

Experimenter Bias in Hypnotist Performance

by

Suzanne A. Troffer

Charles T. Tart

Reprinted from Science,
September 13, 1964,
Vol. 145, No. 3638,
Pages 1330-1331

Abstract. Eight hypnotist-experimenters administered a standardized suggestibility test to subjects under two separate experimental conditions. The experimenters understood the problem of experimenter bias, knew that they were being checked, and felt that they had treated both groups alike, yet judges were able to tell under which condition the subjects were tested by listening to the performances of the experimenters.

The study reported here resulted from the finding, during a check of experimenter consistency, of significant bias among experimenters participating in an investigation of the effect of two conditions - hypnotic induction and lack of it - on responses to standardized tests of hypnotic suggestibility. The main study will be reported elsewhere (1); it is sufficient to know that the subjects in the experimental group of concern here were individually given one of two parallel versions of Form C of the Standard Hypnotic Susceptibility Scales (SHSS) (2) on two successive days. Eight hypnotist-experimenters participated in the experiment, and no subject was tested by the same experimenter on both days. On the 1st day the subjects were told to imagine that the suggested items were true, and no hypnotic induction was given; on the 2nd day the standard eye-closure induction of the SHSS was given immediately prior to the suggestibility tests. One of the major contributions to be made was between subjects' scores with and without hypnotic induction, the hypothesis being that the subjects would experience more hypnotic phenomena, and thus score higher, in the induction condition than in the imagination (no induction) condition.

Recent studies of experimenter bias (3) and Orne's recent discussions of the importance of implicit demands (4) on subject performance indicate that an experimenter may unknowingly influence subjects to perform in accordance with the experimenter's hypotheses. Hence, it seemed necessary to ascertain whether the experimenters were administering the suggestibility tests of the SHSS in the same manner in both conditions, especially in view of the report of one experimenter (C.T.T.) that he found it difficult to speak as "hypnotically" in the

imagination condition as in the hypnosis condition(5). The problem was discussed with the experimenters (all of whom were graduate students or staff members in psychology, experienced in hypnotic work). They were urged to be as consistent as possible and were informed that practically all experimental sessions would be tape-recorded so that their performances could be judged.

Tape recordings were made with remote recorders, so that the experimenters were not aware of whether or not they were being recorded. After eliminating recordings which were unintelligible because of electrical and mechanical noise, 13 pairs of tapes were secured; in each pair the same experimenter was recorded with one subject in the imagination condition and another subject in the induction condition. The tapes in each pair were presented to the judges in random order. Only the experimenter's voice reading item 1 (hand will lower because it feels heavy) of the SHSS was transferred to the tapes which were to be judged. Item 1 was given immediately following instructions in the imagination condition, and immediately following induction in the hypnosis condition. The voices of the subjects never appeared on these tapes; only the voices of the experimenters reading from a standardized form were heard. Thus any inference as to which condition the subject was under would have to be made from the various paralinguistic features of the recording, such as amplitude, rhythm, pitch, and volume changes.

Seven judges (assistants and staff members of the Laboratory of Human Development) were asked to independently rate which half of each pair was more "hypnotic" in quality. Three of the seven judges (designated C, F, and G.) had served as experimenters in the study, although none of the tapes recorded from judge C were included because of unintelligibility. All the judges were personally acquainted with all the experimenters.

The results for each judge are presented in Table 1. Probabilities were assessed on the null hypotheses that either half of a pair was equally likely to be judged as the "hypnotic" half. The binomial probabilities that the number of correct judgments would be as high as or higher than that obtained by chance alone were calculated directly from binomial tables (6) for each judge's results. As shown in Table 1, three of the judges (B, D, E) scored significantly higher than chance ($p = .05$), three others (A, C, F) came close to significance ($p = .13$), and one judge (G) scored at chance expectancy.

A judge judging his own performance had no advantage. Judge F did no better than five other judges, and judge G was the poorest judge of all. It is of interest that judge G was the only judge who felt that the study would show negative results.

The combined results for all seven judges are highly significant, according to the method of Jones and Fiske (7) (χ^2 , 31.602; degrees of freedom, 14; $p < .005$).

The impressions of the judges as to what their judgments were based on are of some interest. They described the voices of the experimenters under the hypnosis condition as being relaxed, somnolent, solicitous, convinced, dramatic, insistent, coaxing, breathy, sing-song, deeper, slower, soothing, sibilant, softer, droning, descriptive rather than commanding, and as having a sense of hushed intimacy, while experimenters' voices under the imagination condition were described as businesslike, casual, conversational, brisk, alert, natural, prosaic, and rational.

Although the data were insufficient to permit detection of small effects, several other analyses of possible variables influencing the experimenters' performance were undertaken. It was hypothesized that the experimenter might carry out the role of hypnotist more effectively if he were being rewarded by a good performance from the subject, but comparisons did not support this hypothesis. Whether or not a subject successfully passed the first item did not affect the analysis of the judging. Subjects more easily hypnotized than others (judged by total score on the SHSS) might have performed better in the induction period and thus encouraged hypnotic behavior by the experimenter, but comparison of the results for experimenters testing the better subjects with those for experimenters testing the poorer ones showed no differences. It appears, then, that the significant identification of a hypnotic quality in the experimenters' voices by the judges can only be associated with the experimenters having carried out a hypnotic induction procedure.

There are two possible interpretations of this finding, which are not mutually exclusive. The first is that going through an induction procedure gave the experimenters time to firmly establish themselves in the role of hypnotist and that this role carried over into the first test item. The second interpretation is that the experimenters, aware of the experimental hypotheses, unknowingly extended themselves more in the induction condition because of their expectancy that subjects would perform better in this condition and because of their wish to confirm the hypothesis. All the experimenters favor the first hypothesis, insofar as they can judge their own behavior, feeling that it was more natural to act like a hypnotist after "warming up" by way of the induction procedure, a sort of psychological "inertia." Regardless of which interpretation is correct, it is apparent that the experimenters were not consistent in their treatment of the subjects. Since the effects of this inconsistency or bias are confounded with treatment effects, it is not feasible to assess how much effect this bias had on the differences found between groups; but because the basic assumption of identical testing of the groups has been found to be false, the main

experiment was repeated with a tape-recorded testing procedure.

Many psychologists have read of the importance of experimenter bias, but probably feel it is something a sophisticated experimenter (like themselves!) can avoid. Yet in this study a group of sophisticated experimenters, aware of the importance of testing all subjects identically, trying to do so, and knowing that their performance was being recorded for later judging, were nevertheless unable to treat all subjects identically. Nor were these experimenters aware that they had treated the two groups differently.

Subtle differential treatment of groups of subjects which are ostensibly being treated identically sets up demands with different characteristics for each group. The findings of this study thus have important methodological implications for all studies in which it is possible that the performances of the subjects may be affected by subtle demands and expectations (particularly true in hypnosis research), and insofar as the present results are validated by later studies it will become incumbent upon other experimenters either to eliminate such possible bias and demands, or to compensate for their effect in analyzing and interpreting their data.

Laboratory of Human Development,
Stanford University, 514 Alvarado,
Stanford, California 94305

Table 1. Judgments of whether a subject had been hypnotized.

Statistic	Judges						
	A	B	C	D	E	F	G
Total correct judgments among 13 comparisons*	9	10	9	10	10	9	7
P (1-tailed)	.13	.05	.13	.05	.05	.13	.50

* Total $\chi^2 = 31.602$; $df = 14$; $p < .005$.

References and Notes

1. E. R. Hilgard and C. T. Tart, in preparation.
2. A. M. Weitzenhoffer and E. R. Hilgard, Stanford Hypnotic Susceptibility Scale, Form C (Consulting Psychol. Press, Palo Alto, Calif., 1962).
3. R. Rosenthal, Am. Scientist 51, 268 (1963); R. Rosenthal and E. Halas, Psychol. Rept. 11, 251 (1962); R. Rosenthal and G. Persinger, Percept. Mot. Skills 14, 407 (1962).
4. M. T. Orne, J. Abnorm. Soc. Psychol. 58, 277 (1959); Am. Psychol. 17, 776 (1962).
5. The results of this experimenter are not included in the analyses of this paper.
6. H. Walker and J. Lev, Statistical Inference (Holt, New York, 1953), pp. 458-459.
7. L. V. Jones and D. W. Fiske, Psychol. Bull. 50, 375 (1953).
8. Supported by grant M-3859 from the National Institute of Mental Health to E. R. Hilgard during the tenure of one of us (C.T.T.) on USPHS post-doctoral fellowship 1-F2-MH-14,622-01.

29 June 1964