EXPERIMENTER BIAS AS A FUNCTION OF TYPE AND STRUCTURE OF THE TASK

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

Morris Shames

September, 1966

ABSTRACT OF THESIS

The purpose of the present research was to test the generality of the E-bias phenomenon under conditions of a different type of task and a more structured person-perception task. Type of task, in this study, was defined by the presence of factual or emotional components in the task, whereas task-structure was defined in terms of the ambiguity of the task for S.

Two studies employing a Rosenthal replication and Numerosity Estimation condition (Study I), and a modified Rosenthal replication and modified Numerosity Estimation condition (Study II), were performed. Photographs of faces were rated by Ss in the Rosenthal replications, and the number of dots per stimulus card were estimated according to a rating scale in the Numerosity Estimation conditions. Positive and negative biases were induced in the Rosenthal replications, and overestimation and underestimation biases in the Numerosity Estimation conditions. It was concluded from the results, that the type of task and task-structure were crucial variables in the transmission of E-bias.

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Drs. J. G. Adair and M. S. Aftanas, without whose help and encouragement this research could not have been accomplished.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The observer's role in any dyadic situation has long been recognized as something more than passive. For instance, in the interests of objectivity in science, Ebbinghaus, as early as in 1913, stated:

> "It is unavoidable that, after the observation of the numerical results, suppositions should arise as to general principles which are concealed in them and which occasionally give hints as to their presence. As the investigations are carried further, these suppositions, as well as those present from the beginning, constitute a complicating factor which probably has a definite influence upon the subsequent results." (Pp. 28 - 29)

The genealogy of this phenomenon, broadly called the "experimenter (E) effect," can be traced to the ingenious investigations by Pfungst (1911) of Mr. von Osten's incredible horse. Pfungst found that the horse's knowledge was a direct function of the cues which were inadvertently transmitted by his questioners. In much the same manner, an E can inadvertently transmit unintended communications to his subject (S) in the paradigmatic psychological experiment.

Although the problem of unintentional E influence in psychological studies was recognized, it was not until 1942 that Stanton & Baker performed an experiment to investigate this phenomenon. They demonstrated that E's bias influenced Ss' responses in the recall of nonsense geometric figures. Subsequent to this study, however, only a few investigators (George, 1938; Merton, 1948; Lord, 1950; Orne, 1953; Hanley & Rokeach, 1956) have attempted to examine the E-bias phenomenon.

The clinician and interviewer, however, have long been concerned with themselves as sources of influence and bias (Mosteller, 1944; Calahan et al, 1947; Lindzey, 1951). Unfortunately, this concern has not been shared by the experimental psychologist, as Kintz, Delprato, Mettee, Persons, and Schrappe (1965) indicate:

> "It appears that experimental psychology has too long neglected the experimenter as an independent variable. By relating some of the findings of clinical and social psychologists, as well as the few experimental studies to date, it is hoped that experimental psychologists will no longer accept on faith that the experimenter is necessary but harmless." (p. 3)

Kintz et al, (1965) further point out that over a 12 year period, when E is being intensively investigated in social psychology, the E-effect continues to be ignored by experimentalists.

Thus, even by the late 1950's, only sporadic and loosely-connected work had been produced in this very critical area (Van Krevelen, 1954; Binder et al, 1958; Criswell, 1958). However, an extensive program of research in this area was initiated by Robert Rosenthal (1958, 1964, 1966), and his systematic attack upon the E-effect has cast doubt upon the generalizability and replicability of some psychological experiments.

The prototypical Rosenthal task consists of Ss rating 10 photographs of faces on a scale ranging from +10 to -10. This scale represents a "success-failure" continuum. The principal investigator instructionally induces a positive (+5) expectancy in one group of Es, and a negative (-5) expectancy in another group. Each E presents the photographs to his Ss and records their ratings from the scale. The E-bias effect is, thus, inferred from Ss' ratings which are consonant with the induced expectations. Investigators in this field have almost invariably obtained significant differences between Es with positive and negative biases.

Largely within the context of the above experimental paradigm, Rosenthal and his co-workers have attacked the problem of the E-bias effect from two general vantage points. First, those attributes, characteristics, and traits of both E and S, which lend themselves most readily to the E-expectancy effect, have been investigated. In addition, the modalities of expectancy transmission, i.e., the ways in which the bias is mediated, have been studied.

Initially, Rosenthal tended to focus primarily upon E and S attributes and their interaction, rather than the modes of expectancy transmission. As a result of several studies on the personal attributes of E, Rosenthal (1963) concludes:

> "...that experimenters will obtain different data from their subjects as a function of how the experimenter is regarded by his subjects on the attributes of (1) likeability, (2) prestige, (3) professional skill, (4) trust, and (5) sex." (p. 330)

More specifically, Rosenthal and his co-workers (Rosenthal, 1963; Rosenthal, Persinger, Mulry, Vikan-Kline, and Grothe, 1964a; 1964b) have found that the sex variable is important in E-bias studies. The results show that male Es consistently exert the biasing effect irrespective of the sex of their Ss, while female Es show the influence only with female Ss. In several instances (Rosenthal et al, 1964a, 1964b) female Ss not only failed to obtain expected data from their male Ss, but obtained data significantly opposite to that expected. Irrespective of the sex of E, male Ss appeared to be less susceptible to bias than female Ss. Friedman, Kurland, and Rosenthal (1965) also found that Es whose experimental behaviour reflected greater interpersonal involvement, task-orientation, and competence, obtained ratings significantly more in accordance with their expectancy. Thus, sex of E amd S, warmth, and professionalism, seem to be crucial variables in E-bias.

Need for social approval and anxiety have also been investigated in relation to E-bias. Using the Marlowe-Crowne Social Desirability Scale and the Taylor Manifest Anxiety Scale, it has been found (Rosenthal, Persinger, Vikan-Kline, and Mulry, 1963) that Es who were more anxious and had a lower need for social approval obtained data significantly in accord with their expectations. However, in a subsequent study (Rosenthal, Kohn, Marks, and Carota, 1963), using the same personality instruments, it was found that less anxious Es with a higher need for social approval obtained more ratings of success from their Ss. Other studies (Marcia, 1961; Mulry, 1962) have similarly shown that a high need for social approval is correlated with Ss' ratings of success, or superior performance. As yet, there are no consistent correlations between extent of bias and E's anxiety level. Similarly, no consistent relationships have been found with the extent of bias given by S and S's anxiety level or need for approval.

From the point of view of Riecken's (1962) postulation of the "deutero-problem", and Ornes (1962) treatment of the "demand characteristics" of the experimental situation, Rosenthal has studied E-bias from the vantage point of S's expectancies interacting with E's. In one such study (Rosenthal, Persinger, Vikan-Kline, and Fode, 1963a), two groups of Es were led to expect high rates of awareness and low rates of awareness from their Ss. One half of the Ss in each were given instructions designed to let them "see through" the experiment and the other half received standard instructions. It was found that the most powerful effects were obtained when the E's bias operated conjointly with S's set. Generally, however, it was found that more subtle demand characteristics had more predictable effects on the responses made by Ss. Unusually "expectant" or cued Ss provided less predictable data.

Rosenthal has also found that unusually motivated Es obtained data opposite to what they had been led to expect. When one-half of the Es were promised $2.00 for obtaining "good data", and the other half were promised $5.00 for such data, the results showed that the more moderately motivated Es obtained data which was in accord with their expectancies (Rosenthal, Fode, and Vikan-Kline, 1960). The extremely motivated Es, however, showed a tendency to obtain data opposite to what they had been led to expect. In another study (Rosenthal, Friedman, Johnson, Fode, Schill, White, and Vikan-Kline, 1964) Es were promised $1.00 and the $1.00 of an unknown partner if they would obtain results better than that of their "partner". However, their $1.00 would be forfeited to their "partner" if he obtained better results. Once again, extremely-motivated Es seemed to obtain data opposite to what they had been led to expect.

In scrutinizing the E-expectancy phenomenon, Rosenthal has also directed his attention to the effects of other more situational or task variables. Rosenthal, Persinger, Vikan-Kline and Fode (1963b) employed accomplices who provided E with expected or unexpected data. The results indicated that these early data returns significantly determined the kind of data E obtained from subsequent Ss. Rosenthal, Kohn, Marks, and Carota (1963) confirmed the above findings in a second study. In another study, however, Rosenthal and his colleagues (Rosenthal, Persinger, Mulry, Vikan-Kline and Grothe, 1964a) have shown that significant expectancy effects are obtained even when Es had their expectancies reversed during the course of their experiments. Rosenthal, Kohn, Greenfield, and Carota (1966) have employed a verbal conditioning study to demonstrate that the expectancy effect is more consistent when the induced desirability of the data and induced bias are consonant. Those Es who wanted but did not expect, and those who had expected but did not want verbal conditioning, did not obtain biased data.

Rosenthal's research has not, however, ignored investigations into the modalities of expectancy transmission. Fode (1960) found that without restriction of visual cues, verbal cues alone were sufficient to mediate E-bias. Thus, much of bias transmission must occur in the instruction phase of the task. Recently, Rosenthal and his colleagues have been analysing sound-motion pictures of the person-perception experiment to explore the modes of E bias transmission. While they have yet to find the specific cues that mediate the "Clever Hans" phenomenon to human subjects (Rosenthal, 1966), analysis of the pictures alone, sound tracks alone, and combined pictures and sound tracks, have led to some interesting, yet not entirely consistent data. While both visual and auditory cues play a role in bias transmission, their relative significance is not constant throughout the E-S interaction. At this point, Rosenthal (1966) feels that "the most compelling and most general conclusion is that human beings can engage in highly effective and influential unprogrammed and unintended communication with one another." (p. 35) The nature of this communication may vary from situation to situation.

Although the major portion of this extensive experimental program has been performed with human Ss, studies investigating expectancy effects and relevant variables have also been performed with animals. In several animal studies (Rosenthal & Halas, 1962; Rosenthal & Lawson, 1963; Rosenthal & Fode, 1963), Es who believed they were running maze-bright

(Berkeley-bred) rats obtained superior performance from their animals than Es who believed they were running maze-dull rats. Since all the rats came from the same colony, Es with different expectancies still obtained the bias effect. For the most part, Rosenthal suggests differential handling and gentling as possible explanations for the obtained results.

In addition to the animal studies, Rosenthal has attempted to explore the E-bias phemonenon in verbal conditioning studies. In several such studies, Rosenthal and his co-workers (Fode, Rosenthal, Vikan-Kline, and Persinger, 1961; Rosenthal, Persinger, Vikan-Kline, and Fode, 1963a; Rosenthal, Kohn, Greenfield, and Carota, 1966) studied the possibility that operant conditioning could account for the development of a set, by S, which was reinforced according to E's will. From the findings of these studies and others (Rosenthal, Fode, Vikan-Kline, and Persinger, 1964), these authors suggest that E, during the brief pre-data-gathering interaction, "greets, seats, 'sets', and instructs S", and this "set" can be perpetuated by E's subtle reinforcement. It is to be noted, however, that the tasks used in these studies were either variations of the usual Rosenthal task, or a similarly ambiguous Taffel task (Rosenthal, 1966). Therefore, with the exception of these few variations, research has employed the usual person-perception task.

In view of Rosenthal's research, the E-expectancy effect must be acknowledged as an undeniable phenomenon which pervades much psychological experimentation; but it remains to be answered if this influence is peculiar to the specific experimental task, or, if it is so robust a phenomenon as to contaminate all psychological experimentation. In answer to this query, Rosenthal (1966) alludes to the possibility of the task-specifity of the E-effects:

> "What effects just such behavior on the part of the experimenter will have on the subject's responses, depends no doubt on the particular experiment being conducted and, very likely, on various characteristics of the subject as well." (p. 15)

He further edifies this allusion by averring:

> "The particular patterns of covert communication which have been described as relevant to the experimenter's communication of his expectancy to his subjects are no doubt specific to the type of experiment being performed. We are in no position to speak for the generality of any of these findings across different experiments, much less for their generality in the other 'real world', that one outside the laboratory..." (p.35)

Yet, in spite of these allusions, he claims, ealier in the same article:

> "One purpose of this paper is to illustrate the fact that unintended covert communications are the norm in psychological experiments." (p.3)

Thus, it would appear that Rosenthal expects the generality of the E-bias phenomenon, but admits that it has not been demonstrated.

In regards to the generality of E-bias phenomenon, it should be noted that not all studies have obtained unequivocally significant differences between Es with differing expectations. Of these, a few have obtained clearly contradictory results (Pflugrath, 1962; Wartenberg-Ekren, 1962).

In her perspicous study, Wartenberg-Ekren (1962) had eight Es administer the Block Design of the WAIS to a total of 32 Ss. Prior to this procedure, Es were led to expect that half their Ss would perform well and half would perform poorly. According to Wartenberg-Ekren:

> "The purpose of the present study is to investigate whether experimenter expactancy influences the performance on a task which does not appear as self-revealing or personality revealing to subjects, e.g., one involving perceptual organization and reasoning. Such a task should sensitize subjects less toward seeking additional cues from the experimenter relative to experimenter expectancy." (p. 4)

She reasoned, therefore, that since it requires more attention and absorption on the part of S, a reasoning task would leave less attention (than Rosenthal's task) for E behaviour. From the results of this study, she concluded:

> "This experiment demonstrated that the experimenter expectancy in regard to a certain subject performance need not have an effect on subject behaviour, and that the generality of the experimenter outcome-bias phenomenon as alleged by Rosenthal needs further exploration." (p. 16)

In another study, Pflugrath (1962) had three sets of three examiners administer the Taylor Manifest Anxiety Scale to a total of 142 Ss. The expectations induced in one group of Es was that their Ss were highly anxious, while the second experimental group was told to expect little or no anxiety in their Ss. A third, control group was given no instructions beyond those required to administer the test. The hypothesis proposed was that the E-expectancy effect would be evinced under this group testing situation. However, the analysis failed to yield significant differences among the groups or among the examiners.

In his explanation of these negative instances of the E-expectancy effect, Rosenthal (1964) proposes:

> "The simplest conclusions to draw from the results of the experiments discussed in this section (Pflugrath, 1962; Wartenberg-Ekren, 1962) might be that the experimenter's expectancies do not affect subjects' intelligence test or anxiety scale performance. Indeed there may be a large array of tasks which will prove relatively resistant to the effects of experimenter expectancies. What tasks are more or less resistant is a question deserving of further research." (p. 104)

In response to these contradictory results, however, Rosenthal (1964) reasons that one can question the efficacy of the induction of E-expectancy in the Wartenberg-Ekren study, since the different expectations were only inferred from Ss' ratings of their respective Es. Normally, this induction of bias can be inferred from both the dependent variable and Ss' ratings of E, but since it was hypothesized that no difference between Es would obtain, the induction of bias could only be inferred from Ss' ratings of E. In this regard, it should be noted that until very recently (Rosenthal, 1966), when sound-motion picture analyses were performed, Rosenthal inferred the experimental induction of E-bias primarily from Ss' ratings of E, because the dependent variable could have been, possibly, the result of variables other than the instructionally-induced bias. Also, Rosenthal claimed that since Wartenberg-Ekren's Es were graduate students who were cognizant of the importance of constancy in psychological experiments, they probably had instituted "double blind" procedures to eliminate the possibility of expectancy effects or any other such contaminating variable. Wartenberg-Ekren cites some evidence, in the case of one E, to support this interpretation.

With regard to the Pflugrath study, Rosenthal (1964) asserted that the examiners in this experiment were graduate counsellors and, thus, therapeutically-oriented. It could be expected that they would bring all their skills to bear on the reduction of their Ss' anxiety, thus eliminating the possibility of an E-expectancy effect in this experiment.

Notwithstanding these valid explanations of the above failures, it is of speculative interest to note that a common thread runs throughout both studies. In both of these studies, the experimental task employed has been decidedly different than that of Rosenthal, and this may, perhaps, be the sine qua non for further experimental exploration. The two studies still indicate, however, that Ss influence E who, in turn, influence Ss. In spite of this complex feedback loop of influence, the efficacy of this influence remains contingent upon whether the experimental format does or does not lend itself to the transmission of bias.

Thus, the <u>type</u> of task, or the properties inherent therein, may dictate whether or not bias may be efficaciously transmitted so as to confound the dependent variable. In this connection, Ferber & Wales (1952) reported a method of detecting interviewer bias in survey data. Twelve experienced interviewers were given a questionnaire to complete, unknowing that they would administer this same questionnaire in a survey several weeks later. Similarity of the interviewers' responses to those of the respondents was the measure of Answer Bias. The results of this study demonstrated that Answer Bias was transmitted to respondents, but predominantly on certain <u>types</u> of questions. Thus, the authors conclude:

> "These results tend to confirm previous findings that interviewer bias is more likely to crop up on attitudinal questions than on questions of fact." (p. 116)

<u>Rationale and Statement of the Problem</u>

Although Rosenthal's photographs were "standardized," there does not exist in his task the consistent agreement as in, for instance, judging the number of lines per page according to a universally agreed-upon numerical scale. Since it entails the evaluative judgment of success and failure on quite an arbitrary scale, the typical Rosenthal task can be considered "attitude-oriented" rather than "fact-centered". <u>S</u>'s attitudes and emotions thus may be contaminating variables in such a task, since <u>S</u> makes an attitudinal judgment and is being influenced by <u>E</u>'s, and his own attitudes and emotions. Greenspoon (1955) has demonstrated that a simple "mmm-hmmm" is sufficient reinforcement for the elicitation and perpetuation of a response. The Rosenthal task appears equally conducive to the facile transmission of cues by <u>E</u> since he can paralinguistically and kinesically transmit "attitude cues", e.g. grimace, intonation, inflexion, and so on. Thus, E can transmit these cues with greater facility in the judgment of faces, than would be possible if the task were purely "fact-centered". An important variable upon which the first of the present studies is based, therefore, is the type of task, i.e. the properties inherent in the task.

In view of the Ferber & Wales (1952) and Wartenberg-Ekren (1962) studies, it is suggested that a different type of task, containing less of an attitude component than that of Rosenthal, should prove less amenable to the paralinguistic and kinesic transmission of cues. Therefore, it is proposed that a numerosity estimation task, with a universally-agreed-upon rating continuum, is more "fact-centered" and should be less coloured by attitude and emotion. Also, the numerosity of the figures used in this study cannot be accurately estimated due to the large numbers to be estimated in so short a period of time (5 seconds). Because of this the task is quite similar to Rosenthal's insofar as they both lack clear structure for S. Although Pflugrath's (1962) Taylor Manifest Anxiety Scale and Wartenberg-Ekren's (1962) Block Design tasks were completely dissimilar to Rosenthal's, the numerosity estimation task employed in the present study is quite similar to that of Rosenthal, with the exception that it is of a different type, i.e. "fact-centered". However, a factual task of the numerosity sort, should not evince the E-effect commonly found in Rosenthal's research. Therefore, it is hypothesized that, in contrast to a replication of Rosenthal's positively- and negatively-biased groups, there will be no significant difference between high and low biased groups in a Numerosity Estimation condition.

Another important variable - and that upon which the second study is based - is the structure of the task. So that the distinction between type of task and task-structure may be clarified, it should be noted that the same type of task may be structured in various ways.

It should be noted that Wartenberg-Ekren (1962) claimed that a reasoning task would require more absorption on the part of S and, thus, deny E the influence of his expectations. Structure of the task, in terms of how it directs S's attention and provides S with a set, may, conceivably, be a crucial variable. Thus, in the typical Rosenthal design, the task is quite structureless, and S may acquire a set from E. This set may persist throughout the judgment of the remaining stimuli and, thus, produce the E-bias effect.

Certain aspects of the Rosenthal task help to create this lack of structure. Since success implies a variety of different things to different people, there is a somewhat poorly-delineated criterion of success and failure. Also, with regard to Rosenthal's Empathy Test Rating Scale, it should be noted that, in the opinion of psychophysics (Miller, 1956; Garner, 1962), a rating scale exceeding 9 or 10 discrete intervals loses its efficacy as a judgmental guide for S. Thus, a 20-point scale lends itself to ambiguity, or lack of structure.

A second study, therefore, was run to investigate E-bias under conditions in which task-structure is increased. A 10-point rating scale was employed, and it was reasoned that this should be a more effective judgment guide than Rosenthal's 20-point scale. Also, in the Numerosity Estimation condition, a more estimable amount of dots were used, i.e., 80 in this study as compared to 200 in Study I. Both of these variables should yield a greater degree of structure of the task for S and consequently minimize the influence of E's bias. Since both the modified

Rosenthal replicate condition and the Numerosity Estimation condition are assumed to be more structured, it is hypothesized that there will be no significant differences between the Es with positive (high) and negative (low) biases in either task.

CHAPTER II

METHOD

Subjects (Ss)

One hundred and sixty-seven females, from the Introductory Psychology, Psychology of Personality and Criminology courses at the University of Manitoba, served as Ss. Ss had previously participated in an E-bias study, employing a different set of stimulus pictures.

Experimenters (Es)

Forty-seven male volunteers, from the Psychology of Personality course at the University of Manitoba, served as Es. Prior to running the experiment, it was stipulated that each E must run a minimum of 3 Ss. Since 7 Es ran only 2 Ss or less, each, their data was discarded and the data of 40 Es and 153 Ss was used for the statistical analysis.

Materials

The Rosenthal replicate conditions employed 10 photographs of men's faces from weekly newsmagazines which were photographed and mounted on 3 x 5 inch white cards for presentation to Ss. These photographs were similar in form to those used by Rosenthal (Rosenthal & Fode, 1963a) and were standardized by a previous study which corroborated Rosenthal's bias (Epstein, 1966).

Two sets of 10 cards each were used in the Numerosity Estimation conditions. One set, employed in Study I, consisted of 9 1/4 x 12 1/4 inch white cardboards upon which were mounted 200 dots each, in irregular patterns. In the second set, used in Study II, there were 80 dots mounted on each 9 1/4 x 12 1/4 inch cardboard. As indicated in the photographs of the sample cards, in Figures 1 and 2, the dots in both sets were arranged in varying densities so as to yield the impression of varying numerosity between stimulus cards.

Design

Es were randomly assigned into 8 groups. Since two studies were being conducted concurrently, four groups of Es comprised each study. There were 5 Es per cell, and an approximate random assignment of Ss to experimental treatment groups was effected in a manner similar to that of Fode (1960). E's free half-hour was partitioned into 5 equal time segments, and Ss signed up for whichever time interval they preferred.

Each study consisted of a 2 x 2 factorial design with 2 levels of bias in each of two tasks. In Study I, the Es in Groups I and II (N = 10) were identical to the basic paradigm of the Rosenthal studies in E-bias. The 10 Es in the Numerosity Estimation condition, Groups III and IV were run in a manner similar to that of Rosenthal's studies, with differential biases being induced according to the numerosity judgment task.

The same experimental format was employed in Study II, with the 10 Es, Groups I and II in the modified Rosenthal replicate condition receiving differential biases. The two Groups (III and IV) in the Numerosity Estimation condition also had differential biases induced.

Procedure

All Es were greeted by the principal investigator and were read a set of instructions describing their task to them (Appendix A). The experimental procedure, which was typed out for E, was essentially identical in all conditions (Fode, 1960). In all conditions, each E read a

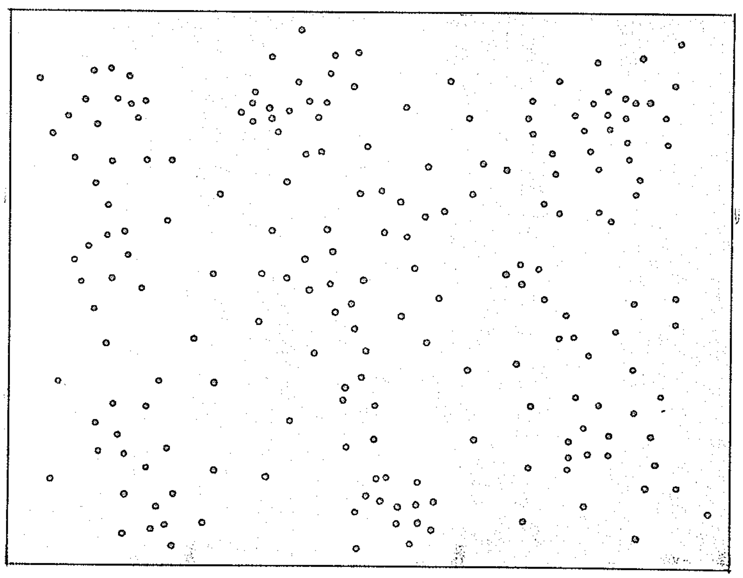

Fig. 1. Sample stimulus card used in Numerosity estimation condition in Study I.

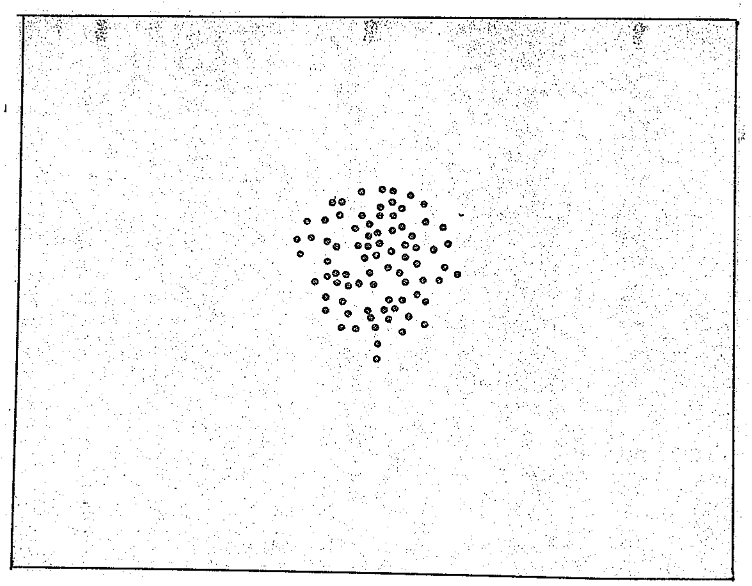

Fig. 2. Sample stimulus card used in Numerosity estimation condition in Study II.

set of standard instructions to S before presenting each photograph or stimulus card individually. E recorded Ss' verbally responded ratings. E's task was structured, basically, as a laboratory exercise to see if they could replicate "well-established" findings. The rudimentary procedure followed by each E was: (1) Greet each S in the hallway and ask S to be seated at the appropriate side of the table. (2) Obtain the factual information from S, as required by the data recording sheet. (3) Read the instructions to S. (4) Place the appropriate set of photographs, or stimulus cards, on the table in front of S. (5) Record S's rating, for each photograph (card), on the data recording sheet. (6) Send S across the hall to the principal investigator to complete a Personal Opinion Inventory (i.e., in reality, an experimenter rating sheet).

Study I

In the Rosenthal replicate conditions, Groups I and II, the usual "Empathy Test Rating Scale" was employed (Fode, 1960). This scale runs from -1 to -10 to indicate degrees of failure and from +1 to +10 to reflect degrees of success. The principal investigator read the instructions to the Es which described their role in the experiment and induced the different expectations according to this scale. A positive bias was instructionally induced in that Es were told that previous research has found that Ss averaged a +5 rating, whereas a negative bias was induced by the expectation of an average -5 rating. These instructions are almost identical to those used by Rosenthal (Rosenthal, Persinger, Vikan-Kline, & Fode, 1963b) and are in Appendix A.

Each E followed the same experimental procedure and began by reading a set of instructions to S. The instructions are similar to those employed by Rosenthal (Rosenthal, Persinger, Vikan-Kline, & Fode, 1963b) and may also be found in Appendix A. The only difference in these instructions from Rosenthal's was E querying S as to whether she had previously participated in a similar experiment. Since the Ss employed in this study had previously participated in a similar experiment, this phrase was inserted for the specific purpose of making E cognizant of this fact and to assure S that this was permissible.

A scale slightly similar to Rosenthal's 20-point scale was used in the Numerosity Estimation condition. This scale, as represented in Figure 3, was typed on an 8 x 5 inch white card and placed on the table in front of S. The principal investigator read the following instructions, to each E, which were designed to describe E's role in the experiment and to induce this expectation of overestimation or underestimation.

"You have been asked to participate in a research project developing a human engineering test. One reason for your participation in this project is 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. The main reason, however, is that there is the problem in psychological research of different examiners getting somewhat different data on the same tests as a function of individual differences in both subjects and experimenters. Therefore, to standardize the tests, it is better methodological procedure to use groups of experimenters.

You will be asked to run a series of subjects and obtain from each ratings of stimulus cards. After you have run each subject, you are to send her across the hall to me where she will fill in a Personal Opinion Inventory. This inventory is a rather gross measure of individual differences. 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 you will be using have averaged a 190 (210) rating. Therefore, the subjects you will be using should also average about a 190 (210) rating.

Just read the instructions to the subjects. Say nothing else to

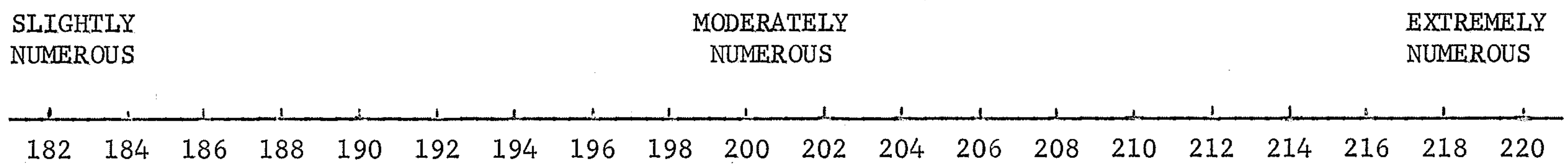

Fig. 3 Numerosity estimation rating scale used in Study I.

them except hello and goodbye. If for any reason you should say anything to a subject other than that which is written in your instructions, please write down the exact words you used and the situation which forced you to use them.

Your subjects will be coming in at about ________ and there is a note outside asking them to wait in the corridor until you get them. When you have finished with them, send them across the hall to me in Room 326.

"GOOD LUCK!"

The Es in the Numerosity Estimation condition, Groups III and IV, were instructionally induced to expect their Ss to overestimate, i.e., give an average 210 rating, or to underestimate, i.e., give an average 190 rating, on this scale.

Upon completion of this phase, each E -- prior to S's expressed ratings -- read the following instructions to each S:

"I am going to read you some instructions. I am not permitted to say anthing which is not in the instructions nor can I answer any questions about this experiment. OK?

We are in the process of developing a human engineering test. This test is designed to show those patterns which lend themselves most readily to accurate judgment of numerosity. I will show you a series of cards. For each one I want you to judge the number of discrete dots on the card. To assist you in your judgments you are to use this rating scale. As you can see the scale runs from 182 to 220. Try to match your estimation of the number of dots on each card with the appropriate number on the scale. Thus, a rating of 182 means that you judge the card to be slightly numerous. A rating of 220 means that you judge the card to be extremely numerous.

You are to rate each card as accurately as you can. Just tell me the rating you assign to each card. All ready? Here is the first card. (No further explanation may be given although all or part of the instructions may be repeated.)"

Study II

The experimental procedure employed in this study was identical to that in the first study with the exception that different rating scales were employed and the instructions to both E and S were slightly altered to accommodate the rating scales (Appendix A).

A modified Empathy Test Rating Scale, as represented in Figure 4, was employed. This was a 10-point rating scale ranging from -5 to +5. The instructions to E in the Modified Rosenthal replicate condition were identical to those in Study I, with the exception that Es in the positive bias condition were led to expect that their Ss would average a +3 rating, while those Es in the negative bias condition were led to expect their Ss to average a -3 rating.

A somewhat similar 10-point rating scale was employed in the modified Numerosity Estimation condition. As indicated in Figure 5, this scale ran from 60 to 105. Again, the instructions to E, in this condition, were identical to those in Study I, with the exception that Es who were led to expect that their Ss would overestimate, were instructed to expect an average 95 rating; while those Es who expected their Ss to underestimate, were led to expect a 70 rating.

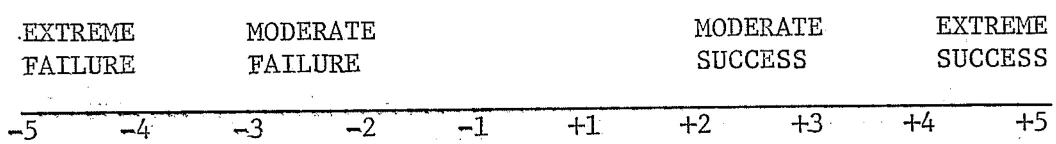

Fig. 4 Modified empathy test rating scale used in Study II.

SLIGHTLY NUMEROUS				MODERATELY NUMEROUS			EXTREMELY NUNEROUS	
60	65	70	75	80	85	90	100	105

Fig. 5. Modified Numerosity Estimation rating scale used in Study II.

CHAPTER III

RESULTS

In both studies, the dependent variable was Ss' responses on the rating scales employed. In Study I, the Rosenthal replicate condition employed a dichotomized scale running from -10 to +10, while in the Numerosity Estimation condition, an absolute scale of 20 points was similarly used. This scale ran from 182 to 220 in intervals of 2. In Study II, the modified Rosenthal replicate condition again employed an algebraic scale running from -5 to +5, while the Numerosity Estimation condition employed a 10-point absolute scale running from 60 to 105 in intervals of 5.

Since obvious differences in the values of the dependent variable precluded comparisons across conditions, a numerical transformation to a uniform scale was planned. The intention was to transform the data from both conditions in Study I to a uniform scale running from 1 to 20. A similar scale running from 1 to 10 for both conditions in Study II was also planned. Thus, the data from the Rosenthal replicate and Numerosity Estimation conditions would be transformed to comparable scale units. An analysis of variance could then be performed and a comparison across conditions, and between groups within each condition, could be made in each of the two studies.

A transformation of this kind was performed and yielded data, as Figure 6 indicates, which were inconsistent with the algebraic data in the Rosenthal replicate conditions. Subsequent t-tests were calculated for both the algebraic and the transformed data, and, the resultant values were not consistent. Thus, the intended statistical procedure could not be employed because of the nature of Rosenthal's scale. Although the interval between any two points on the Empathy Test Rating Scale is assumed to be only one interval, the interval between points -1 and +1 is larger because there is an implicit 0 between these two points. Rosenthal makes the implicit assumption of this 0 point as the scale is employed by his Ss for rating purposes. Similarly, he makes such an implicit assumption in his statistical treatment of the data. However, the possibility of obtaining statistically significant results is increased when an algebraic scale without an actual 0 point is employed since, as the transformation revealed, the error variance is spuriously reduced. Since Rosenthal has employed, both in the experimental session and in the statistical analysis, an algebraic scale, without having made the 0 point explicit in either case, this assumption could not be made in the present study. Consistency would demand that the same scale units which were used by Rosenthal, be used for the transformation. A 21-point scale, with an explicit 0 could not be used since in the Numerosity Estimation condition a scale ranging from 182 to 220, in intervals of 2 was employed. Statistical consistency could not be effected because it would entail the addition of a point to this scale when in fact this is not justified because it was not employed during the experimental session. A transformation on a uniform 20-point scale for all conditions was not possible since this led to an unfaithful transformation, and the equivalence of the algebraic data with the transformed data could not be assumed.

Since the experimental hypotheses suggest no difference in the criterion variable between high and low bias groups in the Numerosity Estimation condition, and a significant difference between positive and

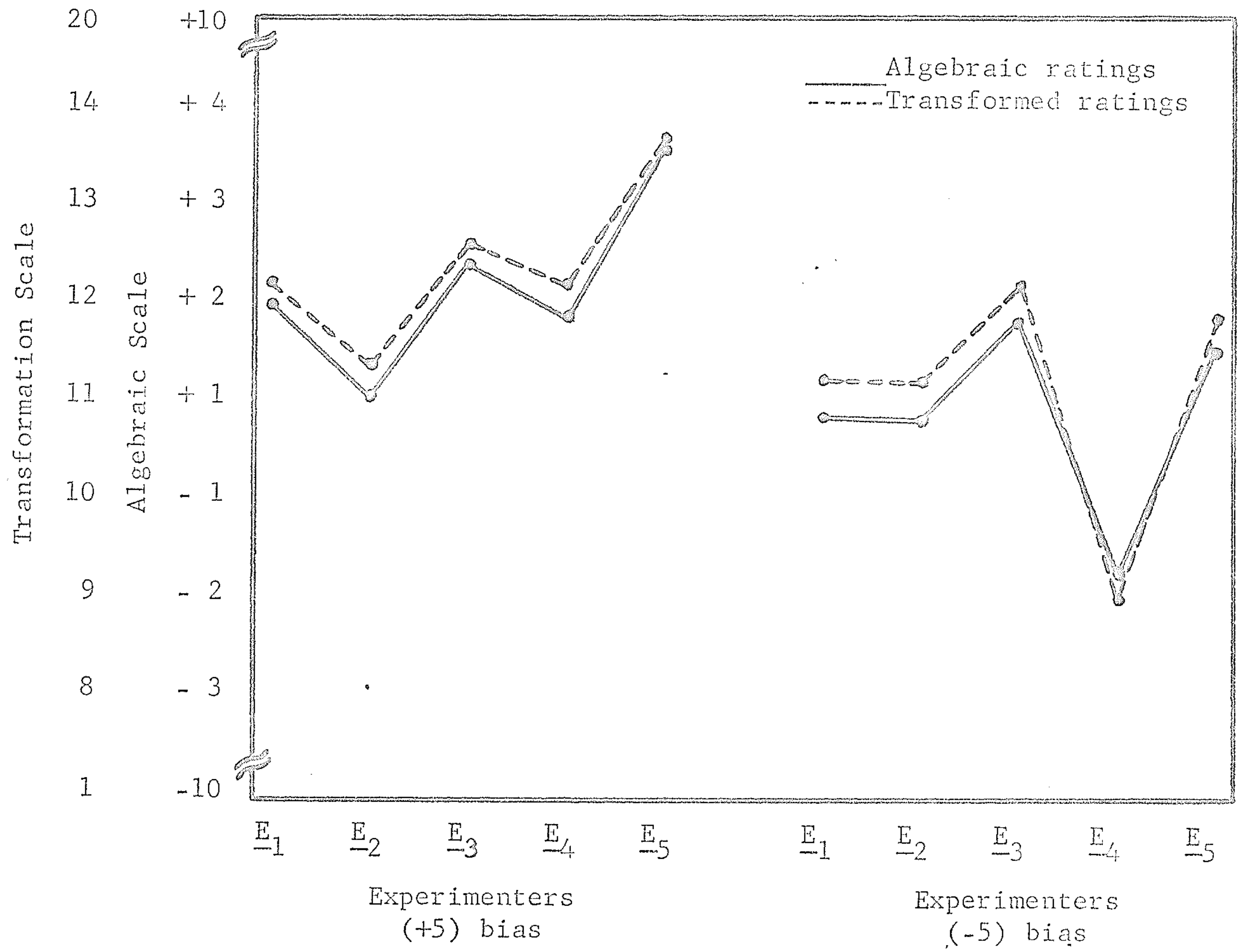

Fig. 6. A comparison of the algebraic data with the transformed data in the Rosenthal replicate condition in Study I.

TABLE 1

S_s' MEAN RATINGS IN STUDY I

		Rosenthal Replication					Numerosity Estimation				
		E_1	E_2	E_3	E_4	E_5	E_1	E_2	E_3	E_4	E_5
+5 Bias	S_1	+4.6	+0.9	+4.4	+1.5	+4.2	194.0	206.4	198.2	195.6	193.6
	S_2	+1.2	+0.3	+5.1	+1.6	+4.7	196.8	201.0	196.6	212.6	185.2
	S_3	+1.8	+1.0	+1.0	+3.4	+3.5	194.4	200.6	196.8	197.0	195.8
	S_4	+0.0	+2.1	+2.6	+1.3	+2.1	183.2				188.2
	S_5	+2.7		-0.1			194.2				
	S_6			+1.7							
Mean		+2.06	+1.075	+2.45	+1.95	+3.625	192.52	202.66	197.2	201.73	190.7
-5 Bias	S_1	+2.4	+3.7	+1.3	-4.8	+0.7	198.4	202.4	189.8	199.8	205.0
	S_2	-1.7	+1.6	+2.4	-0.3	-0.2	199.4	183.8	196.6	199.4	194.8
	S_3	+2.1	-1.8	+3.1	+ .2	+3.1	198.8	195.6	186.6	194.4	199.0
	S_4		+ .1	+1.1		+2.8		200.6	201.4	187.6	187.8
	S_5										195.8
	S_6										
Mean		+0.93	+0.90	+1.975	-1.633	+1.60	198.87	195.6	193.6	195.3	196.48

negative bias groups in the Rosenthal replicate condition, it was decided that t-tests would be run for each pair of groups in Study I. As Table 1 indicates, the raw score data were used and the mean of each S's ratings (10 ratings per S) was treated as the unit of analysis. Thus, each experimental treatment group was comprised of from 18 to 23 scores, a mean for each S within each group.

The results of the t-tests for independent means are summarized in Table 2. A comparison of Groups I and II, in Study I, yielded a "t" of 2.3548, which was significant at the .05 level with 40 degrees of freedom. The t-test employed was one-tailed since the research hypothesis suggested no difference between groups while the alternate hypothesis, based on Rosenthal's previous findings, suggest a difference in the "expected" direction. The comparison between Groups III and IV yielded a "t" of .1326, which did not reach significance at the .05 level with 37 degrees of freedom. This t-test, was also one-tailed. Thus, the results were consistent with the hypothesis proposed in Study I.

TABLE 2

RESULTS OF t-TESTS FOR INDEPENDENT GROUPS

Comparison	Study I		Study II	
	t (one-tailed)	p	t (two-tailed)	p
Group I vs. Group II	2.3548	.05	.7516	N.S.
Group III vs Group IV	.1326	N.S.	1.8169	N.S.

The same analysis was employed in Study II since the experimental hypothesis proposed no significant difference positively- and negatively-

TABLE 3

<u>Ss</u>' MEAN RATINGS IN STUDY II

		Modified Rosenthal Replicate					Modified Numerosity Estimation				
		E_1	E_2	E_3	E_4	E_5	E_1	E_2	E_3	E_4	E_5
+3 Bias	S_1	+1.3	+1.7	-0.2	+2.5	+1.4	77.5	69.0	73.0	76.0	76.5
	S_2	-0.1	+2.8	-1.2	-0.6	+1.7	78.0	61.0	63.5	74.0	83.5
	S_3	+2.3	+0.5	+1.0	+1.1	+2.2	73.5	65.5	66.5	69.0	79.5
	S_4	-0.7		+0.7	+1.4		71.0			73.5	
	S_5										
Mean		+0.70	+1.66	+ .075	+1.10	+1.76	75.00	65.16	67.66	73.125	79.83
-3 Bias	S_1	+1.6	+0.3	-0.4	-1.0	-0.9	78.0	72.0	83.0	75.5	76.0
	S_2	+1.7	-1.7	+1.6	-0.9	+3.7	67.5	68.0	65.5	77.5	86.5
	S_3	+0.7	-1.0	+1.0	-2.1	+1.8	88.0	75.0	80.0	81.5	72.5
	S_4	+3.1		+3.0	+2.1	+0.9	68.0	68.0	76.5	89.5	
	S_5	-0.9							78.0		
Mean		+1.24	-0.80	+1.30	-0.475	+1.375	75.375	70.75	76.60	81.00	78.33

biased Es in the Rosenthal replicate and Numerosity Estimation conditions. However, since no experimental precedents have been established for either of these tasks, the alternate hypothesis did not entail a directional difference, but rather any difference whatever. Thus, unlike the analysis employed in Study I, the t-tests calculated were two-tailed. As Table 3 indicates, each experimental treatment group was comprised of from 17 to 20 scores, a mean for each S within each group.

As represented in Table 2, a comparison of Groups I and II yielded a "t" of .7516, which was not significant at the .05 level with 36 degrees of freedom. A comparison of Groups III and IV yielded a "t" of 1.8169, which was also non-significant at the .05 level with 35 degrees of freedom. Thus, these results tend to confirm the experimental hypothesis.

CHAPTER IV

DISCUSSION

The significant difference between the means of Group I, (+5) bias replicate, and Group II, (-5) bias replicate, in Study I, substantiates Rosenthal's earlier finding (Rosenthal & Fode, 1959) that E-bias can be experimentally induced. It is to be noted, however, that the Ss used in this study participated previously in a similar experiment. These Ss were used primarily because they were the only female S source available. In this regard, it should be noted that the E-bias effect was obtained in spite of Ss' previous participation in a Rosenthal type experiment. Therefore, prior exposure to a different set of pictures in a photograph-rating task does not seem to interfere with the E-bias effect.

The "t" obtained between the means of Group III, (210) overestimation bias, and Group IV, (190) underestimation bias, however, failed to reach significance at the .05 level and, consequently, the experimental hypothesis was accepted. Thus, as predicted, no significant difference between high-bias and low-bias Es in the Numerosity Estimation task was obtained.

In view of these findings, in Study I, it may be suggested that type of task is a crucial variable in E-bias studies. Specifically, these results lend support to the idea expressed by Ferber & Wales (1952) regarding differential influence on questions of fact and on attitudinal questions. It appears, therefore, that a numerosity estimation task is more "fact-centered" than is Rosenthal's task, and that E-bias, as expressed in the dependent variable is in part a function of the type of task employed. These results are also consistent with Wartenberg-Ekren's (1962) allegation that tasks such as Rosenthal's i.e., person-perception tasks, which are "self-revaling or personality-revealing to subjects", are more amenable to Ebias. More factually-oriented tasks, such as a reasoning task or the like, allow for less E-influence.

A similar analysis was made in Study II, comparing the means of Group I, (+3) modified bias replicate, and Group II (-3) modified bias replicate. This analysis yielded a "t" which failed to reach significance at the .05 level. Thus, the experimental hypothesis of no difference between groups in a modified, structured Rosenthal replication was supported. The comparison of Group III, modified (95) overestimation bias, with Group IV, modified (70) underestimation bias, also yielded a "t" which failed to reach significance.

These results lend support to the notion that task-structure is a crucial variable in E-bias studies. It is assumed that since the scale was reduced from 20 points to 10 points, it was a more effective guide for Ss' responses. Thus, S relied more on the scale and less upon E during the judgmental process. It should be noted that this finding is consonant with psychophysical findings (Miller, 1956; Garner, 1962). Thus, the more ambiguous the experimental task, the more readily E-bias is transmitted.

This hypothesis was also supported by the comparison between groups in the modified Numerosity Estimation condition. The task was more structured, since 80 dots were employed in this condition as compared to 200 dots in Study I. Since this is a more estimable amount of dots, the task was considered more structured than in Study I. In conjunction with a more effective 10-point rating scale, these more structured conditions appear to eliminate task ambiguity which facilitates the transmission of E-bias.

The degree of bias induced in Study II, might be offered as an alternate explanation to the results obtained. It may be, in order to obtain the effect, the expectancies induced must be of a magnitude of +5 to -5. By instructing Es to expect almost a (+5) (-5) bias rather than a (+3) (-3) bias, perhaps a significant difference between groups might be obtained in spite of having used a more reduced scale. In the same way, perhaps, more extreme biases might be induced in the modified Numerosity Estimation condition although an E-bias was not anticipated because of the type of task employed.

It is to be noted that in both the modified Rosenthal replicate and Numerosity Estimation conditions, there was a non-significant trend toward data which were in the opposite direction to that which was expected. The lesser magnitude of the difference between induced biases may have been a factor contributing to this trend. More importantly, however, in a previous study (Rosenthal, Persinger, Mulry, Vikan-Kline, and Grothe, 1964) male Ss appeared to give data opposite to those expected by female Ss as an assertion of their "masculine independence." In the present study, a similar phenomenon may have occurred since many of the female Ss were older than the male Es and were, thus asserting a conventional status-hierarchy.

It is equally possible that this non-significant trend was due to the "blatant" demand characteristics of the experimental situation. Since Rosenthal, Persinger, Vikan-Kline, and Fode (1963) found that more subtle demand characteristics had more predictable effects on the responses made by Ss, the data in the present study could possibly have been influenced by the blatent demand characteristics. All Ss had participated previously in a similar experiment and many were cognizant of the experimental hypothesis. Data opposite to that expected has been obtained when the demand characteristics are blatent (White, 1962).

In summary, it might be proposed from the results of Study I, that the type of task employed in studies of E-bias is a crucial variable in determining the influence of E's expectations. Thus, whether a task is "fact-oriented" or whether it is of the more ambiguous person-perception variety, may dictate whether E's expectations will be contributory to the dependent variable. However, it should also be noted that implicit in this discussion of "fact-oriented" tasks, is the inextricably bound property of increased structure. Thus, although the Numerosity Estimation condition was considered structureless, like Rosenthal's task, because accurate judgment of the actual number of dots was a prohibitive task, it should be noted that this condition might still be considered more structured than Rosenthal's task because of a universally-agreed-upon rating scale, and because of its relative resistance to "attitudinal" confounding.

Study II, however, was specifically designed to investigate the task-structure variable. The results demonstrate that this is also a crucial variable in E-bias studies since the use of a 10-point rating scale, instead of a 20-point scale, effectively eliminated the customary bias obtained. Since no bias was evinced in the more ambiguous Numerosity Estimation condition, in Study I, no difference between differentially -biased groups in a more structured Numerosity Estimation condition was also obtained. Thus, both studies confirm the experimental hypotheses that the type of task employed and task-structure are significant variables in determining whether E's expectations will be transmitted to S.

Needs for further research

In view of the present findings, it is suggested that research involving the study of bias-resistant and bias-amenable tasks would be of immense value. Working from within this framework, more light could be shed upon the nature of E-expectancy and its mode of transmission. Thus, by using a variety of tasks other than one which is consistently successful in demonstrating the E-bias effect, more can be learned about E and S attributes which do or do not lend themselves to demonstrating this effect, and whether the generality of E-bias is something more than a speculative fiction. In this respect, more can be learned about E-expectancy, the modes of transmission of this expectancy, and the concomitant attributes involved by collating information from other studies which, while not specifically designed to study the E-bias phenomenon, show the expectancy effect.

Studies replicating Rosenthal's experimental paradigm, but with modifications in structure, might also yield interesting results. Since a transformation in the present study was not possible because of the absence of an explicit 0 point in the rating scale, a study employing the Empathy Test Rating Scale with an explicit 0 is suggested. Whether Rosenthal's current scale favours obtaining a bias effect because of its forced-choice nature, would be investigated. In this regard, a study employing Rosenthal's photograph-judging task with an explicit 0 point in the scale, might serve to demonstrate whether the E-bias phenomenon is as robust as implied, or whether it is the almost-inevitable result of a well-designed experiment.

A replication of Study II, with some improvement, might also prove to be of heuristic value. It might prove to be more appropriate if the number of dots used for each stimulus card was intermediate with respect to the two central rating point values. In this way, the Numerosity Estimation rating scale would be equivalent to Rosenthal's scale with an implicit 0 point. With this change introduced, the variable(s) responsible for the non-significant trend toward obtaining data opposite to those expected might be discovered. It might be demonstrated that under these modified conditions, a significant difference between groups in the Rosenthal replicate condition would be obtained. If a non-significant difference between groups in the Numerosity Estimation condition is obtained, these results would serve to suggest that type of task is a more crucial variable than task-structure.

Finally, studies more directed toward the investigation of S's experimental demeanour would be fruitful since E-expectancy has been infinitely more investigated than S-expectancy. The E-expectancy phenomenon does not occur *in vacuo*. Thus, studies similar in form to that of White (1962), where both E and S biases are instructionally-induced, might prove elucidating. Since S comprises one-half of this well-studied dyad, equal scrutiny should be given to S and his expectancies in the psychological experiment. Further, studies investigating attentional phenomena, or S's set might prove invaluable. Changing E's hypothesis through the course of an experiment has been attempted (Rosenthal et al., 1964b). It might prove equally instructive to interpolate different rating continua for S throughout the course of an experiment to observe if a "set" is operating. Thus, any studies elaborating upon Rosenthal's basic experimental paradigm and expanding into the realm of the generality of the E-bias phenomenon, would be of immense value since these represent unresearched areas.

CHAPTER V

SUMMARY AND CONCLUSIONS

This investigation was concerned with the generality of the E-bias phenomenon. Since the studies in E-bias have almost exclusively employed a person-perception task i.e., judging photographs of faces, it was proposed that both a "fact-centered" type of task as well as a more structured person-perception task would not demonstrate this bias effect. A "fact-centered" task was defined in terms of universality of agreement with minimal "attitudinal", or emotion-provoking properties. Structure, in this study, was defined in terms of the ambiguity the task held for S.

A Rosenthal replication and a Numerosity Estimation condition comprised Study I. Ten photographs of faces were used in the Rosenthal replication and 10 stimulus cards, with 200 dots mounted on each were used in the Numerosity Estimation condition. Positive and negative biases were instructionally induced in the Rosenthal replicate condition, while overestimation and underestimation biases were instructionally induced in the Numerosity Estimation condition. In Study II, a modified, 10-point Empathy Test Rating Scale (Fode, 1960) was used for the purpose of structuring the Rosenthal paradigm. A 10-point scale was also used in the Numerosity Estimation condition along with a more estimable number of dots (80) on each stimulus card. The experimental procedure was essentially identical to that of Study I.

The results confirmed the experimental hypotheses in that no E-bias effect was found for a "fact-centered" task, i.e. a numerosity estimation task and the E-bias effect, similarly, was not found for a more structured person-perception task, i.e. the modified Rosenthal replication. A significant bias effect was present, however, in the usual Rosenthal task.

It was concluded that type of task and task-structure are crucial variables in *E*-bias studies. The importance of these variables, in relation to the person-perception task and the algebraic rating scale used in the paradigmatic Rosenthal experiment were evaluated and discussed. Further experiments are suggested to explore these variables further and to extend our knowledge with regard to the generality of the *E*-bias phenomenon.

REFERENCES

Binder, A., McConnell, D., & Sjoholm, Darcy, A. Verbal conditioning as a function of experimenter characteristics. J. abnorm. soc. Psychol., 1958, 56, 133-135.

Cahalan, D., Tamulonis, Valerie, & Verner, Helen, W. Interviewer bias involved in certain types of opinion survey questions. Int. J. Opin. attit. Res., 1947, 1, 63-67.

Criswell, Joan, H. The psychologist as perceiver. In R. Tagiuri, & L. Petrullo, Person perception and interpersonal behavior. Stanford Univer. Press, 1958. Pp. 95-109.

Ebbinghaus, H. Memory: a contribution to experimental psychology. (1885) Translated by Ruger, H. A., and Bussenius, Clara E. New York: Teachers College, Columbia Univer., 1913.

Epstein, J. The role of verbal cues in the mediation of experimenter bias, Unpublished M.A. thesis, Univer. of Manitoba, Winnipeg, Canada, 1966.

Ferber, R., & Wales, H. G. Detection and correction of interviewer bias. Publ. Opin. Quart., 1952, 16, 107-127.

Fode, K. L. The effect of non-visual and non-verbal interaction on experimenter bias. Unpublished M.A. thesis, Univer. of North Dakota, Grand Forks, 1960.

Fode, K. L., Rosenthal, R., Vikan-Kline, Linda, L., & Persinger, G. W. Susceptibility to influence in a verbal conditioning situation. Unpublished data, Univer. of North Dakota, Grand Forks, 1961.

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

Garner, W. R. Uncertainty and structure as psychological concepts. New York: Wiley, 1962.

George, W. H. The scientist in action: a scientific study of his methods. New York: Emerson Books, 1938.

Greenspoon, J. The reinforcing effect of two spoken sounds on the frequency of two responses. Amer. J. Psychol., 1955, 68, 409-416.

Hanley, C., & Rokeach, M. Care and carelessness in psychology. Psychol. Bull., 1956, 53, 183-186.

Kintz, B. L., Delprato, D. J., Metee, D. R., Persons, C. E., & Schoppe, R. H. The experimenter effect. Psychol. Bull., 1965, 63, 1-23.

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

Lord, E. Experimentally induced variations in Rorschach performance. Psychol. Monogr., 1950, 64, No. 316.

Marcia, J. Hypothesis-making, need for social approval, and their effects on unconscious experimenter bias. Unpublished M.A. thesis, Ohio State Univer., Columbus, Ohio, 1961.

Merton, R. K. The self-fulfilling prophecy. Antioch. Rev., 1948, 8, 193-210.

Miller, G. A. The magical number seven, plus or minus two: some limits of our capacity for processing information. Psychol. Rev., 1956, 63, 81-97.

Mosteller, F. Correcting for interviewer bias. In H. Cantril (Ed.), Guaging public opinion. Princeton, New Jersey, Princeton Univer. Press, 1944. Pp. 286-288.

Mulry, R. C. The effects of the experimenters perception of his own performance on subject performance in a pursuit rotor task. Unpublished M.A. thesis, Univer. of North Dakota, Grand Forks, 1962.

Orne, M. T. The nature of hynosis, artifact and essence. J. abnorm. soc. Psychol., 1953, 48, 215-224.

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

Pflugraph, J. Examiner influence in a group testing situation with particular inference to examiner bias. Unpublished M.A. thesis, Univer. of North Dakota, Grand Forks, 1962.

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.

Riecken, H. A program for research on experiments in social psychology. In N. F. Washburn (Ed.), Decisions, values, and groups. New York: Pergamon Press, 1962. Vol. II, Pp. 25-41.

Rosenthal, R. Projection, excitement and unconscious experimenter bias. Amer. Psychol., 1958, 13, 345-346. (Abstract).

Rosenthal, R. Experimenter attributes as determinants of subjects' responses. J. proj. tech. pers. Assess., 1963, 27, 324-331.

Rosenthal, R. The effect of the experimenter on the results of psychological research. In B. A. Maher (Ed.), Progress in experimental personality research. Vol. I. New York: Academic Press, 1964. Pp. 79-114.

Rosenthal, R. Covert communications and tacit understanding in the psychological experiment. Unpublished manuscript. Harvard Univer., 1966.

Rosenthal, R., & Fode, K. L. (Psychology of the Scientist: V). Three experiments in experimenter bias. Psychol. rep. Monogr., 1963, 3, V12 (a).

Rosenthal, R., & Fode, K. L. The effect of experimenter bias on the performance of the albino rat. Behav. Sci., 1963, 8, 183-189. (b).

Rosenthal, R., Fode, K. L., & Vikan-Kline, Linda. The effect on experimenter bias of varying levels of motivation of Es and Ss. Unpublished manuscipt, Harvard Univer., Cambridge, Mass., 1960.

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

Rosenthal, R., Friedman, C. J., Johnson, C. A., Fode, K. L., Schill, T. R. White, C. R., & Vikan-Kline, Linda. Variables affecting experimenter bias in a group situation. Gent. psychol. Monogr., 1964, 70, 127-296.

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

Rosenthal, R., Kohn, P., Marks, Patricia, &nCarota, N. Experimenter's hypothesis-confirmation and mood as determinants of experimental results. Unpublished manuscript. Harvard Univer., Cambridge, Mass., 1965.

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

Rosenthal, R., & Lawson, R. A. A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats. Amer. Psychol., 1963, 18, 345 (Abstract).

Rosenthal, R., Persinger, G. W., Mulry, R. C., Vikan-Kline, Linda, & 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. W., Mulry, R. C., Vikan-Kline, Linda, & 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. W., Vikan-Kline, Linda, & Fode, K. L. The effect of experimenter outcome-bias and subject set on awareness in verbal conditioning experiments. J. verb. learn. verb. Behav., 1963, 2, 1275-283. (a).

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

Rosenthal, R., Persinger, G. W., Vikan-Kline, Linda, & Mulry, R. C. The role of the research assistant in the mediation of experimenter bias. J. Pers., 1963, 31, 313-335.

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

Wartenberg-Ekren, Ursula. The effect of experimenter knowledge of a subject's scholastic standing on the performance of a reasoning task. Unpublished M.A. thesis, Marquette Univer., 1962.

White, C. R. The effect of induced subject expectations on the experimenter bias situation. Unpublished doctoral dissertation, Univer. of North Dakota, Grand Forks, 1962.

Van Krevelen, Allice. A study of examiner influence on responses to maps test materials. J. clin. Psychol., 1954, 10, 292-293.

APPENDIX A

Principal Investigator's Instructions to _E_

in the Rosenthal Replications

"You have been asked to participate in a research project developing a test of empathy. The reason for your participation in this project is to standardize results of experiments of this type. There is the problem in psychological research of different examiners getting somewhat different data on the same tests as a function of individual differences in experimenters. Therefore, to standardize the tests it is better methodological procedure to use groups of experimenters.

"You will be asked to run a series of subjects and obtain from each, ratings of photographs. After you have run each subject you are to send her across the hall to me where she will fill in a personal opinion inventory. This inventory is a rather gross measure of individual differences. The experimental procedure has been typed out for you and is self-explanatory. (Subject peruses experimental procedure).

"According to preceding research of this nature, the type of subjects you will be using have averaged a (+5) (-5) (+3) (-3) rating. Therefore, the subjects you are running should also average about a (+5) (-5) (+3) (-3) rating.

"Just read the instructions to the subjects. Say _nothing_ else to them except hello and goodbye. If for any reason you should say anything to a subject other than that which is written in your instructions, please write down the exact words you used and the situation which forced you to say them. Your subjects will be coming in at about _________ and there is a note outside asking them to wait in the hallway until you get them. When you have finished with them, send them across the hall to me to Room 326. Good luck!"

$\underline{E}$'s Instructions to $\underline{S}$ in the Rosenthal Replications

"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. O.K? Have you participated in a similar experiment previously? That's fine, thank you.

"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) (-5) to (+10) (+5). A rating of (-10) (-5) means that you judge the person to have experienced extreme failure. A rating of (+10) (+5) 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 (moderate) failure while a rating of (+1) means that you judge the person to have experienced mild (moderate) 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)."

E's Instructions to S in the Modified Numerosity Estimation Condition

"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. O.K?

"We are in the process of developing a human engineering test. This test is designed to show those patterns which lend themselves most readily to accurate judgment of numerosity. I will show you a series of cards. For each one I want you to judge the number of discrete dots on the card. To assist you in your judgments you are to use this rating scale. As you can see the scale runs from 60 to 105. Try to match your estimation of the number of dots on each card with the appropriate number on the scale. Thus, a rating of 60 means that you judge the card to be slightly numerous. A rating of 105 means that you judge the card to be extremely numerous.

"You are to rate each card as accurately as you can. Just tell me the rating you assign to each card. All ready? Here is the first card. (No further explanation may be given although all or part of the instructions may be repeated)."