Reassessing the Validity of Laboratory-Produced Attitude Change

Gary L. Wells
Ohio State University

In their 1974 study, Hendrick and Seyfried found that experimental group subjects given an attitude change treatment (reading a persuasive essay) were more attracted to a stimulus stranger who espoused their posttreatment attitudes than to a stimulus stranger who espoused their pretreatment attitudes. Control subjects, however, remained more attracted to a stimulus stranger espousing their pretreatment attitudes. Hendrick and Seyfried concluded that this attraction transfer demonstrated the validity of laboratory-produced attitude change. The present study added a new experimental group, which did not receive the posttest, in an attempt to eliminate potential demand and/or evaluation apprehension interpretations, which are applicable to Hendrick and Seyfried's results. Hendrick and Seyfried's results were replicated with the original experimental group but not with the new experimental group. It is suggested that Hendrick and Seyfried's results cannot be interpreted as indicative of genuine attitude change.

Byrne (1971) and his associates have consistently demonstrated that people are more attracted to a stranger whose attitudes are similar to their own than to a stranger with dissimilar attitudes. On the basis of this finding, Hendrick and Seyfried (1974) presented an ingenious solution to the problem of demonstrating the validity of laboratory-produced attitude change. They reasoned that genuine attitude change should result in subjects being more attracted to a stimulus stranger who espouses their posttest opinions than to a stimulus stranger who espouses their prettest opinions.

To test this hypothesis, Hendrick and Seyfried (1974) yoked subjects on the basis of their responses to a pretest opinion survey. In a 2 x 2 design, one half of the subjects were assigned to the experimental group, which received a persuasive essay designed to affect attitudes on a student opinion survey, and each experimental subject's yoked partner was assigned to a control group. Later in the quarter, the experimental group received a persuasive communication and again filled out the student opinion survey (the control subjects did not meet that day). The next day, all subjects rated two stimulus strangers who were individually designed for each subject. In the experimental group, the two stimulus strangers exactly matched each experimental subject's own pretest (pretest stimulus stranger) and posttest (posttest stimulus stranger) responses to the student opinion survey. Each control subject's pretest and posttest stimulus strangers were exactly the same as their individually yoked partner's pretest and posttest strangers from the experimental group. As predicted, Hendrick and Seyfried found that the control subjects were more attracted to the pretest stranger than to the posttest stranger, whereas the experimental subjects were more attracted to the posttest stranger than to the pretest stranger. They concluded that the essay-induced attitude scale change was genuine.

There is an important assumption that underlies the Hendrick and Seyfried prediction. This assumption is that the similarity—attraction situation employed is itself free of demand characteristics (Orne, 1962), evaluation apprehension (Rosenberg, 1969), and other artifacts. Suppose, for example, that the posttest attitude scale responses were artificial, due perhaps to apprehensive subjects having expressed an attitude they believed
would make them look good and/or having responded to the demand characteristics of the situation. When they participated in the experiment the second day, the subjects might still have been trying to look good to the experimenter, and the appropriate strategy would have been for the subjects to rate the opinions they recently expressed to the experimenter as having desirable qualities (intelligent, adjusted, etc.). Thus, for the experimental group subjects, it would have been logical to have rated the posttest stranger (who exactly matched each subject's own most recent expression of opinion) in a highly desirable manner. Perhaps the subjects even felt that they were evaluating themselves.

It is not necessary to assume that the expressions of opinion toward the strangers were sincere or genuine. Note that the subjects in the experimental group rated a stranger whose responses were an exact replica of their own recent posttest. This by itself might elicit suspicion. Byrne and his colleagues have abandoned this perfect matching procedure (called "identity mirror" stimulus pattern) in favor of a "constant discrepancy" pattern primarily because of the potential extraneous factor of suspicion (Byrne, 1971, p. 16). It should be noted that Hendrick and Seyfried did check on suspicion by asking their subjects to write a paragraph giving their own conception of what the experiment was about. However, such questionnaires are of unknown validity, and Page (1973) has recently found that questionnaires tend to score aware subjects as unaware. In addition, questions that ask subjects to generate hypotheses may falter because the experimenter ignores hypotheses that are incorrect even though a given subject's hypothesis may have led to the same results as the experimenter's hypothesis.

One way to eliminate the possibility that subjects in the experimental group are trying to make their stated (but not necessarily genuine) opinions look good is to eliminate the posttest. If it is not genuine attitude change, experimental group subjects will not prefer the posttest stranger, since they have not filled out the posttest questionnaire and thus cannot make themselves look good with high ratings of the posttest stranger. However, if the findings of Hendrick and Seyfried were not due to evaluation apprehension, the results should not be changed by the elimination of the posttest. The present study was designed to test this idea. To do this, it was necessary to yoke three subjects and assign the third yoked partner to an experimental group that did not receive a posttest. The posttest stimulus stranger for each set of three subjects was based on the posttest responses of the experimental group member who did receive the posttest. Note that this three-group design allows a direct replication of Hendrick and Seyfried's two-group design and yields a more appropriate comparison between the control group and the experimental-no-posttest group.

Thus, the design is a 3 X 2 mixed factorial. The between-subjects variable was exposure to a persuasive message followed by a posttest (experimental-posttest group) versus exposure to an irrelevant message (control group) or exposure to the persuasive message not followed by a posttest (experimental-no-posttest group). The within-subjects variable was the pretest and posttest stimulus strangers' opinions.

A preference for the posttest stranger in both experimental groups would be relatively strong evidence of attraction transfer and therefore the validity of attitude change. If the experimental-posttest subjects preferred the posttest stranger and the experimental-no-posttest subjects preferred the pretest stranger, this would indicate that Hendrick and Seyfried's findings should not necessarily be taken as evidence for the validity of laboratory-produced attitude change.

**Method**

**Subjects**

The subjects were students in several sections of introductory psychology at Ohio State University and participated in the study in partial fulfillment of the course requirements. In keeping with Hendrick and Seyfried's procedure, those who responded favorably to at least four of the five student opinion survey items and no less than uncertain on the fifth (87% of all subjects) were selected.

At the beginning of the quarter, 173 subjects took the pretest, but only 135 subjects (45 per cell) satisfied both the attitude response criterion above and the yoking criterion, to be described in the Pro-
procedure section. On the first day of the experiment, 10 more subjects per cell were eliminated because of a failure of 1 of the yoked members to attend. The total number of subjects was 35 per cell.

Materials

The current study employed a modification of the stimulus materials and dependent variables materials used in the Hendrick and Seyfried study. These included a counterattitudinal essay arguing against students voting in local elections, a five-item Likert-type scale called student opinion survey, and an interpersonal rating scale. Modifications were necessary in Hendrick and Seyfried's persuasive message and the student opinion survey because the items were prepunished in Kent, Ohio. In addition, pilot testing indicated that the message failed to produce effects of the magnitude observed in the Hendrick and Seyfried study. Therefore, fake statistics about student lack of knowledge were added, as well as a line attributing the message to a "professor in political science." This modification resulted in significant differences between control and experimental groups with a 24-hour delayed posttest (X = 22.4 and 16.4, respectively).

The five student opinion survey items were statements about students voting in local elections (e.g., all students 18 years of age and older who reside in Columbus should be allowed to vote in Columbus elections.) The subjects responded on a five-point Likert-type scale with endpoints labeled strongly agree and strongly disagree. This study also employed an irrelevant message on day-care centers for the control group and a questionnaire on day-care centers for the control and experimental-no-posttest groups.

Procedure

At the first part of the quarter, all subjects filled out the five-item student opinion survey, which was embedded in a larger survey. The five items appeared together on page 3 of a five-page survey. Insofar as possible, triads of subjects were yoked on the criterion of identical response patterns on the pretest. Perfect matching was possible for 32 of the 35 subjects per cell. The remaining subjects varied by only one response alternative on a single item. One member of each triad was randomly assigned to each of the three conditions.

Later in the quarter, the subjects were called and asked to participate in a person perception study. The subjects whose yoked partner(s) did not attend (n = 19) the experiment were given the student opinion survey again to check on possible shift in attitude from the first part of the quarter. Group size ranged from 21 to 27 subjects per session, and each of the five sessions had from 6 to 9 subjects in each condition.

Upon arrival, the control subjects read the essay arguing for the establishment of a day-care center for the children of graduate students and faculty (irrelevant message), which was attributed to a professor in political science. The subjects then rated their impressions of and attraction to the professor on a form attached to the end of the essay and filled out the questionnaire on day-care centers. The experimental-posttest group (designated to replicate Hendrick and Seyfried's findings) read the essay arguing forcefully against student voting in local elections (relevant message), which was attributed to the same political science professor, and subsequently rated their impressions of and attraction to the professor. After these ratings were completed, the five-item student opinion survey (posttest) was again administered, with the written explanation that knowledge of the subject's own reactions was necessary because "your attitudes can influence your responses to the person who wrote the essay." This survey used exactly the same format and instructions as in the pretest. The experimental-no-posttest group was given the same treatment as the experimental group just described except that the day-care survey (irrelevant questionnaire) was administered with the written explanation that "information is also needed on an unrelated project." This irrelevant questionnaire was included to keep the experimental-no-posttest subjects from finishing earlier than the subjects in the other two groups.

The subjects returned the next day and were told, "Yesterday you formed an impression of a person based on an essay he wrote. Today we wish you to form an impression of two more people based on the attitudes they express." Booklets were distributed that contained two completed student opinion surveys (counterbalanced for order), one representing the pretest stranger and the other the posttest stranger. The subjects evaluated the stimulus strangers' intelligence, knowledge of current events, morality, and adjustment. Two items measured attraction: "How much do you think you would like this person?" and "How much would you enjoy working with this person?" Two other items measured perceived similarity: "How similar are your attitudes to this person's attitudes?" and "Excluding this person's attitudes, how similar do you think you would be in other areas, such as interests, values, personality, etc.?" All items were measured on 7-point scales with endpoints appropriately labeled. The items measuring evaluation of stimulus strangers were identical to those employed by Hendrick and Seyfried.

After the ratings were completed, the subjects were asked to try to "fill out the five-item attitude scale exactly as you recall doing at the beginning of the quarter." This question was included because even though Hendrick and Seyfried did not get a measure of recall, they stated that "it seems likely that such recall for the experimental group would yield a mean closer to posttest than to initial pretest attitudes" (p. 870). In addition, the subjects were asked to write a brief paragraph giving their
conception of what the experiment was about. They were then fully debriefed as to the nature of the experiment and were asked to tell no one of the experiment for 2 weeks.

**Results**

**Attitude Change**

The mean scores on the student opinion survey are given in Table 1. These scores are based on the sum of the five items for each subject and thus could range from 5 to 25. The pretest scores for the subjects in all three groups were virtually identical (22.3, 22.4, 22.4), as should be the case due to yoking. The possibility that the subjects in the control group would shift their attitude responses from the first part of the quarter was tested by observing such shift in the subjects who filled out the pretest but whose yoked partner(s) did not attend. The data for these subjects indicated no opinion shift for the group \(n = 19\) mean from pretest to posttest \(X = 22.5\) for both measures. In addition, a 5 (Items) \(\times\) 19 (Subjects) matrix of responses indicated zero shift in 82 of the 95 cells. Thus, attitude stability for the control subjects seemed adequate.

The pretest to posttest shift for the experimental-posttest group indicates that the essay did affect their attitude scale responses \(X = 22.3\) and 16.2, respectively, and \(F(1, 68) = 9.34, p < .01\). It should be recalled here that this study used the student opinion survey score of each subject in the experimental–posttest group as the posttest stranger for their yoked partner in the experimental–no-posttest group. In having done this, this study assumed that the postpersuasion opinion for each experimental–no-posttest subject was closely approximated by the posttest score of that subject's yoked partner from the experimental–posttest group. An estimate of how well a given subject's actual posttest score can predict that subject's yoked partner's posttest score was obtained by an internal analysis of the experimental–posttest group. Each subject in the experimental–posttest group was yoked to another member of the experimental–posttest group on the basis of pretest scores. Of the 35 subjects, 30 met the yoking criteria described earlier with at least one other subject. Of these 30 subjects, 26 (86.6%) had actual posttest scores that were closer to their own yoked partner's posttest than to their own pretest. Thus, since the experimental–no-posttest subjects' posttest stimulus stranger was based on their yoked partners' posttest score, 86% of the experimental–no-posttest subjects should have given higher ratings of attraction to the posttest stimulus stranger than to the pretest stimulus stranger.

**Attraction, Similarity, and Evaluation Scores**

The two attraction items and the two similarity items were correlated .72 and .67, respectively. Thus, as in the Hendrick and Seyfried study, the two attraction items were summed to form a single attraction score and the two similarity items were summed to form a single similarity score. Table 2 gives all six measures for all conditions along with the results obtained by Hendrick and Seyfried. It should first be noted that the data pattern of Hendrick and Seyfried replicated well. In the control condition, the subjects rated the pretest stranger more favorably on intelligence and attraction and felt the pretest stranger to be more similar to themselves than the posttest stranger. Consistent with Hendrick and Seyfried, the subjects in the experimental condition who had received a

---

Note. Pretest = before essay and posttest = after essay.

### Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Experimental with posttest</th>
<th>Control</th>
<th>Experimental without posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
<td>Pretest</td>
</tr>
<tr>
<td>Wells (1976)</td>
<td>22.3</td>
<td>16.2</td>
<td>22.4</td>
</tr>
</tbody>
</table>

2 Hendrick and Seyfried's data indicate no shift in 101 of 145 cells.
TABLE 2
MEAN ATTRACTION, SIMILARITY, AND EVALUATION SCORES FOR HENDRICK AND SEYFRIED AND CURRENT REPLICATION FOR ALL CONDITIONS

<table>
<thead>
<tr>
<th>Item</th>
<th>Experimental with posttest</th>
<th>Control</th>
<th>Experimental without posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
<td>Pretest</td>
</tr>
<tr>
<td>Attraction</td>
<td>8.2a</td>
<td>10.1b</td>
<td>10.5b</td>
</tr>
<tr>
<td></td>
<td>(9.3a)</td>
<td>(11.3b)</td>
<td>(11.6a)</td>
</tr>
<tr>
<td>Intelligence</td>
<td>4.4a</td>
<td>5.4b</td>
<td>5.5b</td>
</tr>
<tr>
<td></td>
<td>(4.8a)</td>
<td>(5.4b)</td>
<td>(5.6a)</td>
</tr>
<tr>
<td>Knowledge of current events</td>
<td>4.5a</td>
<td>5.4b</td>
<td>5.5b</td>
</tr>
<tr>
<td></td>
<td>(4.9a)</td>
<td>(5.2b)</td>
<td>(5.6a)</td>
</tr>
<tr>
<td>Morality</td>
<td>4.4a</td>
<td>5.2b</td>
<td>4.2a</td>
</tr>
<tr>
<td></td>
<td>(4.4a)</td>
<td>(5.3b)</td>
<td>(5.4a)</td>
</tr>
<tr>
<td>Adjustment</td>
<td>4.7a</td>
<td>5.6b</td>
<td>4.9b</td>
</tr>
<tr>
<td></td>
<td>(4.8a)</td>
<td>(5.4b)</td>
<td>(5.7a)</td>
</tr>
<tr>
<td>Similarity</td>
<td>8.1a</td>
<td>11.3b</td>
<td>10.6b</td>
</tr>
<tr>
<td></td>
<td>(8.3a)</td>
<td>(12.1a)</td>
<td>(11.7a)</td>
</tr>
</tbody>
</table>

Note. Pretest = evaluation of pretest stranger and posttest = evaluation of posttest stranger. The values in parentheses represent the results obtained by Hendrick and Seyfried (1974). Within each of the 3 X 2 analyses, means with no common subscripts differed at the .05 level by the Newman-Keuls test.

posttest rated the posttest stranger higher on attraction, similarity, and intelligence.

Analysis of variance indicated a significant Experimental Condition X Stimulus Stranger interaction on all six dependent variables: attraction, $F(2, 102) = 7.94$; intelligence, $F(2, 102) = 7.15$; knowledge of current events, $F(2, 102) = 9.33$; mortality, $F(2, 102) = 13.12$; adjustment, $F(2, 102) = 10.67$; similarity, $F(2, 102) = 10.27$—with $p < .05$ for all measures. In addition, a Newman-Keuls test was performed within the 3 (Conditions) X 2 (Pretest, Recall of Pretest) analyses for each dependent variable question. These results show that the experimental–posttest group's pattern of significance generally replicates the pattern of Hendrick and Seyfried's (1974) study. Most importantly, however, for the experimental group that did not receive a posttest, the ratings of attraction, similarity, and intelligence are significantly higher for the pretest stimulus stranger than for the posttest stimulus stranger. In addition, the experimental–no-posttest group did not significantly differ from the control group on any of their ratings of the pretest and posttest strangers.

Recalled Pretest Attitudes

The attitude each subject recalled holding on his pretest was computed by summing across the five attitude items. Mean recall scores for the experimental–posttest, control, and experimental–no-posttest groups were 19.2, 23.0, and 21.1, respectively. A 3 (Conditions) X 2 (Pretest, Recall of Pretest) analysis of variance revealed a significant interaction, $F(2, 102) = 7.23$, $p < .05$. A Newman-Keuls comparison of pretest to recall for each group indicated that recall for the control and experimental–no-posttest groups did not significantly differ from actual pretest scores ($p = .35$). The experimental–posttest group's pretest recall, however, approached a difference that was nearly significant from their actual pretest score ($p = .06$).

Suspicion

Each subject's paragraph on what they thought the experiment was about was scored by three raters (blind to condition) according to whether the subject mentioned that the experiment may have been interested in attraction (liking, rating, etc.) as a function of attitude change. All three raters independently indicated that four subjects generated this idea and all four subjects were in the experimental–posttest group. In a binomial distribution (the subjects could either generate the hypothesis or not), the probability of observing all four cases in only one of the three conditions by chance is small ($p = .03$).
DISCUSSION

The results of this experiment suggest that Hendrick and Seyfried's results are replicable but may be due to an artifact. The experimental subjects who did not get a posttest failed to rate the posttest stimulus stranger in a desirable manner. In addition, the pilot data preclude an argument that the experimental–posttest subjects were committed as a function of filling out the posttest, and the experimental–no-posttest subjects were free to "slip" back to their initial attitude during the 24-hour interval between reading the persuasive message and the attraction measure. The mean opinion in this study without delay of posttest was almost exactly the mean obtained in the pilot testing with a 24-hour delay (16.2 and 16.4, respectively). Instead, it appears that the elimination of a posttest reduced the differential evaluation apprehension between experimental and control conditions. That is, the experimental subjects did not have a posttest response that they could have tried to make look good by giving high ratings. The fact that the experimental-posttest subjects were the only ones to generate comments regarding both attitude change and attraction also suggests that differential cues were available for conditions. As noted earlier, however, there is generally a false-negative bias in suspicion measures, and there were no doubt more suspicious subjects. This index simply reflects the differential suspicion between conditions.

The recall data are also not supportive of genuine attitude change. Hendrick and Seyfried noted that the experimental subjects, after attitude change, just about "disowned" a stranger represented by their pretest attitude and that "this much of a shift would not be expected unless subjects had largely forgotten their original prepersuasion attitudes" (1974, p. 870). However, in this study, the experimental–no-posttest subjects' recall did not significantly differ from their actual pretest opinions. For the experimental–posttest subjects, the recall of pretest attitude was significantly different from their actual pretest attitude; however, their recall was equally deviant from their actual posttest attitude, and it is possible that the experimental–posttest subjects, having highly evaluated the posttest stranger, distorted their recall intentionally in favor of the posttest stranger.

These results may appear disappointing to attitude change researchers. However, the conclusion is not that laboratory-produced attitude change is invalid. The conclusion is that the important study conducted by Hendrick and Seyfried provides equivocal evidence for the validity of laboratory-produced attitude change.

REFERENCES


(Received October 15, 1975)