

DATA DESIRABILITY, EXPERIMENTER EXPECTANCY, AND THE RESULTS OF PSYCHOLOGICAL RESEARCH¹

ROBERT ROSENTHAL, PAUL KOHN, PATRICIA M. GREENFIELD,

AND NOEL CAROTA

Harvard University

19 male Es employing a Taffel-type task conducted a verbal conditioning experiment with 60 female Ss. $\frac{1}{2}$ the Es were led to expect their Ss to show verbal conditioning, and $\frac{1}{2}$ were led to expect no verbal conditioning. $\frac{1}{2}$ the Es in each of these groups were led to feel that it would be desirable if their Ss showed conditioning, and $\frac{1}{2}$ were led to feel that it would be undesirable. Those Es who (a) both wanted and expected, and (b) neither wanted nor expected their Ss to show increased use of I and WE pronouns obtained significant conditioning ($p = .001$). Those Es who (a) wanted but did not expect, and (b) expected but did not want increased use of I and WE pronouns obtained no significant conditioning ($p = 1.00$). Ss high in need for social approval arrived earlier at the site of the experiment, were less "aware" of the contingency but were no more likely to show conditioning. Ss' ratings of Es' behavior during the experiment showed significant differences between Es in different experimental conditions, between Es who were 1st vs. later born, and between Es who were high vs. low in need for social approval.

A series of experiments has recently been reported which suggests that for a variety of experimenters, subjects, and situations, the experimenter's expectancy or hypothesis may be a significant partial determinant of the results he obtains (Rosenthal, 1964). In most of our studies, the particular data experimenters were led to expect were presumably also desired by them. The effects of experimenter expectancy have therefore been confounded with those of desirability. This seems reasonable from the standpoint of ecological validity. "Real" experimenters ordinarily want to obtain data they expect and do not want to obtain data they do not expect. Nevertheless one can think of research situations wherein experimenters expect to obtain data they do not consider desirable for any of several reasons (e.g., Milgram, 1963). Moreover, one can also think of situations wherein experimenters do not really expect to obtain data which would be highly desirable (i.e., "long-shot" studies).

The general purpose of the present investigation, then, was to study the separate and combined effects on research findings of the

experimenter's expectancy of certain results and the desirability to the experimenter of those results. A more specific purpose of the present study was to learn whether experimenters' expectancies and desires might be partial determinants of the results of studies of verbal conditioning. An earlier study did show that the experimenter's expectancies might be determinants of the degree of his subjects' awareness that they had undergone a conditioning procedure. That study did *not*, however, vary the experimenter's expectancy of whether conditioning would or would not occur; nor did it involve any manipulation of the outcome desirability of either subjects' awareness or subjects' conditioning scores (Rosenthal, Persinger, Vikan-Kline, & Fode, 1963).

Accordingly, half our experimenters were led to expect that their subjects would show verbal conditioning while the remaining experimenters were led to expect no verbal conditioning. Half the experimenters in each of these groups were led to feel that their subjects' verbal conditioning would reflect well on the experimenter while the remaining experimenters were led to feel that their subjects' verbal conditioning would reflect badly on the experimenter.

¹This investigation was supported by Research Grants G-24826 and GS-177 from the Division of Social Sciences of the National Science Foundation.

METHOD

Experimenters

Nineteen male graduate students served as paid volunteer experimenters. All were in their first 2 years of work at Harvard's Division of Engineering and Applied Physics.

Subjects

The experimenters ran a total of 60 paid volunteer subjects, all of them female students at a Boston secretarial school. The experimental task was explained to the subjects as a test of verbal facility.

Experimental Task

The experimental task was a modified Taffel (1955) procedure. The experimenters presented their subjects with four sets of 20 verbs, each of which was to be used in a sentence constructed by the subject. Each sentence was to begin with any of the following pronouns: I, WE, YOU, HE, SHE, and THEY. In order to reduce any errors of observation or recording, subjects wrote down each sentence and then read it aloud to the experimenter who checked to be sure that the subject had read the same pronoun she had recorded. On the last 60 trials (three blocks of 20), the experimenters said "good" whenever I or WE was the pronoun selected. The dependent variable was a measure of the increase in the use of I or WE from the operant level to the subsequent blocks of trials.

Procedure

The experiment was conducted in one day at two different hours. All experimental conditions were represented in each session. The experimenters in each session were trained as a group. They were given factual material about the phenomenon of verbal conditioning. Two reasons were advanced for their participation in this experiment. The first stressed the need for replication by researchers who were not behavioral scientists in order to extend the generality of the findings in the verbal conditioning literature. The second stressed our interest in learning more of the relationship between the experimenters' personality and their subjects' conditioning scores. All experimenters were then administered the Marlowe-Crowne Social Desirability scale, the Taylor Manifest Anxiety scale, a form for establishing birth order, and the first set of 20 verbs they would present to their subjects. This was done to obtain the experimenters' operant level for using I and WE and was obtained of course before the experimenters knew which pronouns they would be reinforcing. The experimenters were told that their subjects would be assigned them on the basis of the subjects' similarity to the experimenters in personality as measured by the tests the experimenters had taken. The same tests were in fact administered to the subjects, but assignment of subjects to experimenters was essentially random.

After each subject finished her experimental task and left her experimenter's research room she filled out two questionnaires designed to define the degree of her awareness that she had undergone a verbal conditioning procedure. The first questionnaire (Q₁) simply asked the subject to state the purpose of the experiment (Matarazzo, Saslow, & Pareis, 1960). The second questionnaire (Q₂) repeated the substance of the first but asked more specific and more leading questions (Levin, 1961). Both questionnaires had been modified for use in an earlier study (Rosenthal et al., 1963). Each subject also filled out a series of 28 rating scales designed to assess her perception of her experimenter. Each scale had 20 points running from +10 (e.g., extremely businesslike) to -10 (e.g., extremely unbusinesslike) with intermediate labeled points. This same questionnaire had been employed in earlier experiments (Rosenthal, Fode, Friedman, & Vikan-Kline, 1960).

Experimental Conditions

The experimental treatments were administered in the form of "last-minute instructions" placed on each experimenter's desk. For half the experimenters the instructions claimed that their subjects had personality characteristics such that they would condition well. The remaining experimenters were led to expect their subjects to condition poorly. The desirability of these two outcomes was implemented by telling half the experimenters that conditionability was highly correlated with general learning ability and by telling the others that it was highly correlated with susceptibility to deliberate manipulation. Since the experimenters believed themselves to be similar to their subjects in personality, the first group of experimenters should find good conditioning data a desirable outcome since it would imply that the experimenter, like his subjects, had good general learning ability. The remaining experimenters should find good conditioning data an undesirable outcome since it would imply that the experimenter, like his subjects, was highly manipulatable.

Those portions of the instructions used to implement data desirability were as follows:

[Desirable]: There are a few things we now know about subjects who are more and less conditionable. Our hope, of course, is to learn a good deal more about that. What we know so far suggests that highly conditionable people tend to have high general learning ability. They pick up new concepts and ideas quickly and have skill in analyzing and solving problems. Poor conditioners, in contrast, tend to have lesser abilities in these areas.

[Undesirable]: There are a few things we now know about subjects who are more and less conditionable. Our hope, of course, is to learn a great deal more about that. What we know so far suggests that highly conditionable people tend to be manipulatable. They are often like putty in the hands of advertisers and salesmen. Poor conditioners, in contrast, tend to be very resistant to

such manipulation; in other words, they seem to have minds of their own.

Within each of the two above conditions of data desirability, half the experimenters received one of the following two additional instructions specifying expectancy:

[Expect]: The particular subjects assigned to you, on the whole, tend to be good conditioners. That is, they will tend to show a significant increase in the number of "I" and "WE" pronouns from the first set of 20 sentences to the later sets.

[Don't expect]: The particular subjects assigned to you will, on the whole, tend to be poor conditioners. That is, they will not tend to show a significant increase in the number of "I" and "WE" pronouns from the first set of 20 sentences to the later sets.

The method described above of implementing our outcome-desirability variable was selected on the basis of an instruction pretest with Harvard undergraduates from a course in motivation. These subjects received eight characterizations of the personality correlates of conditionability. Of these, four were designed to be desirable and four undesirable. The subjects were instructed to imagine themselves to be experimenters running an experiment on the personality correlates of verbal conditioning. Further, they were asked to imagine that their "subjects" were assigned to them on the basis of similarity (to themselves) on important personality dimensions "so that any interpretation of experimental results should apply to [them] . . . as well as to [the] subjects." Under these role-playing conditions subjects were asked to rate the eight characterizations on five scales: the desirability of having highly

conditionable subjects, assuming the characterizations to be correct; the desirability of having subjects who were highly resistant to conditioning, assuming the characterizations to be correct; the probability that a competent psychologist would be right if he predicted their subjects to be highly conditionable; the probability that a competent psychologist would be right if he predicted their subjects to be highly resistant to conditioning; the believability of the characterization if made by a competent psychologist.

The two characterizations used in the present experiment were selected because the patterns of ratings assigned them on our five scales were generally superior to those of the other characterizations. That is, our pretest subjects rated these two characterizations as relatively believable, and relatively desirable, in one case and undesirable in the other. Moreover, they expressed considerable readiness to concur with either of the two opposite predictions by a competent psychologist.

Precautions against Authors' Expectancy Effects

We took a number of precautions to prevent our own outcome expectancies and outcome desires from having major effects on the data collected by the experimenters. These included: randomly assigning rooms to conditions; assigning experimenters to rooms in order of departure from the experimenter reception room; implementing the independent variables in a way involving no contact with the experimenters by persons who were aware of the experimenters' treatment conditions; assigning subjects to experimenters on the basis of (a) order of subjects' departure from the subject reception room, (b) the experimenter's immediate availability to run a new subject, and (c) the number of subjects run by each experimenter up to that point.

The above procedures left all the major investigators except one blind to each experimenter's treatment condition. The lone exception was the author who was in charge of assigning rooms to conditions and ensuring that approximately equal numbers of experimenters and subjects were assigned to each treatment condition. He had no contact with either experimenters or subjects during the course of the experiment. With the foregoing precautions, the likelihood that the major investigators' own expectancies and desires substantially affected the data seems small.

RESULTS

Conditioning

Initially four alternative definitions of conditioning were employed: increase in I-WE usage from the operant level to Block 4; increase in I-WE usage from the operant level to the mean of the subsequent three blocks; increase in I-WE usage from the operant level to Block 4 plus one-third the increase from Block 2 to Block 3; monotonicity of increase

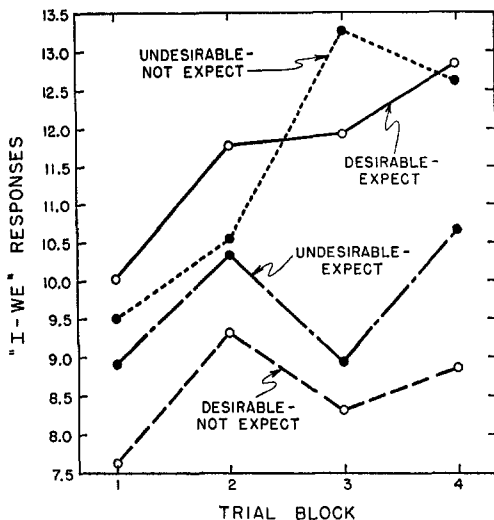


FIG. 1. Frequency of choice of I and WE as a function of experimenter expectancy, outcome desirability, and trial block.

of I-WE usage from Blocks 1 to 4 as measured by rank correlation between number of I-WE responses and block number. The median intercorrelation of these dependent variables was .79. Because the first of these definitions was both the simplest and most highly correlated with the other definitions, it was accepted as the definition of magnitude of conditioning.

Figure 1 shows the mean of the mean number of I-WE responses obtained by experimenters in each of the four experimental groups for each of the four blocks. Table 1 shows the mean conditioning scores (Block 4 - Block 1) and Table 2 the results of the analysis of variance. Only the interaction was significant. Apparently, then, neither experimenter's expectancy nor the desirability of conditioning data alone affect the magnitude of conditioning scores reliably but the congruence between expectancy and data desirability does make a substantial difference. Under the congruent conditions 100% of the experimenters showed a mean increase in their subjects' use of I and WE ($p = .001$). Under the incongruent conditions only 56% of the experimenters showed a mean increase ($p = 1.00$).

Awareness

The awareness questionnaires were independently and blindly scored by two of the authors on a 3-point scale: clearly unaware, 1; vaguely aware, 2; and clearly aware of the response-reinforcement contingency, 3. The reliabilities of Q_1 and Q_2 were .95 and .97, respectively.

Of all subjects 17% were classed as clearly aware,² 8% as vaguely aware, and 75% as clearly unaware (Q_2).

Experimenters who expected conditioning tended to obtain a lower rate of clear awareness (7%) than did experimenters who did not expect conditioning (25%). Because the bulk of subjects in both experimenter expectancy conditions were "clearly unaware," the difference in rates of clear awareness approached significance ($p = .10$) only when

² Among subjects judged as clearly aware a few specifically mentioned their decision to go along or not go along with their experimenters' attempts to influence them.

TABLE 1
EXPERIMENTERS' MEAN CONDITIONING SCORES
FOR FOUR EXPERIMENTAL GROUPS

Outcome desirability	Expectancy	
	Expect	Don't expect
Desirable	3.0	1.3
Undesirable	0.8	3.1

the analysis was limited to subjects who were either vaguely or clearly aware.

Aside from her experimenter's treatment condition, two factors proved to be related to the subject's subsequent awareness: the subject's personality and the order in which she was run by the experimenter. Subjects who scored as more anxious ($r = -.22$, $p = .10$) and as higher in need for social approval ($r = -.30$, $p = .02$) were less likely to become aware subsequently. Subjects run later by a given experimenter were more likely to become aware ($r = .26$, $p = .05$). In addition, all the subjects run in the second experimental session were more likely to become aware than subjects run in the first session ($p = .08$).

Subjects' Perceptions of Experimenters

Subjects had rated their experimenters on 28 variables immediately after the experiment. Those experimenters who had been in the congruent experimental treatment groups were rated by their subjects as more casual ($r = .33$, $p = .01$), more courteous ($r = .27$, $p = .05$), more pleasant ($r = .24$, $p = .08$), more expressive-voiced ($r = .24$, $p = .08$), and as less given to the use of movements of the trunk region ($r = -.26$, $p = .05$). Because of the intercorrelations among these particular variables and among the total set

TABLE 2
ANALYSIS OF VARIANCE OF EXPERIMENTERS'
MEAN CONDITIONING SCORES

Source	df	MS	F
Expectancy (A)	1	0.18	
Desirability (B)	1	0.06	
A × B	1	22.26	10.55*
Within	15	2.11	

* $p = .006$.

TABLE 3
SUBJECTS' CONDITIONING SCORES AND THEIR
PERCEPTION OF EXPERIMENTERS

Variable	<i>r</i>	<i>p</i>
Interested	.43	.001
Businesslike	.43	.001
Professional	.33	.01
Quiet (nonloud)	.31	.02
Enthusiastic	.28	.04
Behaved consistently	.26	.05
Expressive-voiced	.24	.08

of 28 variables, no simple statement is possible of how many of these particular correlations might be attributed to chance. It seems likely, however, that the experimenters' behavior during the experiment, as defined by their subjects' ratings was at least in part determined by the experimental treatment conditions.

Table 3 shows the correlations between the magnitude of the subjects' conditioning scores and their perceptions of their experimenters. We cannot be sure, of course, that these ratings of experimenters actually do reflect differences in experimenter behavior. It is also possible that those subjects who are more susceptible to the interpersonal influence of a reinforcing experimenter simply describe experimenter behavior differently. Or, having been influenced by a reinforcing experimenter, these subjects may have rated that experimenter according to their preconceptions of the sort of person by whom they would permit themselves to be influenced. Assuming for the moment that these ratings accurately described experimenter behavior, subjects were more influenced by experimenters showing a general enthusiastic interest in them; conveying a consistent, professional, businesslike manner; and speaking in a quiet but expressive tone of voice. If these experimenters did *not* in fact behave in this way, at least it seems warranted to believe that more influential subjects ascribe such characteristics to the experimenters by whom they are influenced.

Experimenter Characteristics

Experimenters' birth order, operant level of I and WE, and need for social approval were

not related to their subjects' conditioning scores. Experimenters' anxiety scores were related to their subjects' conditioning in a nonlinear manner. Both high- and low-anxious experimenters obtained greater conditioning than did medium-anxious experimenters ($F = 3.08$, $df = 2/13$, $p = .08$).

While not related to subjects' conditioning, experimenters' birth order appeared to be a significant predictor of experimenters' behavior in the experiment as defined by subjects' ratings. Table 4 shows the correlations between experimenters' birth order and a number of behavioral variables. First-born experimenters were generally rated as fast but reluctant speakers who used fewer body and facial movements and expressions.

Experimenters who used more I and WE pronouns in their pretesting were rated by their subjects as more casual ($r = .34$, $p = .01$), more enthusiastic ($r = .22$, $p = .10$), and more pleasant-voiced ($r = .22$, $p = .10$).

Experimenters scoring higher in need for social approval were rated by their subjects as less personal ($r = -.32$, $p = .02$), less loud ($r = .27$, $p = .05$), less talkative ($r = .22$, $p = .10$), more enthusiastic ($r = .27$, $p = .05$), but less well-liked ($r = -.22$, $p = .10$). None of the subjects' ratings of experimenters' behavior during the experiment showed a correlation with experimenter anxiety which was significant at the .05 level.

Subject Characteristics

Subjects' birth order, anxiety, and need for social approval were found to be unrelated to subjects' conditioning scores. Subjects high in need for social approval were found to be later born more often than first born ($r = -.24$, $p = .08$) and, interestingly enough,

TABLE 4
RELATIONSHIPS BETWEEN EXPERIMENTERS' BEING
FIRST BORN AND RATINGS BY THEIR SUBJECTS

Variable	<i>r</i>	<i>p</i>
Less talkative	.37	.006
Fast speaking	.32	.02
Body use	-.32	.02
Trunk use	-.27	.05
Hand gestures	-.26	.05
Expressive face	-.24	.08

to have arrived earlier at the site of the experiment ($r = .40, p = .003$).

Subjects' operant levels of r and WE responses were not significantly related to their conditioning scores ($r = -.10$). Subjects' operant levels, however, tended to be a function of experimenters' treatment conditions. Those experimenters for whom data desirability and expectancy operated conjointly obtained higher rates of operant level responding from their subjects ($F = 3.22, df = 1/15, p = .10$). Thus while congruence of data desirability and experimenter expectancy was associated both with high operant levels and high conditioning scores, it is clear that the conditioning scores cannot be attributed to the operant levels.

Model Bias

The extent to which a given experimenter's own performance of a task determines his subjects' performance of the same task is the extent to which the experimenter "models" his subjects. A recent summary of experiments testing the hypothesis of modeling effects suggested that in different experiments there might be different orders of magnitude of experimenter modeling effects (Rosenthal, 1964).

In the present study, modeling effects were defined by the correlation between experimenters' own operant levels of r and WE and the mean operant levels of their subsequently run subjects.

Table 5 shows these correlations for each of the four experimental conditions. Two of the four correlations were significant at the .05 level, and the four r 's were significantly different from one another ($\chi^2 = 17.68, df = 3, p < .001$). It appears then, that whether the experimenter expects and/or desires a certain outcome may significantly affect the direction and magnitude of experimenter modeling effects.

Qualitative Analysis of Awareness Questionnaires

Most of the subjects (88%) felt the purpose of the experiment was to assess their personality. Since the subjects had filled out personality questionnaires, this ascription of purpose was natural enough. Many of the

TABLE 5
CORRELATION BETWEEN EXPERIMENTERS' AND SUBJECTS' OPERANT LEVELS FOR FOUR EXPERIMENTAL CONDITIONS

Outcome desirability	Expectancy	
	Expect	Don't expect
Desirable	-.88*	+.997**
Undesirable	+.03	-.74

* $p < .05$, two-tailed.
** $p < .005$, two-tailed.

subjects, however, saw the "test of verbal facility" as a personality test akin to word-association or sentence-completion techniques. More specifically, several subjects saw it as a test of their egocentricity, as measured by the frequency of their use of r .

Only 30% of all subjects believed their experimenters when they told them their verbal abilities were being assessed. Among subjects run by experimenters in the congruent conditions, only 18% believed their abilities were being assessed. In the incongruent conditions, 46% of the subjects believed their abilities were being assessed. The differences in belief rates were significant ($\chi^2 = 5.68, p = .02$) suggesting that the behavior of the experimenters in the congruent conditions made it seem more unlikely to their subjects that their verbal abilities were being assessed.

It has been suggested elsewhere (Rosenthal et al., 1963) that subjects may be interested in their experimenters as people rather than simply as "scientists." Evidence of a "transference reaction" was presented. In the present study, 20% of all subjects made some reference to one or more physical characteristics of their experimenter which were irrelevant to the experimenter's role performance. These included mention of the experimenter's posture, clothing, facial blemishes, wearing of glasses, condition of teeth, and relative attractiveness.

DISCUSSION

The results of the present experiment were both unequivocal and surprising, and their interpretation can at best be only tentative. This was the first experiment in our research program in which the experimenters' expect-

tancies were varied independently of the desirabilities of the outcomes. In most of the earlier research experimenters who expected a given outcome probably also desired it while those who did not expect that outcome also did not desire it. There have, however, been a few studies in which it could be argued that all experimenters desired a given outcome while differing in their expectancy of it. Such would be the case in experiments employing animal subjects in which all experimenters wanted their subjects to perform well since their course grades might depend on it (Rosenthal, 1964). Those experimenters who expected better performance from their subjects obtained better performance than did those who expected poorer performance. For these experiments it could be argued from the results of the present study that if there had been a group of experimenters who neither expected nor desired good performance from their subjects they would have obtained performance as good as that obtained by experimenters who both wanted and expected good performance. Only another experiment can answer this question for us.

But, from one point of view, the present study seems to contradict the bulk of the earlier research in which *opposite* expectancies (coupled with presumably congruent motives) were associated with correspondingly *opposite* results (Rosenthal, 1964). In the present study, on the other hand, opposite expectancies combined with congruent motives produced *identical* results. The present study differed from earlier ones in a number of ways any of which alone or in interaction with one another could account, even if not simply, for the differences. The present study was the first to: employ expectancies about verbal conditioning performances, employ as experimenters graduate students in the physical sciences, and create expectancies about verbal behavior in which experimenters were explicitly taught how such behavior could be intentionally manipulated thereby confounding the unintentional biasing process with the intentional reinforcement process.

Perhaps the simplest tentative explanation is based on a reexamination of the phenomenology of the experimenters in the various experimental conditions. Those experimenters

who both expected and wanted conditioning or neither expected nor wanted conditioning were told by us essentially that we thought they were particularly clever in the one case and that they had minds of their own in the other. Thus the congruent experimenters were complimented by the investigators. On the other hand, experimenters in the incongruent conditions were told essentially that we thought them to be either not too bright or like putty in the hands of manipulators. The experimenters in the incongruent condition then were anything but complimented by their employers. These experimenters could have been emotionally affected to the point where their verbal "reinforcements" lacked sufficient conviction to be positive reinforcers for their subjects. Experimenters in the noncongruent condition were in fact rated by their subjects as less expressive-voiced than experimenters in the congruent condition ($p = .08$) and expressiveness of voice was positively correlated with successful conditioning ($r = +.24$, $p = .08$).

If the interpretation offered to account for our surprising results is correct, then the present experiment in no way contradicts earlier findings, although the relation between the two sets of results requires further clarification. In the bulk of the previous work, affect was not experimentally manipulated, while in this study we must conclude that the experimenter's affect or mood is a more important determinant of his effectiveness as a reinforcer than either his expectancy or the desirability of the outcome in studies of verbal conditioning.

Among those subjects showing some indication of awareness, more clear awareness was shown by subjects whose experimenters had had been led to expect no conditioning. It seems possible, therefore, that some of the ambiguity surrounding the question of awareness rates in studies of verbal conditioning may be associated with the experimenter's expectancy regarding subjects' conditionability as well as his expectancy about subsequent awareness (Rosenthal et al., 1963).

Studies by Crowne and Strickland (1961) and by Marlowe (1962) found that subjects with a greater need for social approval showed greater verbal conditioning effects. The pres-

ent study, like that by Spielberger, Berger, and Howard (1963) found no such relationship. However, those of our subjects with a greater need for approval showed significantly less awareness of the "response-reinforcement" contingency. Quite possibly these subjects recognized that the socially desirable thing to do when a psychological investigator inquires after awareness in a conditioning experiment is to "not see through" the experimental situation. This interpretation is quite consistent with the position that the need for approval is a tendency to respond appropriately to perceived situational demands (Crowne & Marlowe, 1964). Further construct validation of the Marlowe-Crowne scale comes from our finding that subjects higher in need for approval were more likely to arrive earlier at the site of the experiment than subjects lower in need for approval. However, why subjects higher in need for approval should more often be later born than first born is by no means obvious.

The finding that subjects run later by a given experimenter were more likely to be aware is most parsimoniously interpreted as due to later-run subjects having a lower need for social approval. The finding that subjects run in the second experimental session were more likely to be aware is less clearly explained. One likely interpretation involves the possibility of feedback from Session I subjects to Session II subjects. This is not a trivial problem. We may wonder about the effects of feedback from earlier- to later-run subjects in a good deal of behavioral research. Needed are some hard data on the efficacy of the optimistically solicited loyalty oath wherein we swear our earlier-run subjects to secrecy "until the experiment is over."

Asking subjects to describe their experimenter's behavior during the experiment seems to be a useful technique. On the basis of these descriptions we were able to differentiate experimenters under the various experimental conditions. These descriptions suggest

how the preexperimental manipulations of experimenter variables (as well as the experimenter's more enduring personal characteristics) might be translated into unprogrammed experimenter behavior during the experiment. Our data suggest that it is this unprogrammed behavior which is responsible for the experimenter's unintentional effect upon the results of his experiment.

REFERENCES

- CROWNE, D. P., & MARLOWE, D. *The approval motive*. New York: Wiley, 1964.
- CROWNE, D. P., & STRICKLAND, BONNIE R. The conditioning of verbal behavior as a function of the need for social approval. *Journal of Abnormal and Social Psychology*, 1961, **63**, 395-401.
- LEVIN, S. M. The effects of awareness of verbal conditioning. *Journal of Experimental Psychology*, 1961, **61**, 67-75.
- MARLOWE, D. Need for social approval and the operant conditioning of meaningful verbal behavior. *Journal of Consulting Psychology*, 1962, **26**, 79-83.
- MATARAZZO, J. D., SASLOW, G., & PARETS, E. N. Verbal conditioning of two response classes: Some methodological considerations. *Journal of Abnormal and Social Psychology*, 1960, **61**, 190-206.
- MILGRAM, S. Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 1963, **67**, 371-378.
- ROSENTHAL, R. The effect of the experimenter on the results of psychological research. In B. A. Maher (Ed.), *Progress in experimental personality research*. New York: Academic Press, 1964. Pp. 79-114.
- ROSENTHAL, R., FODE, K. L., FRIEDMAN, C. J., & VIKAN-KLINE, LINDA L. Subjects' perception of their experimenter under conditions of experimenter bias. *Perceptual & Motor Skills*, 1960, **11**, 325-331.
- ROSENTHAL, R., PERSINGER, G. W., VIKAN-KLINE, LINDA L., & FODE, K. L. The effect of experimenter outcome-bias and subject set on awareness in verbal conditioning experiments. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 275-283.
- SPIELBERGER, C. D., BERGER, A., & HOWARD, KAY. Conditioning of verbal behavior as a function of awareness, need for social approval, and motivation to receive reinforcement. *Journal of Abnormal and Social Psychology*, 1963, **67**, 241-246.
- TAFFEL, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology*, 1955, **51**, 496-501.

(Received July 10, 1964)