Journal of Personality and Social Psychology
1969, Vol. 11, No. 2, 177–186

SOCIAL PSYCHOLOGY OF A CLASSICAL CONDITIONING OF ATTITUDES EXPERIMENT [1]

MONTE M. PAGE [2]

University of Nebraska, Lincoln

The Staats and Staats deception experiment for conditioning attitudes toward nonsense syllables is reexamined in the context of the social psychology of experiments. It is proposed that the "conditioning" effect was due to subjects becoming aware of demand characteristics and cooperating with them. Contrary to the results of Insko and Oakes, demand awareness is found to correlate better with the conditioned ratings than does contingency awareness. A direct test of learning of the associations between the visual and the spoken word lists also correlates highly with "conditioning" and the two awareness measures. The independent variable of psychological sophistication of subjects (time of semester) and a manipulation of the difficulty of making interlist associations are also significant as predicted. Results are interpreted as challenging a conditioning interpretation of the Staats' effect. The effect seems to be an artifact of the demand characteristics of the experimental situation.

Staats and Staats (1957, 1958) introduced a complex deception experiment which they claim results in the classical conditioning of an evaluative meaning or attitudinal response. Their subjects were told that they were participating in a learning experiment designed to study ability to learn two lists of words simultaneously. One list of six words was presented visually and was repeated many times. Immediately after each visual word, a word was spoken from another much longer list. Each of the spoken words was repeated only once. This procedure was intended to disguise the repeated contiguous association of spoken words having strong positive evaluative meaning (beautiful, healthy) with one of the visual words; and the association of strong negative evaluative meaning (ugly, sick) with another visual word. The other four visual words were included to disguise the associations and were always followed by spoken words of neutral evaluative meaning.

After "conditioning," subjects rated the six visual words on pleasant-unpleasant semantic differential scales. Instructions as to the purpose of the ratings were deceptive. Then subjects wrote down what they thought the purpose of the experiment was. Subjects verbalizing awareness of the relationship between the two lists were dropped from the analysis. The Staats' concern as to subjects' awareness of the association between the lists was a recognition that such awareness might invalidate a classical conditioning interpretation of the results.

Two studies (Cohen, 1964; Insko & Oakes, 1966) have subsequently challenged the Staats' conclusion that their result occurred "without awareness — without cognition [Staats & Staats, 1957]." The issue raised by these studies was in the context of the learning and contingency awareness controversy (Spielberger, 1962; Spielberger & DeNike, 1966). Cohen used the Staats' single-question awareness measure with a more stringent criterion for classifying a subject as unaware. Insko and Oakes used a more extended postexperimental questionnaire. Both studies found strong relationships between postexperimental assessment of awareness of the relation between the two lists (contingency awareness) and the so-called conditioned attitudes. These studies seem to challenge sufficiently the Staats' conclusion as to learning without contingency awareness. However, since neither study obtained evidence as to when the contingency awareness occurred, the correlations by themselves are consistent with two interpretations. Awareness could be causally linked to the marking of the rating scales, or the experimental effect (conditioning) could be

[1] This research was supported in part by a grant from the University of Nebraska Research Council.

[2] Requests for reprints should be sent to Monte Page, Department of Psychology, University of Nebraska, Lincoln, Nebraska 68508.

causing the postexperimental awareness. What is still needed is evidence as to which came first, the awareness or the conditioning.

Assuming for the moment that contingency awareness precedes the marking of the rating scales, and this can be tested by asking aware subjects to pinpoint the time during the experiment when they became aware, there is a broader context in which the attitude conditioning effect may be interpreted. That context is the social psychology of experiments (Orne, 1962; Page, 1968; Silverman, 1968). Contingency awareness may be only the first step in the complex sequence which brings about the "conditioned ratings." Perhaps the crucial factor in mediating the ratings is awareness of demand characteristics, or discovery of the purpose of the experiment, bringing about compliant or cooperative role behavior on the part of the subject (Page, 1968). That such factors are potentially very powerful mediators of behavior in human experiments is implicitly recognized by many experimenters (including the Staats') in that they often go to great lengths to disguise the true purpose of the experiment. In this context the important question with regard to attitude conditioning has to do with whether the deception was effective on all subjects; and if not did those who saw through the deception behave any differently than those who didn't (Kelman, 1967).

In the context of Orne's (1962) demand-characteristics formulation, there could be an alternative explanation of why the Staats' obtained their results. While their subjects were told to learn two separate lists of words, some of them did not follow directions; instead they learned or noted the consistent relation between pleasant and unpleasant spoken words and the visually presented syllables. For these contingency aware subjects the situation changed so that they no longer perceived themselves as participating in a rather dull and routine learning experiment. The experimenter had not fully explained; he must have something subtle and deceptive in mind. With this appraisal of the situation, they may have been set to search for other unexplained relations and to mistrust further instructions; they may even have conceived of the experiment as a test of their intelligence or problem-solving ability. With this or a similar set of expectancies, the aware subjects were presented the rating scales. At that point the purpose of the experiment became apparent to most of them; that is, "The experimenter wants to know if I caught on to the pleasant or unpleasant syllables and this is his way of testing," or "He was trying to condition my feelings," etc. In any case, some of the contingency aware subjects came to believe that the purpose of the experiment was for them to rate the crucial syllables in opposite directions and to the extremes of the scales. Having this knowledge of the purpose of the experiment (demand awareness) each subject then was faced with the problem of whether to cooperate with the experimenter and respond in the way that so obviously was demanded by the situation, or to resist the influence. In the majority of cases, the decision was made to cooperate (Page, 1968) and hence the "conditioned response."

Two previous studies (Cohen, 1964; Page, 1964) have found the distribution of subjects' ratings in attitude conditioning to be peculiarly bimodal. Subjects either rated the critical syllables as negatively or positively as possible or appeared, as a group, not to be affected. The difference between group means was produced by a few subjects conditioning a good deal rather than most of them conditioning a little. In another study (Page, 1968) concerned with awareness of demand characteristics in a deceptive learning situation, a similar bimodal distribution was found. The upper mode of the distribution in that study was populated entirely by subjects claiming to have been aware of and cooperating with demand characteristics. It may be that the bimodal distribution is typical of data mediated by demand characteristics. If so, the Staats' data is suspect. Also, not all subjects who verbalized contingency awareness in the previous studies showed the extreme conditioned rating. This suggests that something in addition to contingency awareness, possibly demand awareness plus motivation to cooperate, is necessary to produce the extreme ratings.

Insko and Oakes (1966) included measures of both contingency and demand awareness in their study. On the basis of their data they concluded that the concept of demand aware-

ness does not add anything beyond simple contingency awareness. They go on to offer an explanation of the Staats' effect in terms of an aware concept formation rather than affective conditioning. Since their measure of demand awareness did not account for all the variance, they suggested that the Staats' phenomenon was not entirely an artifact of demand characteristics. In light of postexperimental interviews conducted earlier by the present author (Page, 1964) a "concept-formation only" interpretation seemed incomplete and a further investigation was considered worthwhile. For example, one female subject in that study said, "I don't really dislike the name Bill, that's my husband's name, but for purposes of this experiment, I marked Bill bad." Other subjects who were also extreme in their "conditioned attitudes" gave similar indications that their ratings were pure artifacts of demand characteristics.

Based on the above considerations, the present experiment was designed. Several predictions consistent with a demand-characteristics explanation were made. First, a strong association between carefully assessed postexperimental reports of having been contingency aware (knew the interlist association) during training and before marking the rating scales and the high conditioning scores was predicted. This is suggested by the Insko and Oakes correlations, but because they did not ask subjects to pinpoint the time at which they noted the contingency, they had no evidence as to direction of causation. Second, the awareness and conditioning association should be strengthened if demand awareness (knew the experimenter expected pleasant and unpleasant ratings) was separated from contingency awareness. This prediction is contradicted by Insko and Oakes, but in this study the questionnaire and scoring procedures were designed on the assumption that demand awareness is basically an either-or dichotomy; rather than attempting to measure awareness as a continuum as did Insko and Oakes. Third, a new "second learning test," which was a direct test of interlist associations given immediately after the rating scales, was predicted to correlate highly with the postexperimental awareness measures and the conditioned ratings.

In addition, two predictions were made with regard to independent variables. First, subject sophistication might significantly facilitate the conditioned rating behavior. This was a significant variable in a previous experiment (Page, 1968), and it is reasonable that subjects who have spent a semester listening to a psychologist lecture, reading a textbook, and participating in experiments (Holmes, 1967) would do better at figuring out what a psychologist might expect them to do in an experiment. This prediction is not necessarily required by the demand-characteristics formulation, but it is consistent with it. If evidence were found in support of this prediction, it would be difficult for the "conditioning" and the "concept-formation only" interpretations to account for.

The last hypothesis concerned a direct manipulation of the difficulty of learning the interlist associations. If this were possible, then it should reduce contingency awareness and consequently the probability of becoming aware of the correct demand characteristics, which in turn should reduce the conditioning. The rationale for this manipulation was similar to that of the color-naming manipulation of Insko and Oakes, although operationally the variables are quite different. For the Staats' interpretation of their experiment, since subjects supposedly didn't know the associations anyway and since the strength of response depends on number of pairings of the syllables with meaningful words, increasing the difficulty of noting interlist associations should have no effect on conditioning as long as number of trials are held constant.

METHOD

Subjects

Subjects were 288 introductory psychology students at the University of Nebraska at Lincoln. They were run in 12 groups of 24 each. Actually 19 other subjects were also run, but they were randomly eliminated, before the data were analyzed, to equalize the *N*s in the 12 groups.

Experimental Design

For an overall description of the basic experimental design, the reader is referred to Staats and Staats (1957, 1958). The present independent variables were: psychological sophistication of subjects (naive versus sophisticated), difficulty of interlist association (2, 4, and 10 filler syllables), and direc-

tion of conditioning (reversal of the syllable-meaning association), making a $2 \times 3 \times 2$ design. In addition, there were the assessed variables of interlist association (second learning test) and postexperimental measures of contingency and demand awareness.

Subject sophistication was varied by running half the subjects (naive) during the first 2 weeks of the university semester. The other half of the subjects (sophisticated) were run 11 weeks later in the semester. To control for possible biasing effects of early and late volunteers, subjects were not allowed to volunteer. The class rolls of three introductory sections were randomly divided and half were asked to participate early and the other half later. A plausible cover story was given for this procedure so as not to arouse suspicion, that is, that some learning experiments require random samples of subjects. Subjects run at the first of the semester were carefully sworn to secrecy until the end of the semester.

Association difficulty was manipulated by varying the number of neutrally paired or filler syllables included with the two crucial nonsense syllables. The crucial syllables were yof and wuh. In the easy condition (AD_1) only 2 filler syllables (laj and giw) were included with the crucial syllables. It was expected that with a list of only 4 nonsense syllables it would be rather easy to discover the relationship between the two lists. For the moderate difficulty condition (AD_2), 4 filler syllables were included (laj, giw, xeh, and qug). For the difficult condition (AD_3), 10 filler syllables were included (laj, giw, xeh, qug, meq, sij, vaf, vec, yim, and xad). It was expected that this amount of extra diversionary material would make the interlist association rather obscure. The effectiveness of this manipulation was checked by the "second learning test" of interlist association. Number of conditioning trials (18 pairings of pleasant or unpleasant words with yof and wuh) was held constant for all three groups.

Procedure

Through the presentation of the semantic differential rating scales the procedure was essentially the same as that of Staats and Staats in their many conditioning experiments. Beyond that point the procedure was changed to accommodate the second learning test and the postexperimental questionnaire. The experiment was conducted in a large classroom and subjects were seated with at least one empty seat between them so as to discourage either talking or copying.

The same deceptive orienting instructions were read to all groups. These were similar, but not identical, to those of Staats and Staats (1957); that is, "This is an experiment to see how well we can learn two separate lists simultaneously through two different sensory modalities." The nonsense syllables were presented by means of a slide projector. Each syllable appeared for 4 seconds with a 1-second change time. The lists of syllables (4, 6, or 12 in length for the three difficulty groups) were repeated 18 times in unsystematic order. One second after a syllable appeared the experimenter gave a spoken word and subjects repeated it in unison. For the crucial syllables (yof or wuh) the spoken words were consistently pleasant for one and unpleasant for the other. The order was reversed for half of the groups. The remainder of the spoken words were of neutral connotation.

Following learning, subjects were given the same deceptive orientation used by Staats and Staats, that is, "How we learn lists of words may be affected by how we feel toward the various words." They then read instructions for marking semantic differential scales. The first learning test consisted of 24 nonsense syllables each on a separate page of a small booklet. Beside each syllable was a blank line and below it was a pleasant-unpleasant semantic differential scale. Subjects rated all 24 syllables and made a check mark on the blank line if the syllable was one they had learned.

To this point the procedure was essentially a replication of the Staats and Staats procedure with the exception of the manipulation of association difficulty and the extra syllables added to the rating booklet so as to make it less obvious. The second learning test, consisting of a sheet of paper containing six syllables with large blank spaces below them, was then introduced. Instructions were as follows:

> Now, I also want to find out how many of the spoken words you can remember. We are passing out a recall test of the spoken words. We want to know how many associations between the spoken words and the written syllables you learned. So, for each nonsense syllable listed, write below it as many spoken words as you remember going with it. You may guess. This is a timed test and you will have only 5 minutes.

This measure represents an indirect index of awareness of the interlist contingency with considerable face validity.[3] It also has the advantage of being presented in the form of a test within the context of the experiment itself, rather than being postexperimental. The rationale is this: If subjects learned or formed the concept "good and bad words go with certain syllables," then they should be able to demonstrate it by writing down many more good and bad associates to the correct syllables than subjects depending on rote recall of something they had not been instructed to learn.

Finally, the postexperimental written questionnaire was introduced with orienting instructions, like those used in a previous study (Page, 1968), to enlist the honesty and cooperation of subjects while releasing them from the demand characteristics accompanying the experiment proper. Since there are many forms of postexperimental questionnaires currently in use, and since the resolution of the contradiction between the data of Insko and Oakes (1966) and those of the present study may involve differences in questioning and scoring, the entire postexperimental interview is reported in Table 1. Each

[3] The author is indebted to Katherine E. Baker for suggesting this measure.

TABLE 1

POSTEXPERIMENTAL QUESTIONNAIRE

1. What was the purpose of this experiment and what were you supposed to do?
2. During the experiment did you ever have the idea that its purpose might be something other than what I was telling you? What?
3. Thinking back to the experiment, did you notice at the time any relationship between certain syllables on the screen and the words that were spoken? What?
4. If you noticed any relationship between the lists, is this something you were actually aware of during the experiment or is it something you thought of while filling out these questions?
5. Do you remember approximately when it was that you noticed this? (1) right away, (2) first $\frac{1}{3}$ of learning, (3) second $\frac{1}{3}$, (4) last $\frac{1}{3}$, (5) while taking the first learning test, (6) while taking the second learning test.
6. What did you think was the purpose of the rating scales at the time you were filling them out, if anything?
7. How did you go about deciding what rating to give the various nonsense syllables?
8. Did you think that the experimenter might have expected that you would rate certain of the nonsense syllables in any certain way? Explain.
9. Was your answer to Question 8 something you were actually aware of before or during the marking of the rating scales, or something that you thought of afterwards?
10. What syllable was always or usually paired with travel words?
    a. How certain are you of this or are you guessing?
       Guessing——:——:——:——:——:——:——Certain
    b. Is this something you were aware of during the experiment or something you thought of since? Please explain if necessary.
11. What syllable was always or usually paired with words of pleasant meaning?
    a. How certain are you of this or are you guessing?
       Guessing——:——:——:——:——:——:——Certain
    b. Is this something you were aware of during the experiment or something you thought of since? Please explain if necessary.
12. What syllable was always or usually paired with words of unpleasant meaning?
    a. How certain are you of this or are you guessing?
       Guessing——:——:——:——:——:——:——Certain
    b. Is this something you were aware of during the experiment or something you thought of since? Please explain if necessary.
13. Were you ever aware during the experiment that yof [wuh for the other group] was always paired with words of pleasant meaning or connotation and that wuh [yof] was always paired with words of unpleasant meaning? And, if so, were you aware of any effect this might have had on you as you marked the rating scales? Explain.
14. Assuming that you knew the pleasant and unpleasant words and what was expected on the marking of the rating scales, rate your attitude while marking the rating scales.
       Resist the influence——:——:——:——:——:——:——Mark the right answers
15. Please make any other comments that you feel might help us understand your reaction to this experiment.
16. Have you had any previous courses in psychology such as in high school?
17. Do you know the meaning of the term conditioning? If so, did you think about it during this experiment?

question appeared on a separate sheet of a small booklet.

The assumption behind this questionnaire is that awareness, like problem solving, is basically a dichotomy. A subject either knows it or he doesn't. Also, subjects aren't likely to tell about such things as demand awareness unless specifically and carefully asked. The use of brief or vague questions may lead to either too many false positives or false negatives or both. These incorrectly scored subjects of either type would reduce any association that might actually be present. Questionnaires should have multiple indicators of awareness, and judges should score in terms of the total context of the questionnaire. Notice Questions 11 and 12, which are crucial questions for contingency awareness. Not only does a subject have to write down the correct contingency, but he has to be reasonably certain without guessing and willing to say he knew it earlier and isn't reflecting back now that the experiment is over. With this

type of multiple-criteria approach, one should be able to separate most of the aware from the unaware subjects.

Scoring Procedures

It was not possible to use the original Latin-square analysis (Staats & Staats, 1957) based on each subject having two scores. Because this study added two extra factors, it was more convenient to assign each subject a single conditioning score (yof rating minus wuh rating) which is equivalent to the previous scoring. This score can vary from −6 to +6 and should average approximately 0 for neutral nonsense syllables if there is no conditioning.

The second learning test of interlist associates was scored so as to maximize the measurement of learning of the correct concept (good words with yof and bad with wuh, or the reverse). Subjects who wrote down several guesses that were neither pleasant nor unpleasant had not learned the correct concept. Therefore, the most sensitive measure of interlist association would be total correct associates for the two crucial syllables minus total incorrect. Because this task proved to be impossible for subjects who hadn't formed the concept, this scoring resulted in a minus number (which depended on amount of guessing) for those who hadn't, and a large positive number for those who had. Actually, a cutoff score of 4 (2 for each syllable) was selected as a minimum criterion for considering a subject to have formed the concept.

The postexperimental questionnaires were read and scored by two independent judges. The judges scored for contingency awareness (did the subject say he was aware, during learning, of the interlist association) on a 4-point scale of (1) clearly aware, (2) probably aware, (3) probably unaware, and (4) clearly unaware. Most subjects fell in either Categories 1 or 4. Only questions from 1 to 12 were considered. The judges made an overall evaluation of the consistency and clarity of the subjects' responses to all questions, but keyed on Questions 11 and 12, one of which had to be answered correctly and with certainty for the subjects to be considered contingency aware. The judges' ratings correlated $r = .96$ and they disagreed as to the dichotomy aware-unaware on only 5 of 288 subjects.

The judges then scored for demand awareness (did the subject say he was aware, before marking the rating scales, of how the experimenter expected him to rate the syllables) on the same 4-point scale. Any subject who was not contingency aware could not logically be considered to have been aware of correct demand characteristics. This may be an important difference between the present measure and the Insko and Oakes measure of demand awareness, because in the present data several subjects knew that the experimenter expected something, but they didn't know what. Also, some contingency aware subjects reported that they realized what was expected too late for it to affect their ratings; these were not considered demand aware. The judges correlated $r = .91$ on their ratings of this variable. While this is a high reliability, the difference in reliability between contingency and demand awareness suggests that the latter was more difficult to judge from the protocols. There were 20 disagreements for aware-unaware for this variable. For both demand and contingency awareness the average of both judges' ratings was used as the index of awareness in the analysis. Cooperation-resistance was scored keying on Questions 13 and 14. Knowing about conditioning and thinking about it during the experiment were scored separately for Question 17.

TABLE 2

MEAN CONDITIONING (DISREGARDING DIRECTION OF CONDITIONING) FOR THE SIX GROUPS OF THE EXPERIMENT

S_1	Association difficulty			M
	AD_1	AD_2	AD_3	
Naive	1.43	1.15	.23	.94
Sophisticated	2.11	2.21	1.08	1.80
M	1.77	1.68	.66	1.37

TABLE 3

SUMMARY OF THE ANALYSIS OF VARIANCE PERFORMED ON CONDITIONING SCORES FOR TWO LEVELS OF SOPHISTICATION, THREE LEVELS OF ASSOCIATION DIFFICULTY, AND TWO DIRECTIONS OF CONDITIONING

Source	*df*	*MS*	*F*
Sophistication (A)	1	24.50	2.74
Difficulty (B)	2	25.45	2.84
Conditioning (C)	1	539.01	60.22**
A × B	2	4.03	—
A × C	1	53.39	5.97*
B × C	2	36.70	4.10*
A × B × C	2	1.94	—
Error	276	8.95	

* $p < .05$.
** $p < .001$.

RESULTS

Results for the independent variables of subject sophistication and association difficulty are presented in Tables 2 and 3. The analysis of variance was performed on the signed conditioning scores described earlier. In Table 2 the data were collapsed over the two syllable-sign conditions so as to give a clearer picture of the differences between the two more important variables. In Table 2, conditioning is indicated by a positive deviation from zero, otherwise the scores used in Tables 2 and 3 are equivalent. The means in each of the cells of Table 2 are based on *N*s of 48.

In Table 3, notice first the strong conditioning effect ($F = 60.22$, $p < .001$) showing that a replication of the Staats' results was obtained. Because of the nature of the experimental design (half of the subjects conditioned in one direction and the other half in the other), the hypotheses regarding subject sophistication and association difficulty are tested by the interactions of these variables with the conditioning effect. In Table 3 the interaction between sophistication and conditioning is significant ($F = 5.97$, $p < .05$). Notice (see Table 2) that the sophisticated subjects showed more conditioning ($\bar{X} = 1.80$) than the naive subjects ($\bar{X} = .94$) as predicted. The interaction between association difficulty and conditioning is also significant ($F = 4.10$, $p < .05$). From Table 2 we see a marked attenuation of conditioning in the difficult condition ($\bar{X} = .66$) relative to the other two conditions ($\bar{X}$s = 1.77 and 1.68) as predicted. Scores on the second learning test paralleled the pattern of results in Table 2, indicating that an attenuation of interlist association did occur in Group AD_3.

Recall the predictions concerning the relationships between the conditioned rating responses, the test for interlist association, and the measures of awareness. Since these variables were all either dichotomies or sharply bimodal, they were made into dichotomies before testing the association between them. Chi-squares were computed and converted into phi coefficients (see Table 4) to obtain a measure of strength of association. All of the relationships were very strong and highly significant as predicted. Under each phi coefficient is given the appropriate phi_{max}. This provides a basis for comparing strengths of association as phi_{max} is the maximum value of phi obtainable with the given marginal proportions (Guilford, 1956).

The tight association between all of the variables in Table 4 is striking. This means that if a subject had a high conditioning score (+4 or greater difference between positive and negative syllable), there is a high probability that he also did well on the test of association. He also was able to state postexperimentally the interlist contingencies and claimed to have been rather certain of them during the learning and before the rating. He also claimed that he knew, before marking the rating scales, how the experimenter expected him to mark them. This is very strong support for the explanation of the Staats' results presented earlier.

While all the variables in Table 4 are tightly bound together, the small differences in strength of association are meaningful. The best predictor of conditioning is demand awareness, not contingency awareness as claimed by Insko and Oakes. In fact, 18 subjects were contingency aware who weren't demand aware, and these subjects did not show high conditioning. The mean for this group, when conditioning was scored as a positive deviation from 0, was .11, which was not significantly different from 0 ($t = .03$, *ns*). The mean of the unaware group was also almost exactly 0 (−.02). The mean for the subjects who were both contingency and demand aware was 4.6, which was quite significantly different from 0 ($t = 37.33$, $p < .0001$). While demand awareness is closely associated with contingency awareness (the first criterion for demand awareness was contingency awareness), the slight discrepancy between these measures makes demand awareness a stronger correlate of the conditioned ratings.

Figure 1 presents the distributions for conditioning scores of demand aware and cooperating versus unaware subjects. The direction of conditioning is folded over in this figure so that positively signed scores mean behavior in the direction of conditioning. Notice the symmetrical distribution of scores for unaware subjects; only two had conditioning scores of 6 while more than 50% of the demand awares had scores of 6. In Figure 2, notice that a distribution of "conditioning"

TABLE 4

SUMMARY OF PHI COEFFICIENTS (OVER THEIR APPROPRIATE PHI_{MAX} VALUES) OF ASSOCIATIONS BETWEEN FOUR DICHOTOMIZED VARIABLES

Variable	2	3	4
1. Conditioned ratings	.67/.90	.81/.93	.69/1.00
2. Contingency awareness		.86/.88	.79/.88
3. Demand awareness			.76/1.00
4. Association learning			

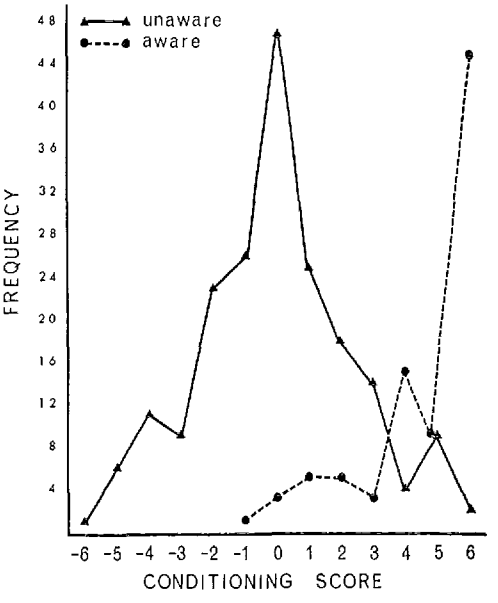

FIG. 1. Frequencies of conditioning scores on experimental syllables yof and wuh (positive minus negative syllable) for the unaware ($n = 195$) versus the demand aware and cooperating subjects ($n = 86$).

scores artificially generated on control syllables looks just like the distribution for unawares on experimental syllables. Only demand awares rated the experimental syllables any differently than control syllables.

What of the demand aware subjects with positive but not extreme scores? A few at 4 were aware of both contingencies, but did not use the ends of the scales. Most demand aware subjects with conditioning scores of 5 or less were only aware of one contingency. If a subject was aware that extreme ratings were expected for only one syllable and rated it accordingly while rating the other syllable according to his actual feelings about a nonsense syllable, then his conditioning score should be positive and would vary according to how he rated the syllable on which he was unaware. This is exactly what happened. All aware subjects with conditioning scores of 6 were aware of both contingencies; most with lower scores were aware of only one. These findings are very consistent with the general formulation which generated the experiment.

Cooperation-resistance has not been discussed in this paper because the number of resistors ($n = 7$) was much smaller than found in a previous study (Page, 1968). However, these few who reported resisting did not show conditioning, corresponding to the previous results.

Recall that subjects were asked whether they knew the meaning of conditioning and if they had thought about it during the experiment. The argument of this paper does not depend upon aware subjects knowing or thinking about conditioning; consequently no predictions were made, but it would be interesting support for the argument if some of them did. The association between sophistication and knowing the meaning of conditioning was highly significant ($\chi^2 = 70.6$, $df = 1$, $p < .001$). Only 10 of 144 sophisticated subjects did not know the meaning of conditioning while this was true for 75 of the naives. Most introductory students apparently do learn something during a semester. What is more interesting is that almost half of the naive group also knew about conditioning, and therefore "naive" wasn't an entirely correct description of this group. The association between knowing the meaning of conditioning and demand awareness was also significant

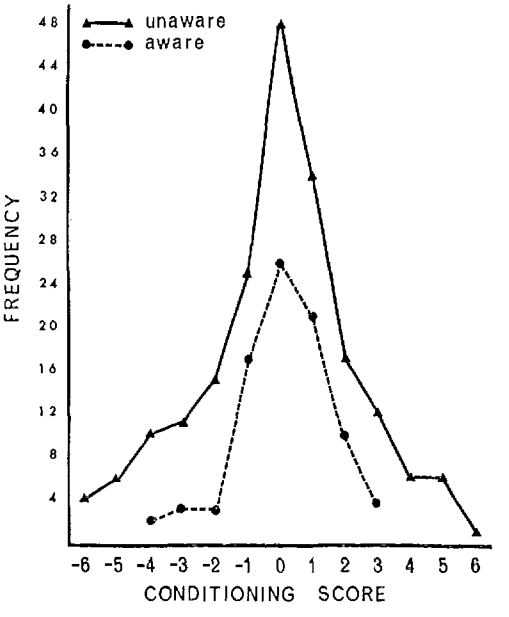

FIG. 2. Frequencies of conditioning scores on control syllables laj and giw (first minus second syllable with order reversed for half the subjects) for the unaware versus the demand aware and cooperating subjects.

($\chi^2 = 9.65$, $df = 1$, $p < .01$), though it doesn't account for a large amount of total variance. Of those ($n = 203$) who knew the meaning of conditioning there was a highly significant association ($\chi^2 = 36.55$, $df = 1$, $p < .001$) between thinking about conditioning during the experiment and demand awareness. Perhaps these data should not be overinterpreted, but they seem to suggest that while knowing about conditioning (and probably other psychology-related concepts) and thinking about it are not required for becoming demand aware in this situation, they do facilitate it.

DISCUSSION

Every hypothesis of this study was clearly supported and all the converging evidence seems to attest to the essential correctness of the formulation which generated the predictions. It appears that the Staats' (1958) classical conditioning of evaluative affect or attitudes interpretation of their data is incorrect. The reaction they evoked in their subjects was far too complex, and apparently cognitively mediated, to be labeled classical conditioning. This critique says nothing about the general theory which they believe was demonstrated by their experiment (Staats, 1967). Attitudes certainly are learned and perhaps sometimes through a process similar to classical conditioning. But this neat little deception experiment, using college sophomores as subjects, does not seem to demonstrate it.

The present study supports the idea that the so-called conditioned attitudes are entirely artifacts of demand characteristics. This is not to say that the social perception, social influence, and other social psychological variables present in the total laboratory situation were not genuine and interesting psychological phenomena, but they are artifactual sources of variance from the point of view of what the original experiments were designed to study. Perhaps the demand-characteristics problem is such a limiting factor on deception experiments that the kinds of hypotheses the Staats' were interested in testing simply cannot be tested in the psychological laboratory at the present stage of development.

From the present vantage point it seems that while attitudes are learned phenomena, it is rather naive to suppose that a brief, deceptive, and highly artificial laboratory experiment could induce genuine attitudes towards neutral stimuli. Perhaps psychologists would do better in the future to recognize the limitations of "quick and easy" deception experiments such as here described. The popularity of such methodology in our discipline seems symptomatic of a kind of worship of experimental manipulations for their own sake (Bakan, 1967) without due regard for the complexities or the actualities of the social context in which these manipulations occur (Sherif, Sherif, & Nebergall, 1965).

The conflict between the present results and those of Insko and Oakes (1966) as to the relative importance of contingency and demand awareness remains to be clarified by further empirical investigation. While they found that demand awareness could not account for all of the variance (Insko, 1967, p. 29), the present study suggests that it can. There are many procedural differences between these two studies; most specifically the different questionnaires and scoring techniques used in assessing awareness. In addition to differences in specific operations there were basic differences in assumptions about the nature of so-called "awareness." In the present study an elaborate and multifaceted written questionnaire was used to answer a simple question: Did the subjects know the approximate purpose of the experiment while it was going on? Insko and Oakes used a simpler oral questionnaire to answer a more difficult question. Under the assumption that demand awareness is a continuum, they attempted to quantify that continuum. They assumed that individual differences in clarity of expression in responding to questions was an accurate reflection of the so-called "continuum of demand awareness," and then classified degrees of clarity on an ordinal scale. If in fact demand awareness is more accurately thought of as a dichotomy, and Insko and Oakes were actually quantifying something only roughly correlated with awareness, then this could account for the lower correlation between their demand-awareness scale and the rating scales. However, which of these sets of assumptions about the nature of demand awareness is

more adequate remains a question for further investigation.

REFERENCES

BAKAN, D. *On method: Toward a reconstruction of psychological investigation.* San Francisco: Jossey-Bass, 1967.

COHEN, B. H. Role of awareness in meaning established by classical conditioning. *Journal of Experimental Psychology,* 1964, **67**, 373–378.

GUILFORD, J. P. *Fundamental statistics in psychology and education.* (3rd ed.) New York: McGraw-Hill, 1956.

HOLMES, D. S. Amount of experience in experiments as a determinant of performance in later experiments. *Journal of Personality and Social Psychology,* 1967, **7**, 403–407.

INSKO, C. A. *Theories of attitude change.* New York: Appleton-Century-Crofts, 1967.

INSKO, C. A., & OAKES, W. F. Awareness and the "conditioning" of attitudes. *Journal of Personality and Social Psychology,* 1966, **4**, 487–496.

KELMAN, H. C. Human use of human subjects: The problem of deception in social psychological experiments. *Psychological Bulletin,* 1967, **67**, 1–11.

ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17**, 776–783.

PAGE, M. M. Role of awareness in the classical conditioning of meaning. Unpublished master's thesis, University of Oklahoma, 1964.

PAGE, M. M. Modification of figure-ground perception as a function of awareness of demand characteristics. *Journal of Personality and Social Psychology,* 1968, **9**, 59–66.

SHERIF, C. W., SHERIF, M., & NEBERGALL, R. E. *Attitude and attitude change.* Philadelphia: Saunders, 1965.

SILVERMAN, I. Role-related behavior of subjects in laboratory studies of attitude change. *Journal of Personality and Social Psychology,* 1968, **8**, 343–348.

SPIELBERGER, C. D. The role of awareness in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham, N. C.: Duke University Press, 1962.

SPIELBERGER, C. D., & DENIKE, L. D. Descriptive behaviorism versus cognitive theory in verbal operant conditioning. *Psychological Review,* 1966, **73**, 306–326.

STAATS, A. W. An outline of an integrated learning theory of attitude formation and function. In M. Fishbein (Ed.), *Readings in attitude theory and measurement.* New York: Wiley, 1967.

STAATS, C. K., & STAATS, A. W. Meaning established by classical conditioning. *Journal of Experimental Psychology,* 1957, **54**, 74–80.

STAATS, A. W., & STAATS, C. K. Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology,* 1958, **57**, 37–40.

(Received May 22, 1968)

ERRATUM

In the article by John T. Lanzetta and James M. Driscoll in the December issue, the institutional connection for John T. Lanzetta on page 479 should read: *Dartmouth College.*