are the same. After debriefing, high-concern subjects do not seem to be more difficult to debrief than low-concern subjects, whether we consider the Sociability Index ($F = .12$, $df = 1/56$) or the Interview Performance Index ($F = .00$, $df = 1/56$).⁴

The data obtained from this experiment leads us to believe that our hypothesis is incorrect. This belief is further strengthened by information that we have not yet presented to the reader. The study presented here is not the first time that this hypothesis was tested. In a very lengthy pilot study conducted by Abrahams (1967) the hypothesis of this paper was tested with only minor variations. (Ostensibly the Abrahams study was concerned with "problem solving ability" and "creativity" rather than with "interview performance" and "sociability.") The data from Abraham's experiment also failed to provide any evidence that highly concerned subjects are more difficult to debrief than unconcerned subjects.

⁴ On both the Sociability index and the Interview Performance index, the three-way interaction (Selected Concern \times Manipulated Concern \times Type of Personality Report) is insignificant as is the two way interaction between selected concern and manipulated concern. All four interaction F s are below 20, $df = 1$ and 56. No significant interaction was present for these variables, and it would have been difficult to know what to make of these interactions if they had been significant.

Although we feel that our hypothesis is incorrect, this does not mean that we have concluded that debriefing is uniformly effective with all subjects. Our data indicate that subjects in the various conditions do exhibit lingering aftereffects of the deception, although they are not the effects we predicted.

Let us first consider the estimates subjects made of their own sociability immediately after having been debriefed. For all subjects, it appears that the personality report had an impact which lasted at least for a few moments beyond the occurrence of debriefing. Regardless of whether concern was manipulated or selected, and regardless of level of concern, subjects who were given favorable sociability reports rated themselves significantly higher on the sociability index than did those subjects who received an unfavorable personality report (main effect $F = 6.12$, $df = 1/56$, $p < .05$). Since the type of sociability report a girl received was determined by chance, we must assume these differences are due to the fact that the subjects were not entirely disabused of the information they received in Experiment 2.

It is possible, of course, that these aftereffects are due to the fact that our experimenters were simply poor debriefers. However, this possibility does not reduce our concern about these results for two reasons: (a) In the Abrahams (1967) experiment four different experimenters conducted the variation of the study reported here. These experimenters also got similar debriefing aftereffects. Those subjects who were told in Experiment

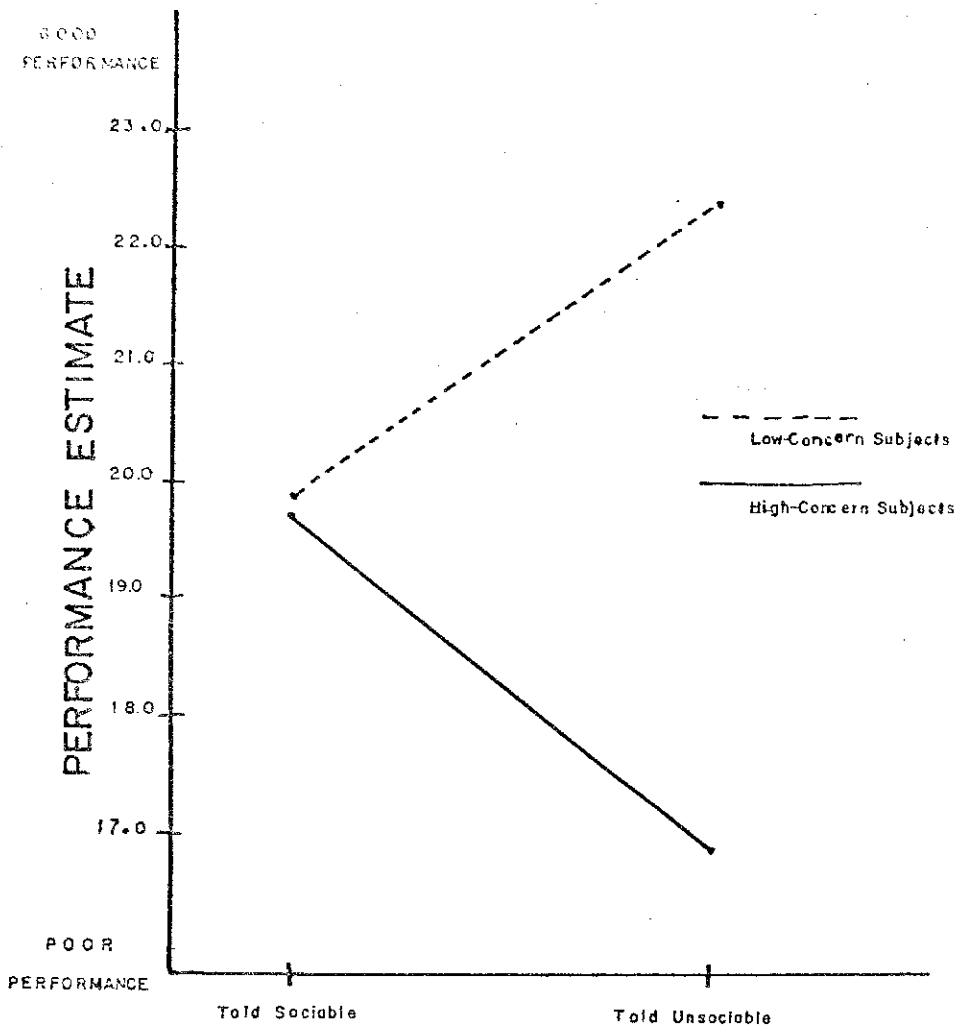


FIG. 1. Interview performance estimates by subjects with high or low fear of rejection scores.

2 that they were immature and uncreative rated themselves as less mature after debriefing ($F = 3.25$, $df = 1/63$, $p = .08$) than did subjects who were told they were mature and creative. The creativity differences were not significant. Since the results of the Abrahams study are so similar to those obtained in this experiment, and since four experimenters were involved in the Abrahams experiment, it is hard for us to believe that our results are simply due to poor debriefing techniques; (b) we, like many other experimenters, believe that we make an unusual effort to instruct our students in debriefing procedures. The six experimenters participating in the two studies discussed in this paper were all grad-

uate students who had been trained to debrief subjects and who had conducted several other experiments. Obviously, doing a good job of debriefing was of unusual importance to them in this experiment, and it will be recalled that Experimenter 2 followed a lengthy debriefing procedure.

When we look at the difference in the interview performance estimates between those subjects who received good sociability reports in Experiment 2 and those subjects who received poor reports, we see that the high-sociability-condition subjects do not think that they performed better than do subjects in the low-sociability condition ($F = .22$, $df = 1/56$). Debriefing does not seem to be

ineffective here. This interview performance estimate differs in two main ways from the sociability estimate. First, the interview estimates were taken a longer period of time after debriefing than the sociability estimates. Perhaps it takes a few minutes for debriefing to "sink in." Secondly, the Interview Performance Index may measure a different aspect of sociability than the Sociability Index.

In addition to the aftereffects previously discussed, there is one additional finding that seems worth commenting upon; though our manipulated-concern measure seemed to have little impact upon the success or failure of debriefing, there is some evidence that selected concern may be of importance.

From Figure 1 and Table 1 it is clear that a subject's initial degree of concern (with her social abilities) has a marked effect on her postdebriefing estimate of her interview performance.

When we consider only the data from *selected high-concern* subjects, we see that even after the passage of time, debriefing does not seem to be totally effective. Selected high-concern subjects who received good sociability reports in Experiment 2, even after being told these reports were false, estimated that they did better in the interview situation than did subjects who received poor sociability reports. *Selected low-concern* subjects who received good sociability reports estimated their interview performance very much as did high-concern subjects who received the good report. However, the interview performance estimates of the low-concern subjects who received a poor sociability report are markedly different from the estimates of comparable high-concern subjects. These low-concern subjects guessed they did *better* in the interview situation than did subjects in any other group. This finding is peculiar.

In an attempt to explain this result, our first thought was to consider carefully what selected high concern and selected low concern must mean. According to Rosenfeld, the Fear of Rejection test was constructed to parallel the structural form of Mandler and Sarason's Test Anxiety Questionnaire and, presumably, measures the extent to which a subject is concerned with social rejection. From an analysis of our data, however, it

appears that in many ways the Fear of Rejection test is a general measure of self-esteem. Subjects classified as low concern by this test tend to say better things about themselves than do high-concern subjects. For example, on both the Sociability Index and on the Interview Performance Index, there is a main effect due to initial level of selected concern of the subject. Subjects low in fear of rejection claim to be more sociable ($F = 6.06$, $df = 1/56$, $p < .05$) and to have done better in the interview situation ($F = 6.20$, $df = 1/56$, $p < .05$) than do subjects high in fear of rejection.

If low concern does in fact reflect high self-esteem, perhaps we can find an explanation for these findings. The apparent failure of debriefing for high-concern subjects is disturbing, but comprehensible. What is peculiar is the high performance estimates made by low-concern subjects who were given low personality results. Perhaps this simply demonstrates that high self-esteem individuals are especially likely to reject unpleasant information about themselves. Cohen (1959), suggested that

different self-esteem groups are differentially able to fulfill the important acquired motive of maintenance of self-esteem at the highest possible level. Through their use of avoidance defenses . . . the highs are able to maintain a high-level equilibrium . . . whereas the lows are more dependent upon experimental variation [p. 117].

Cohen further noted that the high self-esteem individual may "emphasize enhancing experiences, thereby preserving an insulated but positive self-picture [p. 117]." Perhaps our low-concern (high self-esteem) subjects, when faced with a low personality report, simply enhanced their performance in the interview situation in order to maintain their self-esteem.

This explanation is not entirely satisfactory. If Cohen's formulation is applicable there should be *some* evidence in our experiment that the selected low-concern subjects were "less receptive to experimental variations" than selected high-concern subjects, as Cohen suggests is the case. Low-concern subjects seem, during Experiment 2, to accept the negative personality reports to the same extent as do high-concern subjects. Acceptance of the negative or positive evaluations

was measured by the extent to which subjects indicated that they were similar to a sociable or unsociable stimulus girl. Low-concern subjects who were told they had low sociability scores were just as likely to agree that they were similar to a low sociability girl as were high-concern subjects (interaction $F = .03$, $df = 1/56$). Further, even immediately after debriefing, there is no evidence that low-concern subjects, who were told in Experiment 2 that they had low sociability scores, were especially anxious to inflate their estimates of their own sociability (interaction $F = .62$, $df = 1/56$). It was not until they were quizzed several minutes after debriefing about their performance in another situation that low-concern-poor-evaluation subjects overestimated their performance.

We conclude two things from the preceding study:

1. The question of whether or not it is more difficult to successfully debrief concerned subjects than unconcerned subjects of information relevant to their current concerns remains unanswered. We can only say that two lengthy attempts to demonstrate this effect have been unsuccessful. Whether or not a stronger concern manipulation would produce the effect is, of course, a moot question, but we have been unable to produce evidence of even a slight tendency for subjects to behave in the predicted manner.

2. We have presented evidence that debriefing might not be as immediately effective as experimenters have hoped and assumed. This evidence is distressing for a number of reasons. First of all, it is disturbing that in the present experiment and in the Abrahams experiment, even after a very lengthy and thorough debriefing (probably atypical in thoroughness), subjects still behaved to some extent as though the debriefing had not taken place. Subjects behaved in this manner even though they had voiced to the experimenter their understanding that the manipulation was false, their understanding of the true purpose of the experiment, and even though, by their manner and replies, the experimenter had been satisfied that they did indeed understand the nature of the deception.

Even more disturbing is the evidence that the aftereffects of debriefing might be com-

plex, unpredictable, and may depend in part upon the personality traits of the subjects. The nature of the effect of personality traits in the present experiment were not totally explicable to us. Aftereffects in the Abrahams experiment were also present and somewhat inexplicable. The success of debriefing in that experiment was influenced by several significant interactions between sex of subject, sex of experimenter, and treatments. At the time the Abrahams experiment was run, we were willing to conclude that the significant interactions obtained were perhaps due to chance, to experimental error, to measurement error, and so on. The results of the present experiment, however, combined with the results of the Abrahams experiment, have aroused our suspicions and anxiety that there are often residual effects of debriefing, and that these effects appear to be complex and not easily interpreted. What is needed now is the generation of hypotheses and experimentation which would help us predict exactly what effects can be expected, with what kinds of people under what types of conditions, as well as an investigation of the effectiveness of various types of debriefing procedures.

REFERENCES

- ABRAHAMIS, D. The effect of concern on debriefing, following a deception experiment. Unpublished master's thesis, University of Minnesota, 1967.
- COCHRAN, W. G. The comparison of percentages in matched samples. *Biometrika*, 1950, 37, 256-266.
- COHEN, A. R. Some implications of self-esteem for social influence. In I. L. Janis and C. I. Hovland (Eds.), *Personality and persuasibility*. New Haven: Yale University Press, 1959. Pp. 102-137.
- KELMAN, H. C. The human use of human subjects: The problem of deception in social-psychological experiments. Paper presented at the meeting of the American Psychological Association, Chicago, 1965.
- ORNE, M. T. Demand characteristics and their implications for real life: The importance of quasi controls. Paper presented at the meeting of the American Psychological Association, Chicago, 1965.
- ROSENFELD, H. M. Social choice conceived as a level of aspiration. *Journal of Abnormal and Social Psychology*, 1964, 68, 491-499.
- SILVERMAN, I. Motives underlying the behavior of the subject in the psychological experiment. Paper presented at the meeting of the American Psychological Association, Chicago, 1965.
- SURCLIFFE, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychological Bulletin*, 1957, 54, 134-317.

(Received August 25, 1966)