THE ROLE OF VERBAL CUES IN THE MEDIATION OF EXPERIMENTER BIAS

A Thesis

Presented to

the Faculty of Graduate Studies and Research

University of Manitoba

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

Joyce Epstein

September 1966

ABSTRACT

The purpose of the present research was to examine the role of verbal cues as mediators of experimenter bias. Six experimenters tested 120 subjects on a photo judging task. The experimenters were biased to expect high ratings from one half of the subjects and low ratings from the remaining subjects. One half of each of these groups of subjects saw and heard the experimenter giving the instructions for the task; the remaining subjects heard the instructions, but could not see the experimenter.

It was found that the bias of the experimenter was transmitted to the subject, even when visual cues were completely eliminated and the subject could only hear, but not see, the experimenter. It was also found that there was a general raising of the ratings of the photographs when the subject could not see the experimenter, even though bias still appeared.

ACKNOWLEDGEMENTS

The author wishes to express her appreciation to Doctors J. G. Adair and A. Pressey for their encouragement and advice in the preparation of this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

One of the major problems facing social scientists is the attainment of objectivity in research. Usually every attempt is made to ensure maximal objectivity in the methodology and statistical analyses. It is rare, however, to see the same care devoted to the attainment of objectivity in the data collector himself.

The person who collects the data in a psychological experiment brings to the experimental setting a host of ambitions, expectations, hopes, and fears. There is now evidence to indicate that these factors can, and often do, influence the results of otherwise rigorously controlled experiments. The importance for the social sciences of learning more about the operation of such experimenter influences can hardly be overemphasized. Until an understanding of how these factors operate in an experimental setting is attained, only limited confidence can be placed in the results of psychological experiments.

Recently Kintz, Delprato, Mettee, Persons and Schappe (1965) and Rosenthal (1964a) have reviewed the empirical evidence on the experimenters' influence on their data. Perhaps the most striking conclusion of both reviews is that relevant studies are lacking, except in the clinical areas. Kintz *et al.* comment on the "unconcerned attitude toward this phenomenon (that) has been taken, especially by

experimental psychologists." Rosenthal notes that reports on effects of experimenter bias "are not numerous and virtually never published."

While there has not been much experimental interest in the area, it was recognized as a problem some time ago. Probably the first evidence relating to the influence of the experimenter involved Clever Hans, a horse that was apparently able to spell, read and do arithmetic by tapping his hoof. After carefully studying the phenomenon, Pfungst (1911) concluded that the people questioning the horse cued him unintentionally. A forward inclination of the questioner's head served as a signal for Hans to begin tapping. A slight upward motion of the head as the number of taps approached correctness was a signal for Hans to stop. Even skeptical questioners expected Hans to know the correct answers and this expectation was reflected in their subtle signals. As Pfungst stated, the difficulty in understanding the phenomenon of Clever Hans arose from "looking for, in the horse, what should have been sought in man."

Clinical psychologists have long been aware of the difficulties that arise from looking for, in the patient, what should be sought in the therapist. Mann, Menzer and Standish (1950) studied the conscious and unconscious motivations of the therapist and found that these affected certain aspects of treatment. Clinical psychologists have also directed attention to the influence of the psychologist in the testing situation. Joel (1949) noted that projective

The problem of experimenter expectancy effects was again set aside for a number of years. During this time though, studies appeared which demonstrated the effect of various experimenters' characteristics on subjects' responses: experimenters' sex (Binder, McConnell and Sjoholm, 1957), religion (Robinson and Rohde, 1946), race (Williams and Cantril, 1945) and status (Birney, 1958). Other attributes of the experimenter studied were warmth (Ferguson and Buss, 1960), likeability (Sapolsky, 1960), acquaintanceship (Sacks, 1952) and personality (Young, 1959).

Orne (1962) is one of the researchers who has revived concern over the problem of experimenter influences. He has shown that the subject's behavior in an experiment is a function, not only of the experimental variables being investigated, but also a set of variables he has termed the "demand characteristics" of the situation. The demand characteristics include campus rumors about the research, information conveyed during the original solicitation, the setting of the laboratory and, most important for the present study, the person of the experimenter. These are the cues which convey to the subject the true purpose of an experiment. The subject, in his attempts to play the role of a "good subject", is always trying to validate the experimental hypotheses. The demand characteristics cue him as to what these hypotheses might be, and hence how he should behave. The personal attributes of the experimenter provide a source for such behavioral cues.

It is Orne's belief that the demand characteristics cannot be eliminated from the experiment. Their effects can, however, be studied. Orne has proposed several ways of doing this. One way is to examine the subject's perception of the experimental hypotheses. Another is to hold the demand characteristics constant and eliminate the experimental variable by employing simulating subjects to see if judges can distinguish between these and real subjects. A third technique is to vary the demand characteristics and to study (real) subjects' understanding of the hypotheses without administering the treatment. These procedures were intended as a means of assessing the effect of the demand characteristics. The specific contribution of experimenter bias as a single demand characteristic cannot necessarily be assessed by these methods.

McGuigan (1963) has also recognized the effect of the experimenter in psychological research. He has stated the problem in a cryptic but accurate fashion: the experimenter is a necessary but undesirable factor in an experiment. He outlines the difficulties which arise from using either one or many data collectors in an experiment, although he recommends the latter. He encourages further study of the experimenter variable, and proposes as the only other alternative the use of completely automated devices in the place of the experimenter. He points out that the use of a tape recorder, for example, eliminates visual cues and olfactory stimuli associated with the experimenter.

Rosenthal is another researcher concerned with experimenter effects. At the beginning of the present decade, Rosenthal (Rosenthal and Fode, 1963) began a program of research to examine the role of the experimenter (E) as an unintended determinant of the results of psychological experiments. The procedure required subjects (Ss) to rate between -10 and +10 the apparent success or failure of 20 persons pictured in photographs of "neutral" stimulus value. The pictures had been selected so that by chance alone the ratings would not be expected to deviate significantly from zero. Ten psychology students served as Es. One half of the Es were told to expect a mean rating of +5 from their Ss and the remainder were told to expect a rating of -5. The instructions given by all Es to their Ss were identical. The results indicated that the bias or expectancy of the E influenced the Ss' responses. Those Ss tested by Es who expected a rating of +5 rated the photos significantly higher than the Ss tested by Es who expected a rating of -5.

Rosenthal and his colleagues have studied the expectancy effect in relation to many characteristics of E and S. In a discussion of his work (1964) he reported unpublished studies relating expectancy effects to the Es' need for approval, as measured by the Marlow-Crowne Social Desirability Scale. These studies, as well as those employing the L, K and Pt scales of the MMPI (Rosenthal, Persinger and Fode, 1962) show that Es' need for approval is correlated with the magnitude of their expectancy effects. However, Ss'

It is Orne's belief that the demand characteristics cannot be eliminated from the experiment. Their effects can, however, be studied. Orne has proposed several ways of doing this. One way is to examine the subject's perception of the experimental hypotheses. Another is to hold the demand characteristics constant and eliminate the experimental variable by employing simulating subjects to see if judges can distinguish between these and real subjects. A third technique is to vary the demand characteristics and to study (real) subjects' understanding of the hypotheses without administering the treatment. These procedures were intended as a means of assessing the effect of the demand characteristics. The specific contribution of experimenter bias as a single demand characteristic cannot necessarily be assessed by these methods.

McGuigan (1963) has also recognized the effect of the experimenter in psychological research. He has stated the problem in a cryptic but accurate fashion: the experimenter is a necessary but undesirable factor in an experiment. He outlines the difficulties which arise from using either one or many data collectors in an experiment, although he recommends the latter. He encourages further study of the experimenter variable, and proposes as the only other alternative the use of completely automated devices in the place of the experimenter. He points out that the use of a tape recorder, for example, eliminates visual cues and olfactory stimuli associated with the experimenter.

need for approval was not found to be related to susceptibility to E expectancy effects.

Studies relating E's anxiety level to the bias effect have yielded inconsistent results. Rosenthal (1964) reports that some studies have shown that Es with medium levels of anxiety, as measured by the Taylor Manifest Anxiety Scale, showed the greatest expectancy bias; in other studies, Es with high levels of anxiety showed the greatest expectancy bias; and in still others, Es with low anxiety had the greatest expectancy effect. The same inconsistency marks the findings on the anxiety level of Ss and their susceptibility to bias effects.

Several studies (Rosenthal, Persinger, Mulry, Vikan-Kline and Grothe, 1964a; Rosenthal, Persinger, Mulry, Vikan-Kline and Grothe, 1964b) have examined the influence of Es' and Ss' sex on the bias effect. Both studies found that the conditions most conducive to the appearance of the bias effect occurred when Es were male and Ss were female.

One study (Rosenthal, Friedman, Johnson, Fode, Schill, White and Vikan-Kline, 1964) investigated the effects of motivation on the operation of E expectancy effects. Half the Es in this study were excessively motivated by telling them that if the data they obtained were "better" than that of an unknown partner they would be paid not only their own $1.00 but their partner's as well. If, on the other hand, their "partner" did a better job, the partner would get both rewards. Results showed that moderately motivated Es

obtained data more in accord with their expectancies than did excessively motivated Es. Excessively motivated Es, in fact, tended to obtain data opposite to what they had been led to expect. In sum then, Rosenthal has found that Es who have a greater need for social approval obtain more of the sort of data they are led to expect. Moderately motivated male Es with female Ss yield the greatest expectancy effect. Es' and Ss' anxiety level, however, are related to expectancy effects in an as yet undetermined manner.

In addition to studying the bias effect in relation to characteristics of Es and Ss, such as anxiety, motivation, and need for approval, Rosenthal has also examined bias in relation to the experimental situation. In Rosenthal's early research, the Es were all led to expect that their Ss would give one specific type of response. In "real" psychological experiments, Es generally expect one of several different kinds of responses from their Ss, depending upon the condition to which S has been assigned. Thus, for example, E may first collect data from Ss in one experimental condition, and then from Ss in the other condition or conditions. In other studies, data from Ss representing all conditions may be collected on the same occasion, with Ss appearing in random order.

In order to assess whether E expectancy effects occur in situations which more closely approximate those of "real" experiments, Rosenthal conducted the following study (Rosenthal, Persinger, Mulry, Vikan-Kline and Grothe, 1964a).

Initially, five Es were biased to expect +5 ratings (or -5 ratings, depending upon to which group E was assigned) from Ss on the photo judging task. Several weeks later, each E ran another group of Ss. This time, the Es were given a bias opposite to that which they had held before. This was explained to Es simply on the basis that these new Ss were of a different personality type. The results indicated that reversing Es' bias in this manner produced the expectancy effect. When E was biased to expect +5 ratings, the Ss rated the photos significantly higher than when the same E was biased to expect -5 ratings.

In the second part of the experiment, the bias was induced in E in a random fashion. Six Es collected data on the photo judging task from approximately six Ss each. Es were told that their Ss were of two personality types - some would average +5 ratings and some -5 ratings. Before meeting each S, E was told to which "group" S belonged. Results showed that differences in ratings of the Ss under these conditions were not significant. Thus, although random variation of E expectancy did not lead to changes in Ss' ratings, there is support for Rosenthal's suggestion that Es may alter their hypotheses in mid-experiment and obtain data in accord with the revised expectancy.

Rosenthal's research includes the study of a phenomenon he has termed "modeling" effects. The extent to which an S's performance is predictable from his experimenter's own performance on the same task is the extent of the modeling

effect. Rosenthal (1963) found that in a series of eight studies, the modeling effect decreased from the earlier to the later studies until it was actually reversed. This suggests that in later studies Es were more likely to suspect that they themselves were the objects of study. As they became more conscious of their influence, they were perhaps more on their guard to avoid any possible sort of effect. The modeling effects may have been so carefully guarded against as to reverse the bias.

Another phenomenon studied by Rosenthal is the "monitoring" effect. One study (Rosenthal, Persinger, Mulry, Vikan-Kline and Grothe, 1964a) provided evidence to suggest that monitoring of Es' procedure by a senior experimenter (one of the authors) interferes with the appearance of the bias effect. The presence of a third person in the data-collecting situation sometimes increased the bias effect, more often decreased it, and at other times reversed it.

Rosenthal and his students have examined many other aspects of the experimenter expectancy problem. They have studied the effect of early data returns on biased Es (Rosenthal, Persigner, Vikan-Kline and Fode, 1963), demonstrated the role of E bias in verbal conditioning experiments (Rosenthal, Persinger, Vikan-Kline and Fode, 1963a), distinguished between E's expecting and desiring certain results (Rosenthal, Kohn, Greenfield and Carota, 1966) and have employed sound motion pictures to study behavioral characteristics associated with E bias (Friedman, Kurland and

Rosenthal, 1965).

Rosenthal has also extended his work on experimenter expectancy effects to include animal studies (Rosenthal and Halas, 1962; Rosenthal and Fode, 1963a; and Rosenthal and Lawson, 1964). Es who were led to believe that their rats were bright obtained performance from them significantly superior to that obtained by Es who believed their rats to be dull. A post-experimental questionnaire revealed that Es running "bright" rats handled their Ss more which may have facilitated learning. Brogden (1962) and Cordaro and Ison (1963) have confirmed the existence of expectancy effects when Ss were rabbits and planaria, respectively.

The great importance of this large body of research is that it has served to establish the experimenter expectancy effect as a real phenomenon, the existence of which the experimental psychologist can no longer deny. However, although a great deal is now known about the phenomenon, there is still little information as to how the expectancies of the E are transmitted to the S. What are the cues that mediate bias?

Rosenthal, Fode, Friedman and Vikan-Kline (1960) examined Ss' perception of biased Es. They found that E bias correlated with such E attributes as "friendly", "interested", "expressive-voiced" and "use of hand, head and arm gestures". This suggests that kinesic and paralinguistic aspects of E's interaction with his Ss serve to communicate his bias to his Ss. Fode (1960) compared the role of visual (kinesic) and

verbal (paralinguistic) cues in the transmission of E bias. He employed four groups into which Es were randomly assigned. Group 1 (-5 bias replicate) was given the low (-5) bias instructions and was designed to serve as a control for the other three groups as well as a replication of Rosenthal's original study. In Group 2 (non-visual (+5) bias) Es greeted all Ss but then sat down behind a screen across the table from S. The E read instructions to S and recorded S's ratings, but was not visible to S. This group, as well as the remaining ones, had been biased to expect high (+5) ratings. In Group 3 (non-verbal (+5) bias) Es handed Ss the sheet of instructions and remained completely silent until the end of the session. These Es were across the table from S in full view. Group 4 (+5 bias replicate) was a replicate of the high (+5) bias group in Rosenthal's original study.

Fode analyzed the data by comparing Groups 2, 3 and 4 with Group 1. He found E bias in all comparisons except with Group 3. The failure of Group 3 to manifest the expected bias suggested to Fode that E bias may be communicated primarily by verbal factors. He concluded that "verbal cues from E to S are sufficient to mediate E bias" (Fode, 1960, p. 44). However, the difficulty in drawing this conclusion lies in the fact that Groups 2 and 3 differed in another way besides the method of delivering the instructions. In Group 2, which Fode designated as non-visual, E first greeted and seated the S and then disappeared behind a screen, whereas in Group 3, E said nothing at all

to the S. There is a possibility that the brief pre-instructional interaction between E and S in Group 2 contributed to mediating the bias, rather than the exclusive presence or absence of verbal cues during the instructional phase.

An analysis by Rosenthal (Rosenthal; Fode, Vikan-Kline and Persinger, 1964) of three experiments suggested that mediation of the bias may occur before the experiment proper actually begins, that is, during the pre-instructional interaction. This was based on the finding that the bias phenomenon was already in evidence on Ss' very first response in each of the three experiments analyzed. Rosenthal recommended that attention be directed to the brief pre-data-collecting phase of the E-S interaction. The bias may thus be transmitted in part while E "greets and seats" S.

Viewed in the light of these findings, Fode's (1960) results may have been due, not to verbal cues alone, but to the interchange that went on prior to the reading of the instructions. It will be recalled that, before going behind the screen in the non-visual group, Fode's Es first greeted S and seated him.

Statement of the Problem

While it has been found that the expectancies of the experimenter are transmitted to the subject, i.e., that E bias exists, and that paralinguistic and kinesic factors may be involved in the transmission of bias, Fode's (1960) attempt to eliminate kinesic (visual, from S's point of view) factors

and to demonstrate the mediation of bias by verbal factors was inconclusive. Although he concluded that verbal cues are sufficient to mediate E bias, his procedure did not effectively eliminate visual factors because E was visible to S for a brief interaction period prior to the data collection.

The present study examined the role of verbal cues more closely by eliminating all visual cues; those present in the pre-instructional phase as well as those present in the experiment proper. Under these conditions it is possible that the evidence that Fode found would not be obtained.

To examine this, four experimental groups were created. Two groups consisted in the typical Rosenthal bias conditions, i.e., +5 and -5 bias conditions, in which the E was visible (visual condition). The other two groups consisted in remote-controlled taped instructions of biased Es reading the instructions. In the latter condition the Ss heard the instructions, but the E was not visible (non-visual condition). The hypothesis was that the E bias effect would be significant in the visual condition, but not in the non-visual condition. Rejection of the hypothesis would mean that verbal cues alone are sufficient to mediate E bias. Knowledge of the mode of transmission of the bias could have important ramifications for future research. The value lies in possible control of such bias effects when their source is known, or if not, at least awareness that they exist and of how they are communicated.

CHAPTER II

METHOD

Subjects

One hundred and twenty female students, enrolled in an introductory psychology course at the University of Manitoba, served as Ss as part of a research requirement.

Experimenters

Six male students from an undergraduate course in social psychology were employed as Es. All Es were volunteers.

Design

Each of the six Es tested ten Ss on the photo judging task. One-half of the Es were led to expect high ratings from the first five Ss and low ratings from the last five Ss. The remaining Es were led to expect low ratings from the first five Ss and high ratings from the last five Ss. The instructions given by the six Es to their 60 Ss were tape recorded. The taped instructions were then given to another 60 Ss. Thus there were two conditions, visual and non-visual (taped), and two groups in each condition, high bias and low bias.

Group 1 - Thirty Ss were assigned to this group. All six Es were given a high bias (+5), but three of them had previously held low biases. The instructions Es read to Ss were recorded.

Group 2 - Thirty Ss were assigned to this group. All six Es were given a low bias (-5), but three

of them had previously held high biases. The instructions were recorded.

Group 3 - The 30 Ss assigned to the non-visual condition heard the instructions taped by Es with a high bias (+5). There was no E visible at any time.

Group 4 - Also a non-visual condition; the 30 Ss heard instructions taped by the Es with a low bias (-5). There was no E visible at any time.

Materials

A tape recorder was used in both the visual and non-visual conditions. In the former condition it was recording; in the latter condition, it was playing. In neither case was it visible to either E or S. Concealment of the tape recorder was necessitated by the finding (Rosenthal, 1964) that monitoring E interferes with the bias effect.

The rating scale (see below) was placed at the left-hand side of the table at which S was seated to enable S to refer to the scale while rating the photographs. A rating form was available to all Ss on which to record their responses. A sign directed Ss in the non-visual condition to be seated at the table and to await instructions.

EXTREME FAILURE	MODERATE FAILURE	MILD FAILURE	MILD SUCCESS	MODERATE SUCCESS	EXTREME SUCCESS

-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10

Rating Scale

Standardization of Photographs

Forty-six photographs of mens' faces, cut from popular magazines and mounted on 3x5 inch white cards were presented to 100 female introductory psychology students. These Ss were seen individually by the senior experimenter (author) and were instructed to rate each face according to the degree of success or failure the person pictured in the photo appears to have experienced. The exact instructions read to Ss followed those used by Rosenthal (1964, p. 91):

> "I am going to read you some instructions. I am not permitted to say anything which is not in the instructions nor can I answer any questions about this experiment. OK?
>
> "We are in the process of developing a test of empathy. This test is designed to show how well a person is able to put himself into someone else's place. I will show you a series of photographs. For each one I want you to judge whether the person pictured has been experiencing success or failure. To help you make more exact judgments you are to use this rating scale... As you can see, the scale runs from -10 to +10. A rating of -10 means that you judge the person to have experienced extreme failure. A rating of +10 means that you judge the person to have experienced extreme success. A rating of -1 means that you judge the person to have experienced mild failure, while a rating of +1 means that you judge the person to have experienced mild success. You are to rate each photo as accurately as you can. Just tell me the rating you assign to each photo. All ready? Here is the first photo. (No further explanation may be given although all or part of the instructions may be repeated.)"

From the original forty-six photographs, ten of neutral stimulus value were selected. Each of the photographs selected were those which had been rated closest to zero, and none had been rated higher than $\pm$ 1.50. The sum of the ratings of the ten photographs was zero. These ten photos were mounted in a row on a cardboard strip. This procedure

was necessitated by the inability of E to handle the photographs in the non-visual condition. In Rosenthal's original study, the photos were presented to S individually.

Procedure

Each E volunteered for two hours of testing, worked out on the basis of his class schedule. When he arrived at the appointed time, he was assigned by the senior experimenter to either the group with the +5 bias or the group with the -5 bias. This was done on an alternate basis. He was given his instructions in a room some distance from the testing room. E was then left alone while the senior experimenter went to the testing room to turn on the concealed tape recorder. E believed he was being left alone in order to have an opportunity to study the instructions. E was then taken to the testing room and tested five Ss individually. Approximately one week later, E returned for a second session. He was given the opposite bias and then tested five more Ss.

All Es were motivated by their instructor at the time they were recruited by being told that their performance would reflect on their ability as experimental psychologists and that their participation in this study would provide valuable experience. The exact instructions to E when they arrived to take part in the experiment were as follows:

> "You have been asked to participate in a research project developing a test of empathy. The main reason for your participation in this project is to standardize results of experiments of this type. You may be aware of the problem in psychological research of

different examiners getting somewhat different data on the same tests as a function of individual differences. Therefore, to standardize the tests, it is better methodological procedure to use groups of experimenters.

"There is another reason for your participation on this project - namely, to give you practice in duplicating experimental results. In physics labs, for example, you are asked to repeat experiments to see if your findings agree with those already well established. You will now be asked to run a series of Ss and obtain from each, ratings of photographs. The experimental procedure has been typed out for you and is self-explanatory. DO NOT DISCUSS THIS PROJECT WITH ANYONE until your instructor tells you that you may.

"According to preceding research of this nature, the type of subjects that you will be using have averaged a +5 rating. Therefore, the subjects you are running should also average about a +5 rating.

"Just read the instructions to the subjects. Say nothing else to the subjects except what is written in the procedure. If for any reason you should say anything to a subject other than what is written in your instructions, please write down the exact words you used and the situation which forced you to say them.

" GOOD LUCK!"

All Es received the same instructions except that half the Es were biased to expect +5 ratings from their Ss and half were biased to expect -5 ratings. The biases for each E were reversed at the second session. The instructions are essentially the same as those used by Rosenthal (1964).

Each S signed up for the experiment in the usual manner. On the appointment slip was written the number of the testing room and the time of the experiment. Since testing time was about five minutes, Ss in the non-visual condition were spaced at 20-minute intervals to make certain that each S was gone when the next S arrived.

The first five Ss were assigned to Group 1, the second five to Group 2, the third five to Group 1, and so on for 60

Ss. The last 60 Ss were assigned similarily to Groups 3 and 4. This arrangement for assigning Ss was necessitated by the experimental design. Each of the six Es tested ten Ss, five Ss at a time. Thus the first 60 Ss had to be assigned to each bias condition in groups of five. Although it would have been possible to assign the next 60 Ss in random order, since they received the taped instructions, it was considered more consistent to assign them to the bias conditions in the same manner as the first 60 Ss.

Ss in Groups 1 and 2 were greeted by E and asked to be seated. E then began to read the instructions. Ss in Groups 3 and 4 saw a prominently-placed sign when they entered the room asking them to be seated and informing them that a tape-recorded message would give them their instructions. The instructions were then played, the tape recorder being operated from behind a one-way screen.

The exact instructions to Ss were as follows:

> "I am going to read you some instructions. No questions can be answered about this experiment. OK?
>
> "We are in the process of developing a test of empathy. This test is designed to show how well a person is able to put himself into someone else's place. Before you are a series of photographs. For each one, I want you to judge whether the person pictured has been experiencing success or failure. To help you make more exact judgments, you are to use the rating scale on your left. As you can see, the scale runs from -10 to +10. A rating of -10 means that you judge the person to have experienced extreme failure. A rating of +10 means that you judge the person to have experienced extreme success. A rating of -1 means that you judge the person to have experienced mild failure while a rating of +1 means that you judge the person to have experienced mild success. You are to rate each photo as accurately as you can. Just mark down the rating you assign to each photo on the form

provided. Indicate the number of the rating and the sign. All ready? You may begin."
(Rosenthal, 1964)

All Es read the same instructions to all Ss.

CHAPTER III

RESULTS

Each E obtained ratings of the ten photographs from ten Ss in the visual condition and from ten Ss in the non-visual condition. Es were given a +5 bias for one half of their Ss and a -5 bias for the remaining Ss in each condition. Each S's rating of the ten photographs were summed and the means were computed. Thus there were 20 mean ratings for each of the six Es, five in each of the four treatment combinations. These 120 measures constituted the data for the experiment and are presented in Table I, with the means for each E under each condition. A graphic representation of the results is shown in Figure 1.

There were several alternative designs available for the analysis of the data. In a 3-factor (ABC) design it is necessary to have variables with fixed, known values. Since the six Es in this experiment were regarded as a random sample drawn from a population of such Es, the ABC design was not considered to be appropriate. Because the same E was used in all four treatment combinations (two modes of presenting the instructions and two types of bias) it seemed best to conceptualize the experiment as one in which the same 2-factor study was replicated six times. This is the ABR design discussed by Lindquist (1953). Table II presents a summary of the analysis of variance.

The triple interaction in this analysis was not significant. This means that the corresponding interaction effects

TABLE I

Mean Ratings of Photographs Obtained by Experimenters Under Each of the Four Treatment Combinations

	Experimenter 1				Experimenter 2				Experimenter 3			
	A_1		A_2		A_1		A_2		A_1		A_2	
	B_1*	B_2	B_1	B_2	B_1	B_2*	B_1*	B_2	B_1	B_2*	B_1	B_2
	3.7	2.5	3.9	4.0	4.8	1.0	7.0	3.0	3.5	1.1	6.6	5.9
	4.7	1.3	4.4	5.8	5.8	2.1	6.0	3.9	1.1	2.2	7.0	6.2
	3.4	-1.8	6.7	1.2	2.5	1.0	4.8	4.1	6.7	2.0	4.9	1.6
	2.6	3.1	4.8	.7	.5	.9	4.7	5.0	3.5	- .2	6.6	5.0
	4.0	2.0	7.0	3.7	3.5	2.2	5.2	5.0	3.6	.1	7.3	6.6
$\bar{x}$	3.7	1.4	5.3	3.1	3.3	1.5	5.5	4.2	3.5	1.0	6.4	5.0
$\bar{x}$	2.6		4.2		2.4		4.9		2.4		5.8	

CODE - A_1 = visual condition, A_2 = non-visual condition,
B_1 = high (+5) bias, B_2 = low (-5) bias

* Bias condition E experienced first

TABLE I (Continued)

Experimenter 4				Experimenter 5				Experimenter 6			
A_1		A_2		A_1		A_2		A_1		A_2	
B_1	B_2^*	B_1	B_2	B_1^*	B_2	B_1	B_2	B_1^*	B_2	B_1	B_2
.5	-1.1	.1	1.3	2.0	.1	- .5	4.1	1.9	3.5	.3	1.6
3.2	1.5	7.0	-1.7	.1	2.7	4.0	6.2	5.3	3.1	2.2	.7
1.5	3.0	6.0	.5	4.3	3.2	1.3	1.2	1.6	- .7	2.4	2.7
.7	2.2	4.7	1.6	1.2	3.9	1.4	2.7	1.8	1.6	-1.3	1.2
.5	1.4	4.8	- .5	2.1	.9	5.0	.7	.8	- .5	5.2	2.4
$\bar{x}$ 1.2	1.4	4.5	.2	1.9	2.1	2.2	2.9	2.3	2.4	1.7	1.7
$\bar{x}$ 1.3		2.4		2.1		2.6		1.8		1.7	

CODE - A_1 = visual condition, A_2 = non-visual condition,
B_1 = high (+5) bias, B_2 = low (-5) bias

* Bias condition E experienced first

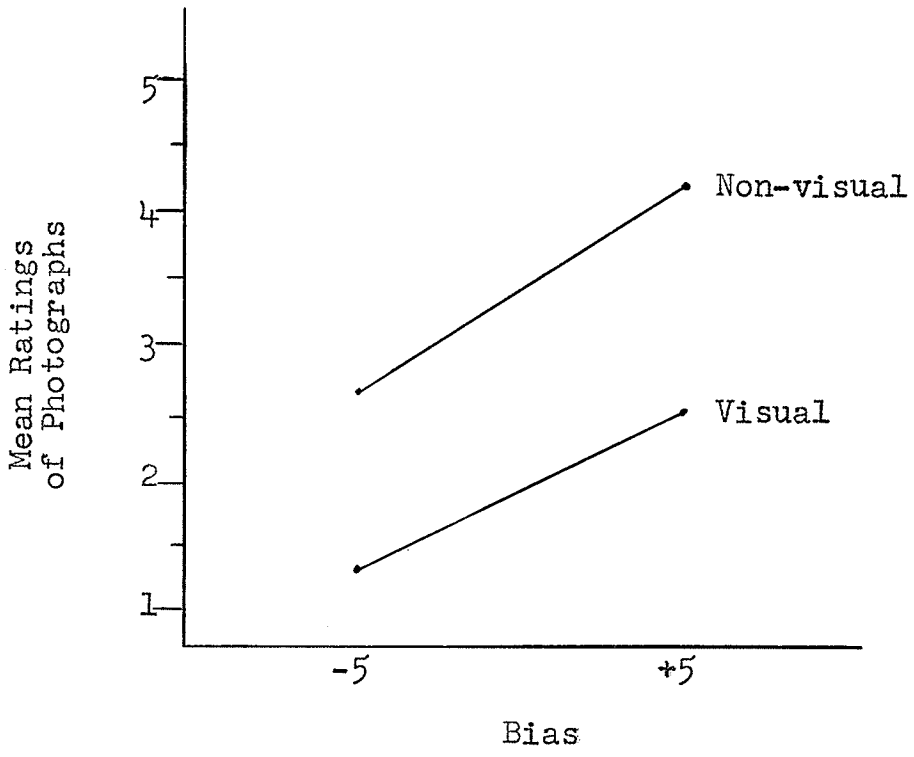

Fig. 1. Mean ratings assigned to the photographs by subjects of experimenters biased to expect +5 and -5 ratings, in visual and non-visual conditions

TABLE II

Summary of Analysis of Variance

Source	df	SS	M.S.	F
Mode of Presentation (A)	1	68.49	68.49	8.55**
Bias (B)	1	54.57	54.57	9.22**
Experimenters (R)	5	96.12	19.22	
Mode of Presentation x Bias	1	.50	.50	.09
Mode of Presentation x Experimenters	5	40.06	8.01	2.81*
Bias x Experimenters	5	29.63	5.92	2.07
Mode of Presentation x Bias x Experimenters	5	27.20	5.44	1.95
Within Cells	96	274.34	2.85	
TOTAL	119	590.91		

* Significant at .025 level

** Significant at .05 level

for the various bias by mode of presentation combinations did not differ significantly across Es (replications). Since the ABR interaction was non-significant, it was possible to look at the AB interaction, which is most relevant to the hypothesis.

The hypothesis was that the bias effect would not be significant in the non-visual condition, but that it would be significant in the visual condition. In analysis of variance terms, this implies that there would be a significant bias by mode of presentation interaction. The analysis of variance indicates that the AB interaction was not significant. Thus the hypothesis that there would be a bias effect in one condition and not in the other may be rejected.

The rejection of this hypothesis, as originally stated, is further supported by the occurrence of a significant main effect for bias ($F = 9.25$, $p < .05$). Es in whom a high expectancy was created obtained higher ratings from their Ss than Es in whom a low expectancy was created. This was true for both the visual and the non-visual (taped) conditions.

It should be noted here that the effects of bias did not differ across Es. The non-significant BR interaction shows that the bias effect may be considered as basically similar for all Es.

It should also be noted that even though the present study employed an entirely different set of photographs than that used by Rosenthal, the bias effect still appeared.

Thus the bias phenomenon was not peculiar to the photographs Rosenthal employed. The present photographs did differ from Rosenthal's in that, in the standardization procedure, they received mean ratings no higher than ± 1.50. The pictures Rosenthal selected had been rated, during the standardization, no higher than ± 1.00. Another difference, and this may have been due to the difference in ratings during standardization, was that the present photographs received mean ratings under all experimental conditions that were generally higher than those reported by Rosenthal.

Although the bias effect did not differ for the two modes of presenting the instructions, there were differences among Es in the ratings obtained under the two conditions. In analysis of variance terms, this means that the AR interaction was significant ($F = 2.81$, $p < .025$). This interaction is presented graphically in Figure 2. From this it may be seen that three Es obtained much higher ratings in the non-visual or taped condition than in the visual condition, two Es were slightly higher, and one E was lower in the non-visual than in the visual condition.

Figure 2 also shows that there was greater variability of ratings obtained by Es in the taped condition, with mean scores ranging from 1.7 to 5.8. The spread in the visual condition (from 1.3 to 2.6) was not so large. Thus it may be seen that the scores obtained by most Es in the visual condition differed from their scores in the non-visual condition.

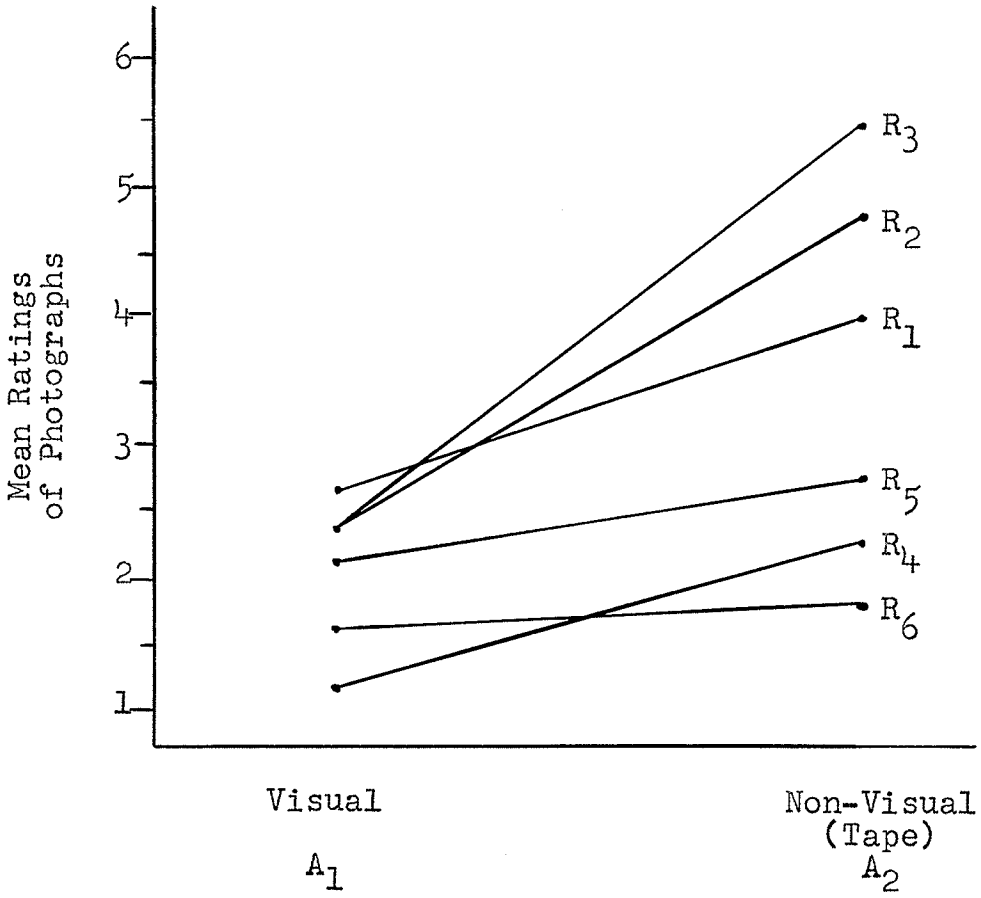

Fig. 2. Mean ratings obtained by experimenters (R) in the visual and non-visual (taped) conditions

A significant main effect for mode of presentation was also found ($F = 8.55$, $p < .05$). The photographs were rated differently by Ss in the visual condition from Ss in the non-visual condition. From Figure 1 it can be seen that mean ratings were higher when instructions were presented on tape than when they were presented in person, even though the bias effect still appeared. This main effect for mode of presentation was primarily due to the performance of only three of the Es, as evidenced by the significant AR interaction.

CHAPTER IV

DISCUSSION

The results clearly provide support for Fode's (1960) suggestion that verbal cues are sufficient to mediate E bias. In the non-visual condition of the present experiment, visual cues were completely eliminated, since at no time did S ever see E. Nevertheless, the expectancies of E, whether for high or for low ratings of the photographs, were conveyed to S. Elimination of visual cues did not prevent the transmission of E bias.

The suggestion that verbal cues are important in the communication of bias receives additional support from related studies in other areas. Troffer and Tart (1964) reported a study in which hypnotist-experimenters administered a suggestibility test to Ss under two conditions. In one condition, the test was given after E had hypnotized S. In the other condition, the test was given after E had simply told S to imagine that the suggested test items were true. In both cases, the hypnotist-experimenter was cautioned to be as consistent as possible in administering the suggestibility test. Administration of the test was tape recorded. Seven judges rated the recordings as to the amount of "hypnotic" quality they possessed. Six out of the seven were able to pick out, at better than chance level, which items belonged to the "hypnotic" condition and which to the "imagination" condition, even though all Es felt they had

spoken the same whether or not S was hypnotized. Troffer and Tart concluded that, "the experimenters, aware of the experimental hypotheses, unknowingly extended themselves more in the induction (hypnotic) condition because of their expectancy that subjects would perform better in this condition..." This expectancy was transmitted by subtle verbal cues to the judges.

Another study which emphasized the importance of verbal cues in the transmission of E influence was that of Barber and Calverley (1964). In this study, variations in E's tone of voice produced clear-cut variations in Ss' responsiveness to suggestions. When E presented suggestions from the Barber Suggestibility Scale in a forceful tone, Ss manifested higher levels of suggestibility than when the suggestions were presented in a lackadaisical tone.

Thus there seems to be considerable support for the idea that verbal cues are an important factor in the mediation of E influence. However Rosenthal (1966), in his most recent and as yet unpublished research employing motion pictures to analyze experiments on E bias, has stated that "we have yet to find the specific cues that mediate the 'Clever Hans' phenomenon to human subjects." By using motion pictures it was possible to predict whether an E would subsequently influence his Ss to respond in accordance with his hypothesis, simply from observing E's initial interaction with his Ss. Although correlations were not high (from .30 to .43), they were better for those observers

using silent films than for those using the sound track to make the prediction. Thus tone of voice variables, in this instance, seemed to be of little consequence. Another unpublished study by Rosenthal, however, showed a number of tone of voice variables to be significant predictors of E's biasing effects and very often the prediction judged from the sound track was opposite to that based on the silent film. Thus, there seemed to be some "channel discrepancies", and if verbal cues are not the only mediators of bias, they are at least important mediators of bias.

Another interesting finding of the present study was that there was a significant difference between presenting instructions "in person" and presenting them on tape. The photographs were rated on the average as more successful by Ss in the non-visual condition, even though the bias effect was still present.

There was also a difference in variability of ratings between the visual and taped conditions. Figure 2 illustrates the greater variability in the non-visual condition.

It should be noted that in the present study the design required Es to change their expectancy from +5 to -5 ratings (or vice-versa) in mid-experiment. Es first tested five Ss under one bias condition, then another five Ss under the opposite bias. The more typical Rosenthal method employs different Es for each of the bias conditions. Only in one previous study (Rosenthal, Persinger, Mulry, Vikan-Kline and Grothe, 1964a) has there been an attempt to change E's

bias in mid-experiment. In that study, the change in bias did result in a corresponding change in Ss' ratings. The present study lends further support to the plausibility of the suggestion that Es may alter their hypotheses in mid-experiment and then obtain data in accord with the revised expectancy. The bias effect appeared, even though the same Es were used in both bias conditions.

Since the present study found that E influence cannot be removed by removing visual cues, the problem of how experimenter influence can be eliminated remains. McGuigan has expressed concern over this problem. He has encouraged psychologists to study the experimenter variable more carefully. According to McGuigan, this should not be very difficult, since experiments usually employ more than one data collector, making it a simple matter to analyze and report the data as a function of Es. Interactions between Es and treatments may then be tested.

McGuigan has proposed the removal of the E from the data collecting situation as the only solution to elimination of the E bias variable. This could be accomplished by the use of automated devices in the place of E. He has recommended the use of a tape recorder, for example, to eliminate the influence of E in an experiment.

The present results show that this may not always be successful, but the way in which the present study used a tape recorder is not typical of the methods employed in other studies. In most procedures involving tape recorders,

Es are fully aware that their voice is being taped. In fact, E very often operates the tape recorder himself. The present study employed a concealed tape recorder, so that E was unaware that his instructions were being taped. It was feared that if E knew his voice was being recorded, this would interfere with the bias effect. Since the purpose of the study was to examine the mode of transmission of bias, and not to control it, E was kept ignorant of the tape recorder.

The suggestion that awareness of the recorder would interfere with the bias effect was based on Rosenthal's (1964) research on monitoring effects. He found that when Es were biased to expect certain results, and were subsequently supervised by a third person while testing their Ss, the usual bias effect did not occur. Taping the instructions may be regarded as a form of monitoring. Thus, it is possible that using taped instructions, recorded with E's full awareness may be a more appropriate control of the bias effect than using instructions taped without E's knowledge, as in the present study.

To test the value of automated devices in research as McGuigan suggested, it would be necessary to tape Es with a fixed bias. E would be given an expectancy either for high or low ratings, and would read the instructions to the Ss, knowing that his voice was being taped. The data would then be analyzed to determine whether the bias effect occurred. If the data indicated that there was no significant bias, then

McGuigan's suggestion that the use of a tape recorder removes E influence would be supported.

Another way that tape recorders may be used successfully in psychological research is to control variability. Barber and Calverley (1964) have recommended the use of tape recorders for this purpose. If the instructions for a given experiment were taped, the same instructions could be given to all Ss, and variability among Es, or within a single E at different times, would be reduced.

A second possible method of overcoming E influence may be to eliminate sources of verbal cues by presenting instructions in written form, either on paper or projected on a screen. This could be done with or without E present. S would read a set of written instructions, rather than hear E read them. Fode's (1960) study gives some support to the validity of this procedure. Ss read the instructions from a prepared sheet while E was present in the room. A significant bias effect did not occur under these conditions.

A third possibility would be to keep the data collectors ignorant of the experimental hypotheses. The difficulty in this is that data collectors who are kept ignorant of the hypotheses may become bored by the experiment. Disinterest may affect the results as much as bias does. A way of avoiding the hazards of boredom would be to make the data collector aware of the over-all hypothesis, but to present Ss in such an order that E does not know to which conditions each S belongs and hence what he should "expect" of each S. In one

study reported by Rosenthal (1964) such a procedure was instituted by at least one E in a conscious attempt to avoid bias. This E employed a double blind procedure simply by not looking at the code identifying the treatment condition of his Ss. No significant bias effect appeared in this particular study.

A final method of controlling E bias is derived from Rosenthal's (1964) finding that random variations in E's expectancies interferes with the appearance of the bias effect. In this particular study, Es were aware of the hypotheses and of the condition to which each S belonged. However, Ss from the several conditions appeared in random order, thus E's expectancy was constantly changing. Under these conditions, the bias effect sometimes appeared and sometimes did not appear. Thus in those experimental designs in which it may be impossible for E to be kept ignorant of the experimental condition to which S belongs, Ss from the several conditions may be tested in random order.

These are a few of the possible methods of controlling the influence of the experimenter in psychological research. However, each would have to be tested experimentally to determine its value in eliminating the bias effect.

Another problem raised by the present study concerns the finding that there were significant differences between presenting instructions "in person" and presenting them on tape. The photographs were rated higher by Ss in the taped condition, and ratings had greater variability in the taped

condition. What is different about the two modes of presenting the instructions? Although the scope of the study does not permit a complete or conclusive answer to this question, several possibilities present themselves.

One possible reason for the raising of the ratings was that the mood of the non-visual condition was different from that of the visual condition. Ss did not expect the room to be empty when they entered and it seemed to the author, observing the Ss from behind a one-way screen, that they seemed anxious, or irritated, or even amused in the non-visual condition. One can only guess at why this change in mood should cause a general raising of the ratings. However, since this difference between the two conditions was noticed, its possible role in the observed difference in ratings should not be overlooked.

Another possible source of difference may derive from the presence or absence of the E. The effect of having an experimenter in the room, especially one who is inexperienced and probably somewhat nervous, may be to depress responding or to inhibit extreme ratings, in the visual condition. Casual observation of the data revealed that the variability seemed larger in the non-visual than in the visual condition.

Finally, one can wonder what effect the slight difference in tonal quality between "real" and taped voices had on Ss' ratings. Why the more mechanical sounding taped version should lead to a raising of the ratings is not known, but the difference in tonal quality does exist, and is thus

reported as a possible source of difference between the two modes of presenting the instructions.

The difference in variability of ratings between the two conditions may have been caused by a number of factors. The increased variability in the non-visual condition may have been due to the absence of the E in the testing room, as already mentioned. Another factor which should not be overlooked is that the quality of the tapes varied somewhat for each E. Some tapes were clearer than others, due to differences in speaking ability of Es and differences in background noise. This may have contributed to the greater spread of the ratings found in the non-visual condition.

Figure 2 also illustrates the different pattern of visual-non-visual means obtained by the various Es. Three Es' scores are clustered at the top, and the other three scores are closer to the bottom of the rating scale. Thus, E differences appeared in the non-visual condition. This is the kind of variability, discussed by Barber and Calverley (1964), that could be controlled by using instructions taped by a single E for all Ss.

A final point of interest is that none of the Es in the present study obtained negative mean ratings of the photographs. All means were positive, even though Es had been biased to expect both negative and positive ratings. Negative bias simply lead to a lower positive rating than did positive bias. While there is no apparent explanation for this, it is consistent with previous research in that

most of Rosenthal's studies yielded the same results. It would be of interest for future research to employ a no-bias control group to learn whether unbiased Es obtain ratings in the photo-judging task which fall in between, above, or below ratings obtained by high and low biased Es. Perhaps the results of such an experiment could shed some understanding on the question of why negative ratings fail to occur.

effects of manipulating E biases in various ways.

The results of the study demonstrate how compelling the bias effect is. Under such restricted conditions as those created by complete elimination of visual cues, the subjects were still able to receive and interpret subtle cues transmitted in the voices of biased Es. The influence of the E is an important feature to be dealt with in psychological research.

REFERENCES

Barber, T. X., and Calverley, D. S. Effect of E's tone of voice on "hypnotic-like" suggestibility. Psychol. Rep., 1964, 15, 139-144.

Binder, A., McConnell, D., and Sjoholm, N. Verbal conditioning as a function of experimenter characteristics. J. abnorm. soc. Psychol., 1957, 55, 309-314.

Birney, R. C. The achievement motive and task performance: a replication. J. abnorm. soc. Psychol., 1958, 56, 133-135.

Brogden, W. J. The experimenter as a factor in animal conditioning. Psychol. Rep., 1962, 11, 239-242.

Cohen, Edwin. Is there examiner bias on the Wechsler-Bellevue? Proc. Okla. Acad. Sci., 1950, 31, 150-153. (abstr)

Cordaro, Lucien, and Ison, James R. Psychology of the scientist: X. Observer bias in classical conditioning of the planarian. Psychol. Rep., 1963, 13, 787-789.

Ferguson, D. C., and Buss, A. H. Operant conditioning of hostile verbs in relation to experimenter and subject characteristics. J. consult. Psychol., 1960, 24, 324-327.

Fode, Kermit L. The effect of non-visual and non-verbal interaction on experimenter bias. Unpublished master's thesis, University of North Dakota, Grand Forks, 1960.

Friedman, N., Kurland, D., and Rosenthal, R. Experimenter behavior as an unintended determinant of experimental results. J. proj. tech. pers. Assess., 1965, 29, 479-490.

Friedman, Pearl. A second experiment on interviewer bias. Sociometry, 1942, 5, 378-379.

Joel, Walther. The interpersonal equation in projective methods. Rorschach res. Exch., 1949, 13, 479-482.

Kintz, B. L., Delprato, D. J., Mettee, D. J., Persons, C. E., and Schappe, R. H. The experimenter effect. Psychol. Bull., 1965, 63, 223-232.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin Co., 1953.

Lindzey, G. A note on interviewer bias. J. appl. Psychol., 1951, 35, 182-184.

Mann, James, Menzer, D., and Standish, C. Psychotherapy of psychoses; some attitudes in the therapist influencing the course of treatment. Psychiatry, 1950, 13, 17-23.

McGuigan, F. The experimenter: neglected stimulus object. Psychol. Bull., 1963, 60, 421-428.

Orne, Martin T. On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. Amer. Psychologist, 1962, 17, 776-783.

Pfungst, O. Clever Hans (the horse of Mr. von Osten); a contribution to experimental animal and human psychology. Translated by Rahn, C. L. New York: Holt, 1911.

Robinson, D., and Rohde, S. Two experiments with an anti-Semitism poll. J. abnorm. soc. Psychol., 1946, 41, 136-144.

Rosenthal, Robert. Experimenter modeling effects as determinants of subjects' responses. J. proj. tech. pers. Assess., 1963, 27, 467-471.

Rosenthal, Robert, Experimenter outcome-orientation and the results of the psychological experiment. Psychol. Bull., 1964, 61, 405-412(a).

Rosenthal, Robert. The effect of the experimenter on the results of psychological research. In Progress in Experimental Personality Research, 1964, 1, 80-114.

Rosenthal, Robert. Covert communication and tacit understandings in the psychological experiment. Unpublished manuscript, 1966.

Rosenthal, R., Fode, K., Friedman, C., and Vikan-Kline, L. Subjects' perception of their experimenter under conditions of experimenter bias. Percept. mot. Skills, 1960, 11, 325-331.

Rosenthal, R., and Fode, K. Psychology of the scientist: V. Three experiments in experimenter bias. Psychol. Rep., 1963, 12, 491-511.

Rosenthal, R., and Fode, K. The effect of experimenter bias on the performance of the albino rat. Behav. Science, 1963, 8, 183-189(a).

Rosenthal, R., Fode, K., Vikan-Kline, L., and Persinger, G. Verbal conditioning: Mediator of experimenter expectancy effects. Psychol. Rep., 1964, 14, 71-74.

Rosenthal, R., Friedman, C., Johnson, C., Fode, K., Schill, T., White, R., and Vikan-Kline, L. Variables affecting experimenter bias in a group situation. Genet. psychol. Monogr., 1964, 70, 271-296.

Rosenthal, R., and Halas, E. S. Experimenter effect in the study of invertebrate behavior. Psychol. Rep., 1962, 11, 251-256.

Rosenthal, R., Kohn, P., Greenfield, P., and Carota, N. Data desirability, experimenter expectancy and the results of psychological research. J. pers. soc. Psychol., 1966, 3, 20-27.

Rosenthal, R., and Lawson, R. A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats. J. psychiatr. Res., 1964, 2, 61-72.

Rosenthal, R., Persinger, G., Mulry, R. C., Vikan-Kline, L., and Grothe, M. Changes in experimental hypotheses as determinants of experimental results. J. proj. tech. pers. Assess., 1964, 28, 465-469(a).

Rosenthal, R., Persinger, G., Mulry, R. C., Vikan-Kline, L., and Grothe, M. Emphasis on experimental procedure, sex of subjects, and the biasing effects of experimental hypotheses. J. proj. tech. pers. Assess., 1964, 28, 470-473(b).

Rosenthal, R., Persinger, G., and Fode, K. L. Experimenter bias, anxiety and social desirability. Percept. mot. Skills, 1962, 15, 73-74.

Rosenthal, R., Persinger, G., Vikan-Kline, L., and Fode, K. L. The effect of early data returns on data subsequently obtained by outcome-biased experimenters. Sociometry, 1963, 26, 487-498.

Rosenthal, R., Persinger, G., Vikan-Kline, L., and Fode, K. The effect of experimenter outcome-bias and subject set of awareness in verbal conditioning experiments. J. verb. learn. verb. Behav., 1963, 2, 275-283(a).

Sacks, E. L. Intelligence scores as a function of experimentally established social relationships between child and examiner. J. abnorm. soc. Psychol., 1952, 47, 354-358.

Sapolsky, A. Effect of interpersonal relationships upon verbal conditioning. J. abnorm. soc. Psychol., 1960, 60, 241-246.

Stanton, F., and Baker, K. Interviewer bias and the recall of incompletely learned materials. Sociometry, 1942, 5, 123-134.

Troffer, S. A., and Tart, C. T. Experimenter bias in hypnotist performance. Science, 1964, 145, 1330-1331.

Williams, F., and Cantril, H. The use of interviewer rapport as a method of detecting differences between 'public' and 'private' opinion. J. soc. Psychol., 1945, 22, 171-175.

Wohl, Julian. Traditional and contemporary views of psychological testing. J. proj. tech. pers. Assess., 1963, 27, 359-365.

Young, R. K. Digit span as a function of the personality of the experimenter. Amer. Psychologist, 1959, 14, 375.