8-1-1970

Experimenter Bias Effect as a Function of Ambiguity

Gary D. Barko

EXPERIMENTER BIAS EFFECT AS A FUNCTION OF AMBIGUITY

by
Gary D. Barko

Bachelor of Arts, University of North Dakota 1965
Master of Arts, University of North Dakota 1967

A Dissertation

Submitted to the Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota

August
1970

This dissertation submitted by Gary D. Barko in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota is hereby approved by the Faculty Advisory Committee under whom the work has been done.

R. H. Kolstoe
(Chairman)

E. L. Hales

L. R. Q.

H. E. Edenstrom

Alice T. Clark

A. William Johnson
Dean of the Graduate School

Permission

Title EXPERIMENTER BIAS EFFECT AS A FUNCTION OF AMBIGUITY

Department Psychology

Degree Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the Chairman of the Department or the Dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Signature Gary D. Baker

Date July 23, 1970

ACKNOWLEDGEMENTS

I would like to thank my committee members, particularly Dr. Ralph Kolstoe, for the time they have spent in reading this paper and giving constructive criticism.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Previous research and discussion has suggested that task ambiguity and experimenter-subject familiarity with the task are important factors affecting the generality and strength of the experimenter bias effect. These factors were conceptualized as inter-task ambiguity (inherent in the task) and two types of intra-task ambiguity (experimenter's familiarity with the task and subject's familiarity with the task). It was hypothesized that experimenter bias would: (1) be a significant factor in a more ambiguous task and not a significant factor in a less ambiguous task, (2) more likely be communicated to the subjects who were less familiar with the task, and (3) more likely be communicated to the subjects as the experimenters became more familiar with the task.

Nine upperclassmen experimenters were randomly assigned to one of three bias level, high, medium, and low (expect +5, 0, or -5 mean ratings on the person perception task). Each experimenter tested 10 subjects (female volunteers from introductory psychology) in a random order. Five subjects rated high ambiguity photos (mean ratings near 0 on the success-failure dimension under a no-expectancy condition) and five viewed low ambiguity photos (mean ratings high in the success direction under the no-expectancy condition).

Five analyses of variance were computed; a significant experimenter bias effect was found with the high ambiguity stimilus photos but not with the low ambiguity photos. However, the bias levels were

not communicated in the predicted order. The high and low expectancy experimenters were both significantly higher than the medium bias experimenters. The hypotheses pertaining to intra-task ambiguity were not supported by the data.

From this study one can conclude that the experimenter bias effect is a function of ambiguity inherent in the task. The role of intra-task ambiguity is less clear and no conclusion can be stated based on the present data.

CHAPTER I

INTRODUCTION

The last 10 years have seen a marked rise of interest in the role of the experimenter in psychological research. Psychologists have only recently begun to take cognizance of experimenter-subject interaction in the social setting of psychological research. McGuigan (1963) in an article entitled "The Experimenter: A Neglected Stimulus Object" discussed this problem and made a strong case for putting the experimenter back in the experimenter-subject interaction.

Experimenters seemingly have accepted research findings without due regard to the dangers of what Merton (1948) has called "self-fulfilling phophecy." Merton discussed the phenomena of self-fulfilling prophecy in relation to minority group stereotypes more than psychological research *per se*, but the conceptualization is readily applicable to the latter. He saw researchers and the general public as sometimes intentionally, but more often unintentionally affecting situations so that they come to reflect the *a priori* hypotheses or prophecies of the person. Merton strikingly illustrated this point with an anecdote about bank failure. Depositors in a solvent bank caused the bank's failure through a mistaken belief in its lack of solvency.

Orne (1962) has discussed at length the problems that have developed in the social sciences from using the basic research paradigm developed to deal with phenomena in the physical sciences. The problem

derives from the fact that the physical sciences deal with inanimate objects--passive responders. In the social sciences the objects are anything but inanimate passive responders. Orne stated that the subjects' motives in the experimental situation, his perception of the experimental situation, and what he is responding to have often been ignored or taken for granted in social science research. Orne discussed the usual subject in psychological research as being the "good" subject who works to please the researcher. The "good" subject sees his task as finding out what the experimenter "really" wants and following through so as not to upset the experimenter or give him unsatisfactory results. The totality of cues which might convey the experimental hypothesis to the subject are termed the "demand characteristics" of the situation by Orne. These demand characteristics include rumors about the research, the person of the experimenter, the laboratory setting, and more important for the purposes of this study, explicit and implicit communication between the experimenter and the subject during the experiment. Where the "good" inanimate subject responds ideally to only the experimental variables, the animate "good" subject responds to the demand characteristics of the situation as well as the experimental variables.

Orne and Scheibe (1964) and Raffetto (1967) investigated the role of the demand characteristics of the situation in the sensory deprivation effect. Orne and Scheibe found significantly more reports of hallucinatory experiences and the other usual sensory deprivation effects when a "panic button" was available to the subjects and when the research was carried out in a medical setting with an "emergency tray" present. Raffetto found more reports of hallucinatory experiences

from the subjects when the experimenters expected them than when the experimenters did not expect them. Seemingly the experimenters' expectancies were covertly being communicated to the subjects and the subjects were responding accordingly in the Raffetto study. The two studies successfully showed that the sensory deprivation effect may in part be due to the subjects responding to the demand characteristics of the situation and not to sensory deprivation per se.

The demand characteristics of the situation are the totality of cues available to the subjects. The focus of this study is the experimenter and his interaction with the subject. The focus is upon the experimenter as in the Raffetto (1967) study more than upon the physical aspects of the situation as in the Orne and Scheibe (1964) study.

The experimenter variables have been divided into two broad classes by Barber and Silver (1968a) and Rosenthal (1966). The first class of experimenter variables are the essentially inherent aspects of the experimenter such as age, sex, race, status, ethnic characteristics and personality traits. These aspects of the experimenter have been conceptualized as part of the experimenter personal attributes effect by Barber and Silver and simply the experimenter effect by Rosenthal. These variables, while important in the experimenter-subject interaction, are of only indirect interest as control variables in this study. The second class of experimenter variables have been conceptualized by Barber and Silver and Rosenthal as the experimenter bias effect. This effect has at various times been called the experimenter outcome-orientation effect, the experimenter expectancy effect,

the "Clever Hans Phenomenon," as well as the experimenter bias effect by Rosenthal (Barber & Silver, 1968a). This effect is conceptualized as being produced by the expectancies, desires or biases of the experimenter.

In almost all psychological research the experimenter expects or hopes for different responses from different subjects or groups of subjects. The experimenter rarely takes specific measures to preclude implicit communication of these a priori hypotheses and expectations to the subjects. If research hypotheses or expectations are easily and often implicitly communicated to subjects, the results of much psychological research are open to question. The generality of the experimenter bias effect is a very important question for future research as well as for the evaluation of past research.

The focus of this paper is on the role of task variables, specifically the ambiguity of the task, in the occurrence and magnitude of the experimenter bias effect. Intuitively, the more ambiguous the task, the more likely the experimenters' biases will be a factor in determining the subjects' responses.

CHAPTER II

HISTORY AND STATEMENT OF THE PROBLEM

The literature pertaining to unintended experimenter effects is very rich and diverse. Indirectly relevant information could be included from many areas. The literature review will be limited as much as possible to directly relevant studies which utilize standardized treatment and assessment procedures. This has the effect of reducing the scope of this study to the experimenter bias effect as discussed and investigated by Rosenthal (1966). Studies not based on this paradigm are to be included only if directly relevant to the emphasis on task variables.

The concept of unintended experimenter effects broadly defined has been of major interest to clinicians and psychotherapists (Goldstein, 1962; Troffer and Tart, 1964), educators (Rosenthal and Jacobson, 1968) and survey interviewers (Ferber and Wales, 1952). Typically, researchers in these areas have not used standardized procedures. In the Rosenthal and Jacobson (1968) study, for instance, elementary school teachers were led to believe that certain students had unusual potential for intellectual growth while others were only average. Standard treatment procedures were lacking in that the teachers were permitted to treat the children differently.

Several additional lengthy reviews of the general literature dealing with unintentional experimenter effects are available. These

include a book-length treatment by Friedman (1967), a lengthy doctoral dissertation by Fode (1967), a journal review by Kintz, Delprato, Mettee, Persons and Schappe (1965), and numerous journal articles by Rosenthal (1963, 1964a, 1964b, 1967a, 1969). The prime source in the area is a book entitled "Experimenter Effects in Behavioral Research" by Rosenthal (1966). Recent more specialized review articles in the area include one by Silver (1968) on experimenter modeling and one by Glixman (1967) on effects of examiner, examiner-sex and subject sex upon categorizing behavior.

Seemingly the earliest study dealing specifically with unintended experimenter effects in a straightforward experimental task was one done by Stanton and Baker in 1942. Stanton and Baker presented 12 nonsense geometric figures to 200 undergraduate subjects. Retention of these figures was measured by five experienced workers some time after the presentation session. The experimenters were given keys which contained correct responses but also some incorrect responses. The experimenters were explicitly warned to avoid any bias associated with having the keys before them while interviewing the subjects. Stanton and Baker found that what the subjects retained tended to be in agreement with what the experimenters believed to be correct. The subjects significantly more often gave correct responses when the experimenter had the correct key than when the experimenter's key was incorrect.

Lindzey (1951) replicated the Stanton and Baker study, but emphasized to his experimenters to keep the key out of sight from the subjects. The results failed to confirm the Baker and Stanton findings. Friedman (1942) also failed to obtain the statistical significance obtained in the

original study. Stanton (1942) strengthened the conclusions stated in the original Stanton and Baker (1942) study. The responses expected by experimenters were found more often than expected by chance. Seemingly, the experimenters were cueing the subjects to what was expected or recording what they expected to hear in the subject's responses.

Most of the recent work on unintended experimenter effects has been based on Rosenthal's (1966) person-perception task. The subject is shown a series of photographed faces and is asked to rate each on a scale ranging from -10 to +10 whether the person has been experiencing failure or success. The subject is instructed to rate a +10 for extreme success and a -10 for extreme failure with intermediate labeled points. The photographs were originally standardized by administering them under a "no-expectancy condition" to a large sample of undergraduate students. The photographs selected for presentation to the experimental groups averaged near zero under the no-expectancy condition. Typically the subjects and experimenters were students. Before seeing the student subjects, the student experimenters were usually told to expect subject ratings to average +5 or -5. The student experimenters were also led to believe that they were replicating "well established findings."

In the first experiment based on this paradigm (Rosenthal, 1966) 10 advanced undergraduates and graduate students of psychology served as experimenters. Each student experimenter was assigned about 20 subjects from an introduction to psychology course. The 10 photos used as stimulus items had a mean no-expectancy rating near zero. Half the student experimenters were told that people generally rated the photos as moderately successful (ratings of +5) and half the experimenters were told that people generally rated the photos as moderately unsuccessful

(ratings of -5). The rest of the instructions given to the experimenters were the same for both groups. They were given identical instructions to read to their subjects and all were cautioned not to deviate from these instructions. The experimenters were told that they were participating in the study to see how well they could duplicate "well established findings."

According to Rosenthal (1966) and Rosenthal and Jacobson (1968) the results were clear. The experimenters that expected higher average ratings found significantly higher average ratings. Two replications were conducted which confirmed the original findings. The experimenters tended to obtain the results that they expected to obtain.

Subsequent research reported by Rosenthal (1966) and his associates dealt with the generality of the phenomena, mediating variables, and parametric studies involving experimenter variables. The results seemed clear to Rosenthal (1963) as he stated that the experimenter expectancy effect is "both a fairly general and fairly robust phenomenon" (p. 271).

Rosenthal and Jacobson (1968) in discussing the individual differences among experimenters stated:

> Those experimenters who show greater self-fulfilling effects of prophecies tend to be of higher status in the eyes of their subjects, and they conduct their experiments in a more professional, more competent manner. They are more likeable and more relaxed, particularly in their movement patterns, while avoiding an overly personal tone of voice that might interfere with the business at hand (p. 28).

Friedman, Kurland, and Rosenthal (1965) also discussed what they called "professionalness" as a factor in experimenter bias communication. The Friedman, Kurland and Rosenthal article was based on the examination of moving pictures of experimenter-subject interactions.

Rosenthal's statements about the generality of experimenter bias have recently been criticized in a series of articles by Barber and his associates (Barber, 1969; Barber & Silver, 1968a, 1968b; Barber, Calverly, Forgione, McPeake, Chavis, & Bowen, 1969). Barber and his associates criticized the earlier studies in the area on the basis of the statistical analyses used. They pointed out that many of the studies (seemingly the majority of them) did not clearly demonstrate the experimenter bias effect. The faults in the analyses of results included failure to perform an overall statistical analysis to exclude the null hypothesis and failure to avoid "probability pyramiding" when postmortem tests were performed. Barber and Silver (1968a) concluded that "the experimenter bias effect appears to be more difficult to demonstrate and less pervasive than was implied" (p. 23). They also felt that in some studies purporting to show the effect the student experimenters misjudged, misrecorded or misreported the results. Some studies they analyzed did show the experimenter bias effect, however.

Rosenthal (1968) defended his earlier conclusions. The defense was based on combining the probabilities from 12 studies which showed the experimenter bias effect. The combined probability was less than 1/1,000,000. Rosenthal held that analyzing results of single experiments as Barber and his associates had done was a misleading procedure. He stated that the studies should be viewed as a run of experiments and the combined probability for the 12 studies was only 1/1,000,000 that the results were a chance happening.

Rosenthal (1969) also criticized the Barber et al. (1969) failures to replicate the findings of the earlier Rosenthal studies.

Rosenthal felt that the Barber *et al.* studies could not be regarded as serious attempts to replicate because of differences in experimenters and subjects from his studies.

Other recent studies based on the Rosenthal paradigm have also failed to replicate the original findings. Jacob (1968) included a zero expectancy control group and directly probed the subjects' perception of the experimenters' expectancies. The overall *F* based on the subjects' mean ratings of the photos was not significant ($F<1.0$). The +5 subjects, when directly asked, thought that their experimenters expected a mean of .38, the -5 subjects a mean of 1.17, and the zero expectancy control group thought that their experimenters expected a mean of 3.04. The *t* between the +5 group prediction of experimenter expectancy and the zero group was significant; the other *t*s were not significant. Jacob concluded: "Coupled with the Barber and Silver conclusions (Barber & Silver, 1968a), the present findings indicates that assuming the stability of the effect and focusing on particular modes of mediation is rather premature, and it is suggested that subsequent research be directed toward specification of conditions under which the effect actually emerges" (p. 240). Another recent study (Kennedy, 1969) also failed to detect the Rosenthal effect in a verbal conditioning task. Kennedy's experimenters were given more extended pretask indoctrination than is usual; the author thought that this might have been a factor that attenuated the effect.

Verbal Mediation

Verbal communication seems very important in the experimenter bias effect. Adair and Epstein (1968) and Epstein (1966) reported on

a study in which verbal cues alone were sufficient for communication of the experimenters' expectations. Six male student experimenters administered the basic person-perception task to 60 female student subjects. Each experimenter ran five subjects under the low expectancy (-5) condition and five under the high expectancy (+5) condition. The reading of the instructions was tape recorded during theses sessions and 60 more female subjects run using only the tape recorded voice as the experimenter. A significant experimenter bias effect was found in both the visual and non-visual situations. The voice of the experimenter reading the instructions seemingly was enough to communicate the expectancies to the subjects.

Fode (1960, Rosenthal & Fode, 1963) also indicated the great importance of verbal cues in experimenter bias communication. Fode restricted experimenter-subject visual contact in one group and the experimenters' verbal communication in a second group and compared the subjects' performance in these groups with conditions in which there were no restrictions. Fode concluded that while verbal cues are sufficient to mediate experimenter bias, visual cues increased the effect.

Intra-task and Inter-task Ambiguity

Ambiguity can be conceptualized as being of two kinds. The first type, intra-task ambiguity, refers to differences in ambiguity in one task over time. For example, a task completely new to an experimenter and subject would be more ambiguous than a very familiar task. As the initially unfamiliar task becomes more familiar over the course of the study the ambiguity would decrease. Intra-task ambiguity for a

subject would be less on later pictures in the person-perception task than on earlier pictures. For the experimenter intra-task ambiguity would be less with subjects seen later in the study than with subjects seen earlier in the study.

Inter-task ambiguity refers to characteristics of the task itself. Some tasks are just more ambiguous than others. Riecken (1962) suggested that there are two general types of experimental tasks "that provide subjects with different kinds of hints as to how to put their best foot forward, or, in effect, urge subjects to adopt one or another 'set' toward the experiment" (p. 35). He called one type "task ability" and the other "self-quality." Riecken stated that:

> A "task-ability" set is characteristically adopted when the experimenter presents the work to be done as involving some ability, skill or capacity to perform. The task may be motor or mental, simple or complex, familiar or strange, e.g. estimating the number of dots on a card, judging "auto-kinetic" movement distances, judging the personality of another or solving "human relations problems." The outstanding feature of such assignments is that there is no upper limit on the amount of skill or capacity the subject "ought" to display (p. 35).

In task-ability problems the positively valued end of the ability continuum is generally known to the subjects. The subject usually works to his limit in the task; it is impossible for him to do more than his best. The subject can misrepresent his performance in only one direction and that is to do less than he is capable of doing.

The self-quality problems are quite different according to Riecken. They "can be characterized in general as being concerned with opinions and beliefs; with responses to frustration, insult, and failure; with conformity-independence, choice-rejection or others; or with qualities such as dogmatism, authoritarianism, punitiveness and the like" (p. 35). In general the self-quality tasks are more open-ended, more

ambiguous than the task-quality problems according to Riecken. The task-quality problems involve one dimension of "good-bad" performance while the self-quality problems "tend to have two bad extremes and a good point located somewhere between the extremes, though not necessarily in the 'middle'" (p. 36). In order to be a "good" subject in a self-quality task the "subject must either draw from his pool of common sense knowledge about what 'anybody knows' . . . or . . . he would have to know the scheme of relevance that the experimenter is employing: the hypothesis being tested, the categories into which the behavior will be placed, the criteria for such placement, and the value assigned to category" (p. 36). Generally the experimenter conceals this information from the subject and this tends to "maximize the negotiation" between subject and experimenter. The person-perception task generally used in the area of experimenter bias would be considered a self-quality situation.

Intra-task Ambiguity

The experimenter's experience with a particular task seems to be a relevant variable in the experimenter's performance. Brogden (1962) reported that naive experimenters differed in the speed with which they conditioned rabbits but that the initial differences disappeared as the experimenters became more experienced. Cordaro and Ison (1963) in a study of observer bias in classical conditioning of the planaria stated that while a significant bias effect was found, it was to be attributed to experimenters using different response criteria. The experimenters were given little information as to when a response should be recorded as a response. Shinkman and Kornblith

(1965) suggested that the degree of observer bias is not as great with more experienced experimenters. Rosenthal and Halas (1962) reported that even experienced researchers showed significant differences when asked to record turning responses and contraction responses in planaria under a no-false expectancy condition. However, the "experienced" researchers in the Rosenthal and Halas study were not particularly experienced in the area of observing planarian behavior. The differences may have been less if the response criteria had been more stringently defined and the experimenters had been given more training. The task of planaria observing may be just too ambiguous for human observers to perform reliably. Cordaro and Ison (1963) suggested that cameras be used to more objectively record planarian responses.

Ingraham and Harrington (1966) required 27 initially naive experimenters to condition "dull" and "bright" rats in a bar-press task. The rats were actually randomly assigned to the student experimenters. The authors reported no significant overall experimenter bias effect but found evidence to indicate that early trials showed some bias effect. They concluded that the bias present was an initial response and not a continuing response in a decreasingly ambiguous situation. Other factors pointed out as possibly having a role in attenuating the experimenter bias effect were the relatively long experimenter pretraining period and the fact that the experimenters were given mixed expectancies. The individual experimenters were led to expect different performances from different rats. Ingraham and Harrington suggested that bias is more often shown where the experimenter is given only one expectancy instead of mixed expectancies.

Rosenthal (1967b, 1967c) reanalyzed the data from the Ingraham and Harrington (1966) study and suggested that the results did demonstrate experimenter bias. The authors responded (Harrington, 1967; Harrington & Ingraham, 1967) to Rosenthal's criticisms. They defended their original conclusions and questioned Rosenthal's analysis of the data. They restated their original conclusion that the bias shown was an initial but not a continuing response in a decreasingly ambiguous situation.

Two studies based on the person-perception task indicate that the bias effect is operating on the earliest photos of a subject's trials. Rosenthal, Fode, Vikan-Kline, and Persinger (1964) looked at the temporal aspects of bias communication in three earlier studies. In general, they found that the magnitude of the expectancy effect was somewhat greater for the first photo alone, than for all ten photos combined. The subjects seemed to be somewhat more affected by the experimenters' bias on earlier trials when the task was relatively unfamiliar than on later trials when the task was more familiar. The tendency for the experimenter bias effect to decrease over the series of photos was more strikingly shown in a study by Weick reported by Rosenthal (1966). Weick required two experimenters to administer the person-perception photo rating task to 10 introductory psychology students. One of the experimenters presented the cards to five subjects under a high bias expectancy and the second experimenter saw the other five students under a low bias expectancy. The experiment was conducted in front of Weick's class in experimental social psychology. The subjects rated the photos differently ($t = 2.93$, $p = .01$, one-tail) in line with the experimenters' biases, but the first responses were more affected than were the subsequent responses. The bias effect was significant on the very first

photo but diminished significantly (p = .08) for the last 10 photos of the standard 20-photo set.

In the Weick study, the effect of subjects' intra-task ambiguity, conceptualized as familiarity, is clearly shown. As the ambiguity of the task decreased through increasing familiarity, the subjects showed significantly less experimenter bias effect. The mean ratings of the subjects being tested by the high bias experimenter tended to approach the mean ratings of the subjects being tested by the low bias experimenter.

There is very little research that is directly relevant to the question of the role of the experimenter's experience in the communication of the experimenter's expectancies involving the person-perception task. Vikan-Kline (1962) in an investigation of the effect of the experimenter's perceived status on the mediation of experimenter bias found no order effect among her lower status experimenters but did find an order effect among her higher status experimenters. The higher status experimenters showed a significant tendency (p = .01) to influence subjects seen later in the study more than subjects seen earlier in the study. Rosenthal (1966) discussed the effect as learning to communicate unintentionally. He concluded that "although the evidence is not conclusive, it does seem that, on the whole, later-contacted subjects are more influenced by the experimenter's expectancy than earlier-contacted subjects" (p. 301-302).

Inter-task Ambiguity

Many authors have suggested that the ambiguity of the task is an important factor controlling the generality of the experimenter bias

effect. As Barber and Silver (1968a) put it: "The hypothesis suggested . . . is that the effects of experimenter's expectancies on the results of his research vary directly with the ambiguity, lack of structure, or non-factualness of the experimental task. Further research is needed to test this hypothesis" (p. 26). Masling (1966) strongly emphasized the importance of task ambiguity: "The stimulus variable is crucial. The more clearly the stimulus can be perceived, the less opportunity for projecting experimenter or S (subject) bias" (p. 92). Shames and Adair (1967) also discussed task ambiguity: "the type of task is a critical factor limiting the generality of experimenter-bias effects" (p. 6). Ambiguity has also been discussed in the context of interviewer bias. Ferber and Wales stated that "interviewer bias is more likely to crop up on attitudinal questions than on questions of fact" (1952, p. 116).

While many persons have talked about the role of ambiguity in experimenter bias, few studies have dealt with ambiguity per se. As one would expect, however, those tasks that are more fact oriented seem not to show the experimenter bias effect as often as the more attitude-oriented tasks.

Ekren (1962) studied the effect of experimenter bias on a relatively unambiguous task, the block design segment of the Wechsler Adult Intelligence Scale. The block design task would be of the fact-oriented type (Ferber & Wales, 1952) or the task-ability type (Riecken, 1962). No significant experimenter bias effect was found. Pflugrath (1962) found no overall significant experimenter bias effect in a standardized paper and pencil test, the Taylor Manifest Anxiety Scale. The scale was administered to groups of subjects, however, so experimenter-subject interaction was minimal. Pflugrath concluded that "Examiner bias

in the group testing situation is probably not a particularly robust phenomenon" (p. 33). The questions on the Taylor Manifest Anxiety Scale tend to be of the fact-oriented type; the bias effect may be a factor only in individual testing situations, if at all.

The ink-blot projective tests seem to serve as ready vehicles for the demonstration of experimenter bias. Masling (1965) required 14 graduate student volunteers for a quick course in Rorschach technique to administer the cards to undergraduates. The experimenters were led to expect different results. Half were told that experienced experimenter found more human than animal responses on the Rorschach task; the other seven experimenters were told the opposite--experienced experimenters found more animal than human responses. The data were in agreement with the induced expectancies. Marwit and Marcia (1967) based their study on achromatic reproductions of Holtzman ink blots. Undergraduate psychology-major experimenters administered the task to introductory psychology-student subjects. One group of experimenters formulated their own expectancies about the number of responses the subjects would give to the stimulus cards. The members of the second group were differentially indoctrinated. Some were led to expect many responses while others were led to expect few responses. The experimenters tended to get the results they expected, even when the expectancy was self induced as with the first group of experimenters. The authors concluded that "Bias (was) found to be an especially strong phenomenon" (p. 253). Strauss (1968), however, found no significant experimenter bias effect in the Rorschach ink-blot task. Strauss stated that "E (experimenter) expectance does not appear to affect a centrally significant Rorschach variable in a personality assessment situation" (p. 129). The variable in question was movement

domination or color domination of the movement to color ratios in the responses. The crucial difference between the Strauss study and the preceding two was that Strauss attempted to approximate a "real-life" assessment situation. The graduate student experimenters had completed a full course in projective techniques. The experimenter expectancies were not specifically induced as in the two preceding studies and the experimenters had different expectancies for different subjects. One must conclude that while the ink-blot projective tests can be readily used to demonstrate the experimenter bias effect, the effect is probably less powerful, less general in actual personality-assessment situations.

Shames and Adair (1967) studied the effect of type of task on the experimenter bias effect. The first task was the person-perception problem and the second, a more fact-oriented task, involved estimating the number of dots on a card. The person-perception task was a replication of the basic Rosenthal paradigm. Some of the student experimenters were led to expect judgments of success (+5) from their student subjects, and the rest were led to expect judgments of failure (-5). The experimenters were led to believe that they were replicating "well-established findings." The numerosity estimation task required the subjects to estimate how many dots were on each of 10 stimulus cards. The individual cards contained 200 dots. One group of experimenters expected overestimates (average 210 rating) and the second group expected underestimates (average 190 rating). A significant experimenter bias effect was found on the person-perception task, but not on the less ambiguous numerosity estimation task.

Weiss (1967, 1969) attempted to vary task ambiguity by varying tachistoscopic exposure time of 15 slides containing differing numbers

of dots (7 to 10 per slide). For high ambiguity the exposure time was .1 second, for moderate ambiguity .5 second, and for low ambiguity 5 seconds. The volunteer student experimenters were divided into three groups and given different expectancies. One group expected the subjects to underestimate, the second group expected average estimates, and the third group expected overestimates. A basic problem found was that all subjects underestimated the number of dots when the exposure time was shortest, .1 second. No firm conclusions can be drawn from the studies.

Wessler (1968, 1969; Wessler & Strauss, 1968) explored the relationship of the experimenter bias effect to various types of tasks. In general, no strong experimenter bias effect was found in any of the tasks studied. One of the tasks was the original Rosenthal person-perception task. In two attempts to replicate the original findings no significant experimenter bias effect was found. In the first attempt the results were in the opposite direction from what was expected; in the second, the results were in the predicted direction but not statistically significant. No significant experimenter bias effect was found in a reaction time study either. The reaction time task would be considered a task-ability situation (Riecken, 1962). In a comparative study of three different tasks, person-perception, judging of line lengths, and the tapping of dots into circles, no overall significant experimenter bias effect was found. The tasks were selected on the basis of decreasing ambiguity. The person-perception task was thought to be most ambiguous, the judging of line lengths of moderate ambiguity, and the dot tapping task of least ambiguity. While no overall significant bias effect was found on any of the tasks, subjects showing

bias in the predicted direction on the most ambiguous task, the person-perception task, were significantly biased on the moderately ambiguous task but not significantly biased on the least ambiguous task, Wessler concluded that "These trends are congruent with the hypothesis that the more obvious the correct response is to S (subject), the less susceptible S's (subject's) performance is to E (experimenter) expectancy effects, probably because S need not seek information about how his performance will be evaluated by E " (1969, p. 66).

Masling and Rabie (Masling, 1966) studied the effect of varying the ambiguity of the stimulus items within a person-perception situation. From a pool of 70 high school class pictures two sets of seven pictures were chosen to serve as the stimulus items. The pictures were selected on the basis of perceived attractiveness as judged by female subjects. While the two groups of pictures had the same mean attractiveness, they varied considerably in homogeneity. One group was fairly homogeneous; the pilot subjects were in fairly good agreement about the degree of attractiveness displayed. The second group of pictures showed much more variability in judged attractiveness. The low variability photos were less ambiguous than the high variability set of photos. The experimenters were told that the task was designed to study the relationship between the self-concept of freshman female subjects with the ratings of attractiveness the subjects would assign to the photos. Half of the experimenters were told that their subjects were high in self-concept, would show good acceptance of others, and therefore tend to give high attractiveness ratings. The other half of the experimenters, a control group, were given no expectancies. The experimental hypothesis was that the subjects in the experimental group would show a significant

difference between the two sets of stimulus items; the more ambiguous items would show biased ratings, be rated more attractive. The control group would show no significant difference. The results of the study were in the predicted direction but did not attain statistical significance.

Recapitulation

Several studies involving animal subjects (Brogden, 1962; Cordaro & Ison, 1963; Ingraham & Harrington, 1966; Rosenthal & Halas, 1962; Shinkman & Kornblith, 1965) have presented evidence that an experimenter's or observer's familiarity with a task is a factor in his performance. The general tenor of the conclusions was that less experienced researchers were more likely to bias their results in animal studies. Very little research has been done on the role of the experimenter's experience involving the person perception task. Vikan-Kline (1962) and Rosenthal (1966) suggested that more experienced experimenters are more likely to display a bias effect on the person perception task.

The general conclusion about a subject's performance relative to his experience seems to be that as a subject becomes more familiar with a particular task he tends to be less affected by the experimenter's expectancy. This tentative conclusion was supported by research conducted by Rosenthal, Fode, Vikan-Kline, and Persinger (1964) and Weick (Rosenthal, 1966).

Several authors have suggested that inter-task ambiguity is an important factor in the experimenter bias effect (Masling, 1966; Weis, 1967, 1969; Wessler, 1968, 1969; Shames & Adair, 1967; Wessler & Strauss, 1968). Only one of the authors, Masling (1966) attempted to vary

inter-task ambiguity within the person perception task; the results were inconclusive.

Statement of the Problem

This study will investigate the role of inter-task ambiguity and intra-task ambiguity in the experimenter bias effect. Past studies have suggested that task ambiguity is a crucial factor limiting the generality of the experimenter bias effect. Past studies have also suggested the importance of intra-task ambiguity. As experimenters and subjects become more familiar with a task their responses change; previous studies have led to tentative conclusions about these changes.

Hypotheses

1. The experimenter bias effect is a function of inter-task ambiguity.
2. The experimenter bias effect is a function of intra-task ambiguity.
    a. The experimenter bias effect decreases as a function of the subjects' increasing familiarity with a particular task.
    b. The experimenter bias effect increases as a function of the experimenters' increasing familiarity with a particular task.

CHAPTER III

METHOD

The Stimulus Photos

Two groups of 10 photos selected from an initial pool of 63 were used in this study. The initial pool of photos, taken from weekly news magazines, included both known and relatively unknown persons. In the eyes of the author, the initial pool of photos sampled all points of the success-failure dimension in terms of facial expression, reputation, clothing, and pictured background.

The initial pool of 63 photos were shown by the author to 27 female volunteers from the introduction to psychology class in individual sessions under a no-expectancy condition. The pilot study subjects were asked to rate the photos on a 20-point scale (-10, extreme failure, to +10, extreme success, with no 0 point) as to the degree of success or failure that the pictured person was experiencing. To insure against expectancy communication between the author and the pilot subjects, the author avoided looking at which particular photo the subject was viewing and recorded results by noting a randomly assigned identification number on the back of the photo. The pilot photos were shuffled between subjects to avoid serial effects.

The two groups of 10 photos selected for use in the study were chosen on the basis of the no-expectancy mean ratings. The photos selected for the high ambiguity group had mean ratings near 0, the

neutral point. Ratings of success-failure for these tended to vary symmetrically about the mean rating. Photos selected for the low ambiguity group were those with the highest mean ratings. The ratings of these high success photos were less free to vary due to the +10 ceiling for ratings. It was thought that the highly rated photos would be a problem more of recognition that attitude for raters while the photos rated near the neutral point would represent more of an attitude-oriented problem. The highly rated photos tended to be relatively well known persons such as Nixon, Dustin Hoffman, Johnny Carson, and Dick Cavett; the high ambiguity group tended to be relative unknowns. A second difference between the groups was that the persons pictured in the low ambiguity photos tended to have more "smiley" expressions; the pictures with means near the neutral point, in the eyes of the author, had expressions that were much more enigmatic.

The mean ratings and standard deviation of the photos are contained in Table 11 of Appendix B. As can be seen from the table, the photos selected for the more ambiguous set had mean ratings of from -.9 to +.9 while the mean ratings of photos selected for the relatively unambiguous set ranged from +6.1 to +8.2. The SDs for the ambiguous photos tended to be larger than the SDs for the unambiguous photos. The mean SD for the ambiguous photos was 4.29 and the mean SD for the unambiguous photos was 3.16.

The selected photos were mounted on 8 x 7 inch tabbed pieces of cardboard and placed in individual slots in an accordian-type expanding folder. The pictures were randomly ordered to guard against having all the ambiguous photos in a group and all of the unambiguous photos in

another group. Letter identifications were written on the tabbed portions of the cards.

Experimenters

The experimenters were nine male volunteers from the abnormal psychology class. All nine were psychology majors in their junior or senior year of study. In soliciting for volunteers, the author stated that the study involved research on "empathy." The volunteers would get actual research experience as research assistants. As a further inducement to volunteer, the students were told that if the research turned out well, it would be published with the volunteer research assistants' names in a footnote. The students that volunteered were asked to wear coat and tie for the actual study.

Subjects

The subjects were 90 female volunteers from the introduction to psychology class. Volunteers were solicited through recitation instructors. Women students were asked to volunteer for a "picture" study that would take only 10-15 minutes of their time. They were asked to sign up for one of nine experimental sessions scheduled an hour and a half apart (6:30, 8:00, and 9:30) on three succeeding evenings. It was found to be quite difficult to get volunteers for the 9:30 sessions. The author had to call students personally to get enough volunteers for that time slot. Each volunteer was called shortly before her schedule time to remind her of when and where the study was scheduled.

Procedure

The nine student experimenters were randomly assigned to one of three bias conditions, low bias (expect -5 mean ratings), neutral bias

(expect 0 mean ratings), and high bias (expect +5 mean ratings). Three experimenters, one from each bias condition, were scheduled to test individually 10 subjects each night of the study. The three experimenters from each bias condition were scheduled in different time slots over the three nights of the study. The experimenters were scheduled 90 minutes apart so that there was no overlap between groups; each experimenter took about 45 minutes to test his 10 subjects.

The experimenters were read standardized instructions by the author before their testing sessions. The instructions were as follows: You have been asked to assist me in a research project developing a test of empathy. I am using different experimenters and subjects to approximate real test situations. Your task will consist of showing a series of photos of persons to about 10 female subjects from the introduction to psychology class and asking them to rate the photos as to the degree of success or failure that the pictured person has been experiencing. The subjects differ rather markedly on tested anxiety level and on tested self-confidence. These basic personality differences are hypothesized to relate to how empathetic the person is--how well he is able to identify with others. For example, highly anxious-low self-confidence subjects will tend to be generally pessimistic and tend to rate others as experiencing relative failure. Low anxiety-high self-confidence subjects are thought to generally view others in a more favorable light and consequently tend to give ratings of relative success. The pictures are to be shown to the subjects in pre-arranged random orders as indicated on the subjects' rating sheets. The instructions you are to read to the subjects are self-explanatory. Try to put the subjects at ease while reading the instructions and going over the information at the top of the rating sheet. If

everything is working as expected your subjects should give average ratings of about (+5, -5, 0). Please do not discuss this project with anyone until all the subjects are run which will be by this weekend. If it gets around what the study is all about, it may contaminate the results.

The only difference in the instructions read was the figure given for the average rating expected. The number read to the student experimenter was in accordance with which bias condition he represented.

The 10 subjects seen by each experimenter were randomly divided into two groups. Five subjects were shown only the 10 high ambiguity photos and 5 subjects were shown only the 10 low ambiguity photos. Each experimenter tested his 10 subjects in a prearranged random order and the 10 photos each subject rated were presented in a prearranged random order.

All testing was done in the same one-way mirror observational room. The subjects were met in a waiting room across the hall from the testing room and ushered into the testing room one at a time by the chief investigator. The testing room contained a chair on each side of a small table on which was placed the mounted photos in their container. A copy of the rating scale to be used was in constant view so that the subjects could refer to it during the course of their testing. A copy of this rating scale, based on one used by Rosenthal (1966), is contained in Appendix A.

The instructions read to the subjects by the experimenters were also based on those used by Rosenthal (1966). They were as follows: 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 in numerical terms--for example, -10, +10, +3 or -3. All ready? Here is the first photo.

Each subject carried into the testing room a rating sheet which indicated the pictures the experimenter was to present and a place for the experimenter to record the ratings given. The upper part of the rating sheet asked for routine information such as name, age, major, recitation number and instructor, and year in school. These questions were answered in the waiting room before the subjects were ushered into the testing room.

The experimenters were instructed to not remove the photos completely from their slots. This speeded up the task as it was faster to merely drop the photo back into its slot and it kept the experimenters from seeing what picture they were presenting. With the card only partly removed from its slot, all the experimenter could see from his side of the table was the identifying letter on the back of the mounted photo.

After testing, the subjects returned to the waiting room and were asked to fill-out a short questionnaire. The questionnaire asked the following questions: (1) How would you describe the attitude of the experimenter? (Cold, Warm, Distant, Calm, Hurried, etc.) (2) What do you think was the real purpose of this study? (3) What do you think was the average rating expected by your experimenter on a scale -10 (extreme failure) to +10 (extreme success)? At the bottom of the questionnaire was an admonition to not discuss the study with others who might take part in it.

A confederate of the author was stationed behind the one-way mirror at the start of the study to check on the accuracy of the student experimenters in recording their ratings. The assistant stationed behind the mirror found that it was very difficult to hear the ratings. As a result, the information that he recorded may have been less accurate than the ratings as recorded by the student experimenter. On the second evening of the study a concealed microphone was tried but still the ratings were very difficult to hear behind the one-way mirror. Rather than move the microphone nearer to the experimenter and the subject, making it obvious that they were being observed, it was thought best to discontinue this aspect of the study.

CHAPTER IV

RESULTS

The raw data are contained in Tables 12, 13, and 14 in Appendix B. A constant of 10 was added to each rating to convert all ratings to positive values for the data analysis.

The first analysis of variance computed was a 3 x 3 x 2 with the second factor, experimenters, nested under the first factor, bias condition. The third factor was ambiguity level, the set of photos used to elicit the ratings. The dependent variable in this analysis was the mean ratings given by the subjects; there were five subjects in each of 18 cells. The cell means and standard deviations are contained in Table 1. The analysis was carried out using procedures described by Winer (1962, pp. 258-263). This overall analysis of the data is contained in Table 2.

Two preliminary checks on the data were performed to assess the tenability of the assumptions underlying the analysis of variance model. To test for homogeneity of variance, a Cochran's *C* statistic was computed. The observed *C* was .132 with a *C* of .192 required to reject the hypothesis of variance equality at the .05 level where *k*, the number of cells, is equal to 20. To test whether a data transformation was advisable, a Pearson Product-Moment Correlation Coefficient, *r*, was computed between the cell means and associated cell variances. The resulting *r*, -.273, was not significant.

TABLE 1

SUMMARY OF THE MEAN RATINGS

Bias	E		Ambiguous Photos	Unambiguous Photos
Low	3	M	10.22	15.02
		SD	.65	2.01
	4	M	10.50	13.86
		SD	.38	2.18
	8	M	10.74	15.30
		SD	3.03	1.77
Medium	2	M	8.92	12.70
		SD	2.66	2.21
	6	M	9.78	15.12
		SD	2.78	1.51
	7	M	9.18	13.66
		SD	1.92	1.58
High	1	M	11.74	14.90
		SD	1.83	1.04
	5	M	12.60	13.68
		SD	2.04	2.36
	9	M	10.54	14.24
		SD	1.61	1.72

TABLE 2

OVERALL ANALYSIS OF VARIANCE BASED ON THE SUBJECTS' MEAN RATING

Source	df	MS	F	P
Bias	2	15.73	4.22	<.05
Es within Biases*	6	3.69		
Ambiguity	1	326.05	87.41	<.01
Bias x Ambiguity	2	7.72	2.07	NS
Es w. Biases x Ambiguity*	6	2.60		
Within Cell*	72	3.82		
Pooled Error	84	3.73		

Note.--After preliminary checks at the .10 level of significance, the starred sources of variance were pooled into a common error term.

As can be seen in Table 2, the bias effect was significant beyond the .05 level and the ambiguity effect was significant beyond the .01 level. The interaction of ambiguity and bias, which would show the differential effect of bias on low and high ambiguity photos, was not significant. Internal analysis of the data showed that there was a significant difference between the high bias mean (12.95) and the medium bias mean (11.56). The observed *t* was 2.80; with 84 degrees of freedom, the probability of this *t* occurring by chance is less than .005. The difference between the low bias mean (12.60) and the medium bias mean (11.56) was also significant. The probability of a *t* this large with 84 degrees of freedom is less than .025. The difference between

the high bias mean and the low bias mean was not statistically significant (t = .704).

The second general part of the analysis of the data consisted of two 3 x 5 analyses with one repeated measure, designed to get at intra-task ambiguity from the experimenters' point of view and the detailed relationship of task ambiguity and bias. The same data used in the overall analysis of variance above, the subjects' mean ratings, were used in this part of the data analysis but the data were split into two parts. The mean ratings from the subjects who were asked to rate the ambiguous photos were analyzed separately from the mean ratings of the subjects who rated the unambiguous photos.

For these analyses of variance the mean ratings were ordered in the sequence in which the subjects were tested by the experimenters. The repeated factor was the subject order and the nonrepeated factor was the bias condition. The data were analyzed following procedures described by Winer (1962, pp. 302-307). The cell means and standard deviations are contained in Tables 3 and 4. The data are presented graphically in Figures 1 and 2. Summaries of these analyses of variance are contained in Tables 5 and 6.

As can be seen in Table 5 and Figure 1, the mean ratings of the ambiguous photos did not show a significant experimenter bias effect. The order effect was significant beyond the .05 level. Figure 1 shows that this significant subject order effect reflects a general decline in the mean ratings given by the later seen subjects.

TABLE 3

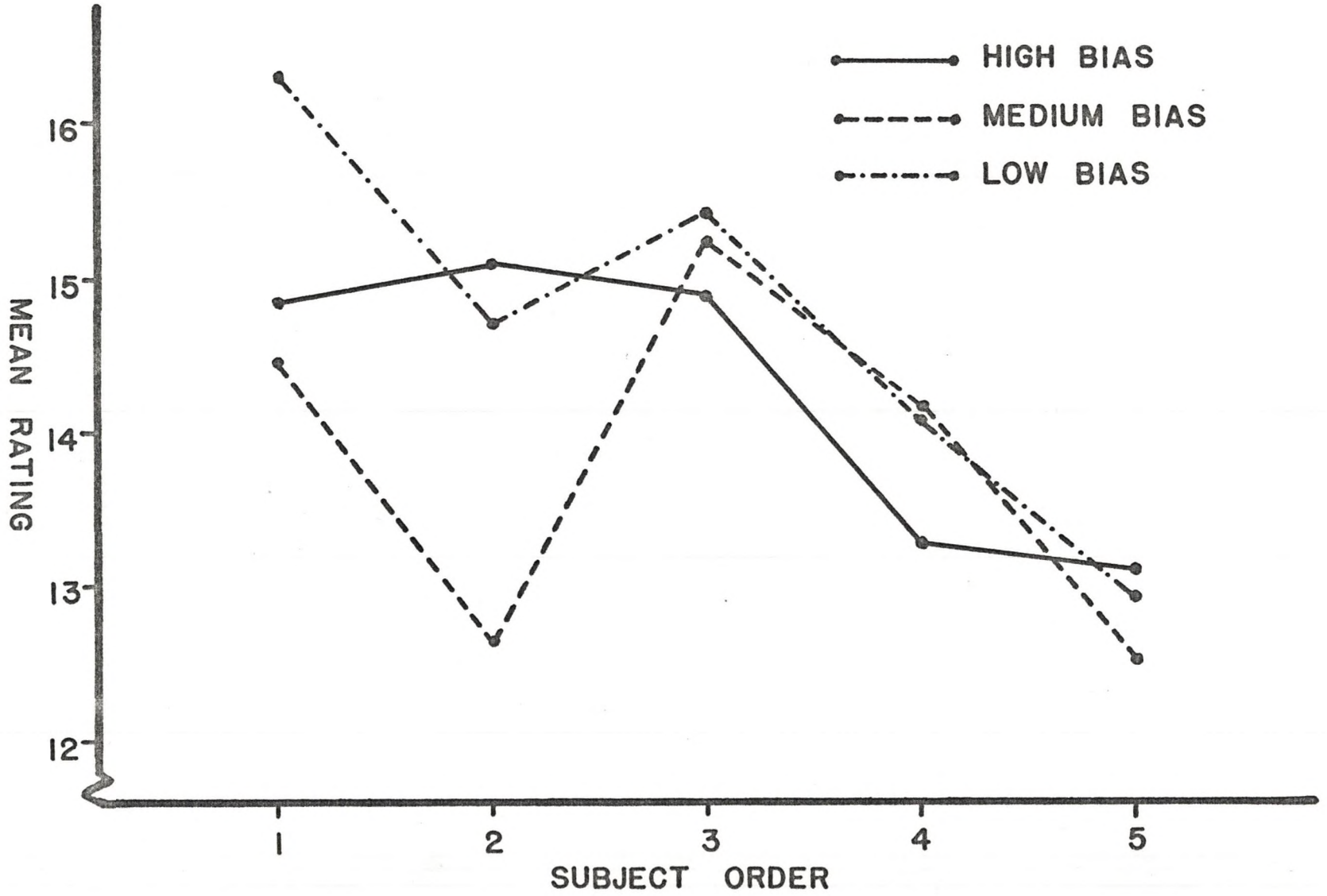
SUMMARY OF THE MEAN RATINGS OF THE UNAMBIGUOUS PHOTOS

Bias		Subject Order 1	2	3	4	5
Low	M	16.36	14.73	15.46	14.10	12.96
	SD	2.15	2.43	.83	1.47	1.78
Medium	M	14.46	12.66	15.23	14.20	12.56
	SD	1.35	1.26	2.08	.60	3.23
High	M	14.86	15.13	14.93	13.30	13.13
	SD	1.75	.30	1.76	1.60	2.54

TABLE 4

SUMMARY OF THE MEAN RATINGS OF THE AMBIGUOUS PHOTOS

Bias		Subject Order 1	2	3	4	5
Low	M	11.77	11.63	9.46	9.60	9.96
	SD	1.59	2.34	1.81	.36	.90
Medium	M	10.23	9.90	9.16	9.23	7.93
	SD	1.14	1.34	3.13	3.96	2.13
High	M	13.06	11.73	9.83	12.33	11.16
	SD	.45	1.50	.77	2.57	2.60


HIGH BIAS
MEDIUM BIAS
LOW BIAS
MEAN RATING
16
15
14
13
12
1
2
3
4
5
SUBJECT ORDER

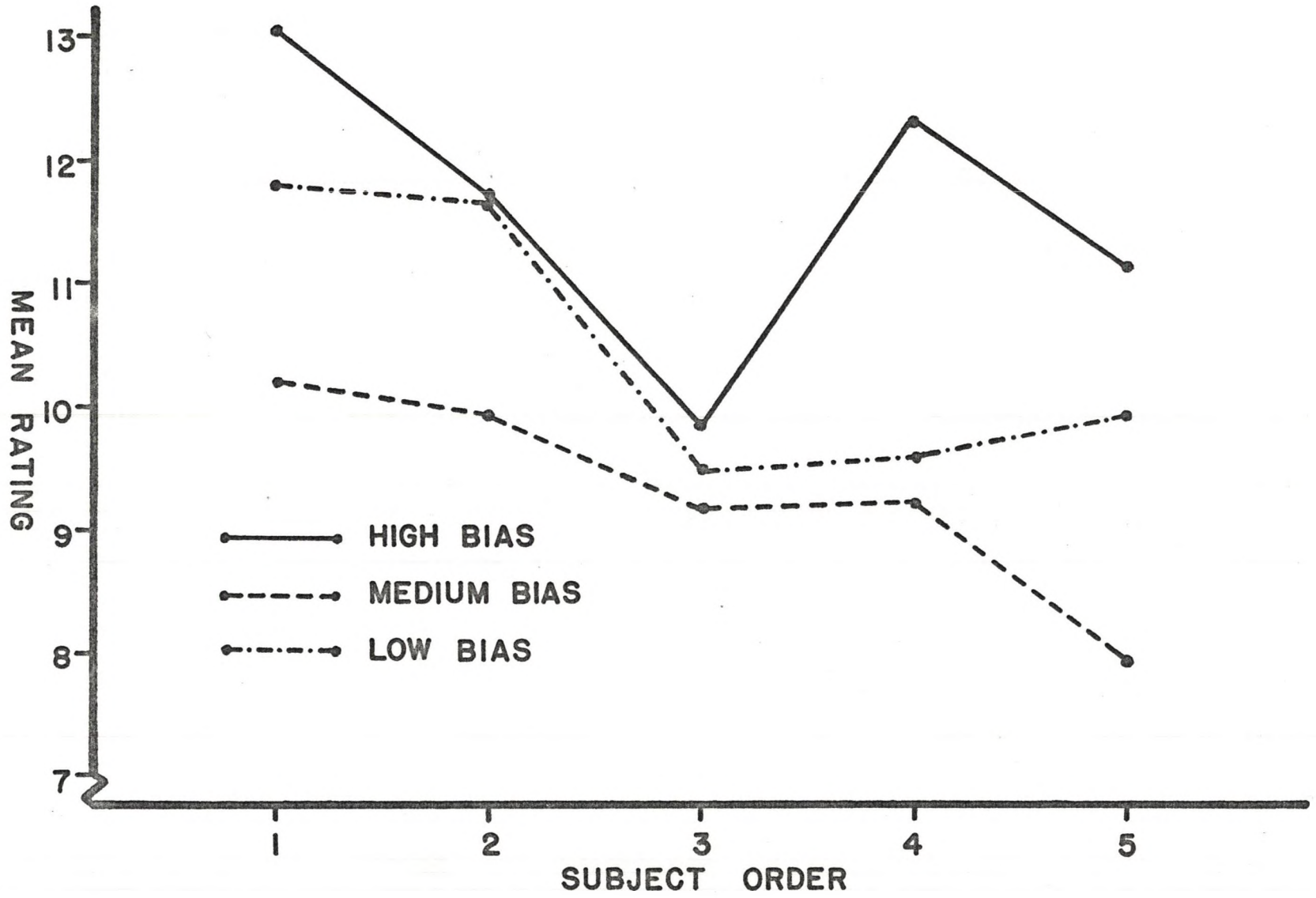

13
12
11
10
9
8
7
MEAN RATING
HIGH BIAS
MEDIUM BIAS
LOW BIAS
1
2
3
4
5
SUBJECT ORDER

TABLE 5

ANALYSIS OF VARIANCE OF THE MEAN RATINGS OF THE UNAMBIGUOUS PHOTOS

Source	df	MS	F	P
Between				
Bias	2	3.04	.767	NS
Es within Biases	6	3.96		
Within				
Subject Order	4	8.74	3.05	<.05
Bias x Order*	8	1.76	.541	NS
Bias x Es within Biases*	24	3.25		
Pooled Error	32	2.87		

Note.--After preliminary checks at the .10 significance level, the starred sources of variance were pooled into a common error term.

TABLE 6

ANALYSIS OF VARIANCE OF THE MEAN RATINGS OF THE AMBIGUOUS PHOTOS

Source	df	MS	F	P
Between				
Bias	2	20.42	9.19	<.025
Es within Biases	6	2.22		
Within				
Subject Order	4	7.75	1.68	NS
Bias x Order*	8	1.44	.310	NS
Bias x Es within Biases*	24	4.61		
Pooled Error	32	3.82		

Note:--After preliminary checks at the .10 significance level, the starred sources of variance were pooled into a common error term.

Table 6 and Figure 2, based on mean ratings given by subjects who viewed the ambiguous photos, reflect a significant experimenter bias effect; the subject order effect was not significant. The order of the bias levels is not as predicted, however. As can be seen in Figure 2, the low bias mean ratings are between the high bias mean ratings and the neutral or medium bias mean ratings rather than lower than the medium bias mean ratings.

The third general part of the data analysis was an examination of the relationship of bias to intra-task ambiguity from the point of view of the subjects. This data analysis involved two 3 x 10 analyses with the second factor a repeated measure. The second factor represented the individual ratings of the 10 photos rated by each subject in the order presented. The first factor was the three levels of bias. The data, as in the second general part of the data analysis, was split into two groups. The data from the subjects who rated the unambiguous photos were analyzed separately from the data from the subjects who rated the ambiguous photos. The cell means and standard deviations are given in Tables 7 and 8. These analyses of variance are summarized in Tables 9 and 10 and the data are presented graphically in Figures 3 and 4.

From the numerical analysis of the data from the unambiguous photos, contained in Table 9, one can see that there is a significant bias x picture order interaction but also a significant picture order main effect. The bias main effect was not significant; the F was only .870. The graphic presentation of this data, contained in Figure 3, shows that while there is much crossover of the lines representing the

TABLE 7

SUMMARY OF THE INDIVIDUAL ORDERED RATINGS OF THE UNAMBIGUOUS PHOTOS

Bias		Picture Order 1	2	3	4	5	6	7	8	9	10
Low	M	16.73	15.13	15.00	14.47	15.73	14.60	14.13	14.23	13.53	13.73
	SD	2.58	3.66	4.16	4.58	2.40	4.98	3.81	3.41	3.93	5.22
Medium	M	13.93	13.13	14.60	14.87	13.33	15.60	13.60	14.33	11.40	13.47
	SD	5.30	5.07	3.83	3.40	4.32	5.01	3.52	2.99	5.05	3.38
High	M	16.07	15.73	15.80	17.13	14.40	11.87	11.87	13.47	12.87	13.53
	SD	3.28	3.61	3.34	2.39	2.82	4.32	4.85	5.28	4.31	3.23

TABLE 8

SUMMARY OF THE INDIVIDUAL ORDERED RATINGS OF THE AMBIGUOUS PHOTOS

Bias		Picture Order 1	2	3	4	5	6	7	8	9	10
Low	M	9.13	10.73	10.87	8.93	9.67	10.67	11.00	12.07	8.33	13.27
	SD	3.23	4.56	6.22	5.61	4.65	5.65	2.83	4.88	4.48	3.13
Medium	M	9.13	10.13	8.40	7.60	9.47	8.33	10.93	7.60	10.07	11.27
	SD	3.98	5.25	5.45	4.44	6.75	6.23	6.20	5.15	5.98	6.80
High	M	11.87	12.13	11.00	10.73	10.80	9.93	13.20	12.60	12.47	11.53
	SD	4.53	4.10	3.76	3.95	5.12	3.79	3.95	4.50	4.03	5.30

TABLE 9

ANALYSIS OF VARIANCE OF THE ORDERED INDIVIDUAL RATINGS OF THE UNAMBIGUOUS PHOTOS

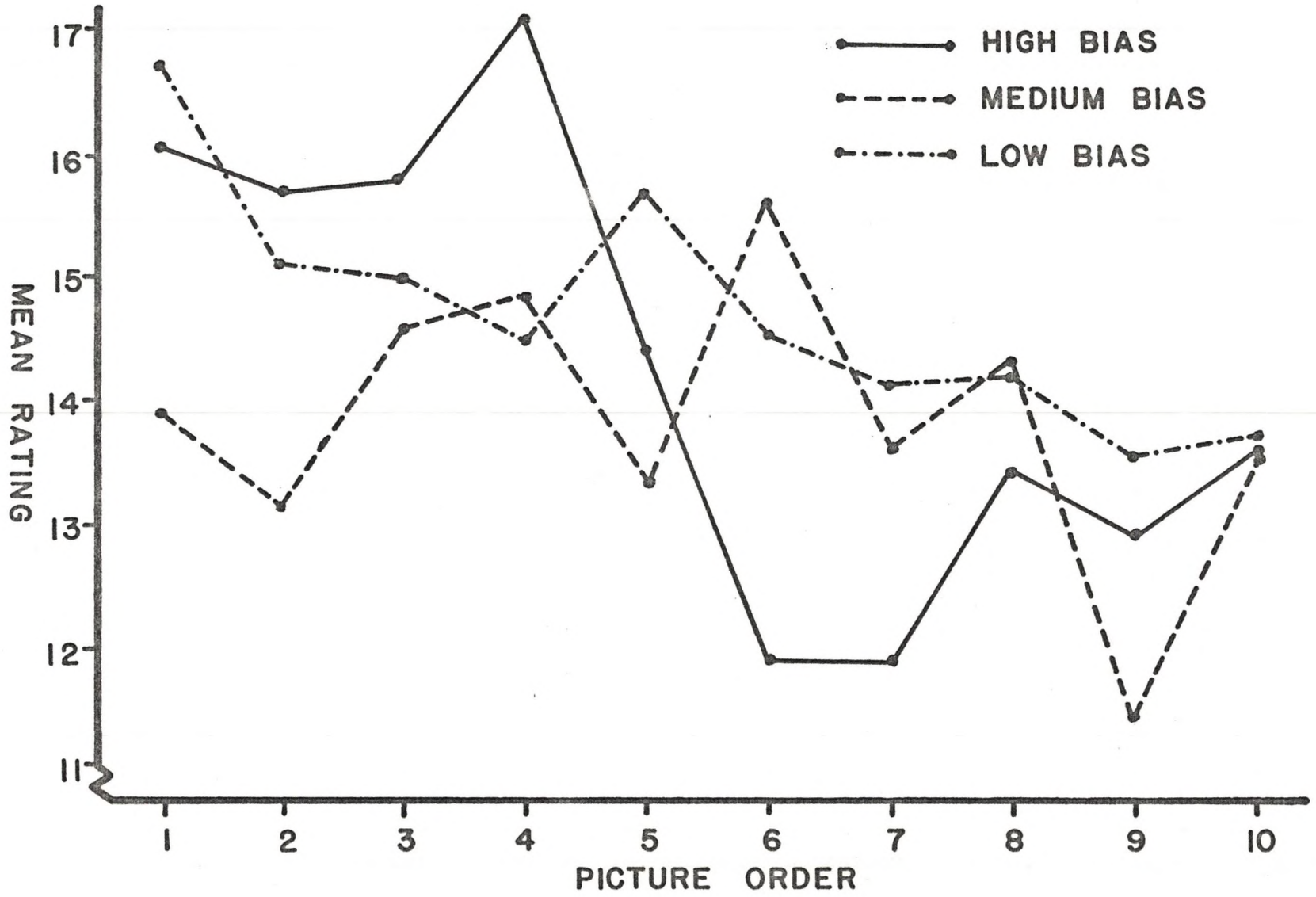
Source	df	MS	F	P
Between				
Bias	2	30.83	.870	NS
Ss within Biases	42	35.56		
Within				
Picture Order	9	43.41	6.00	<.001
Bias x Picture Order	18	20.68	2.86	<.001
Bias x Ss within Biases	378	7.24	2.86	

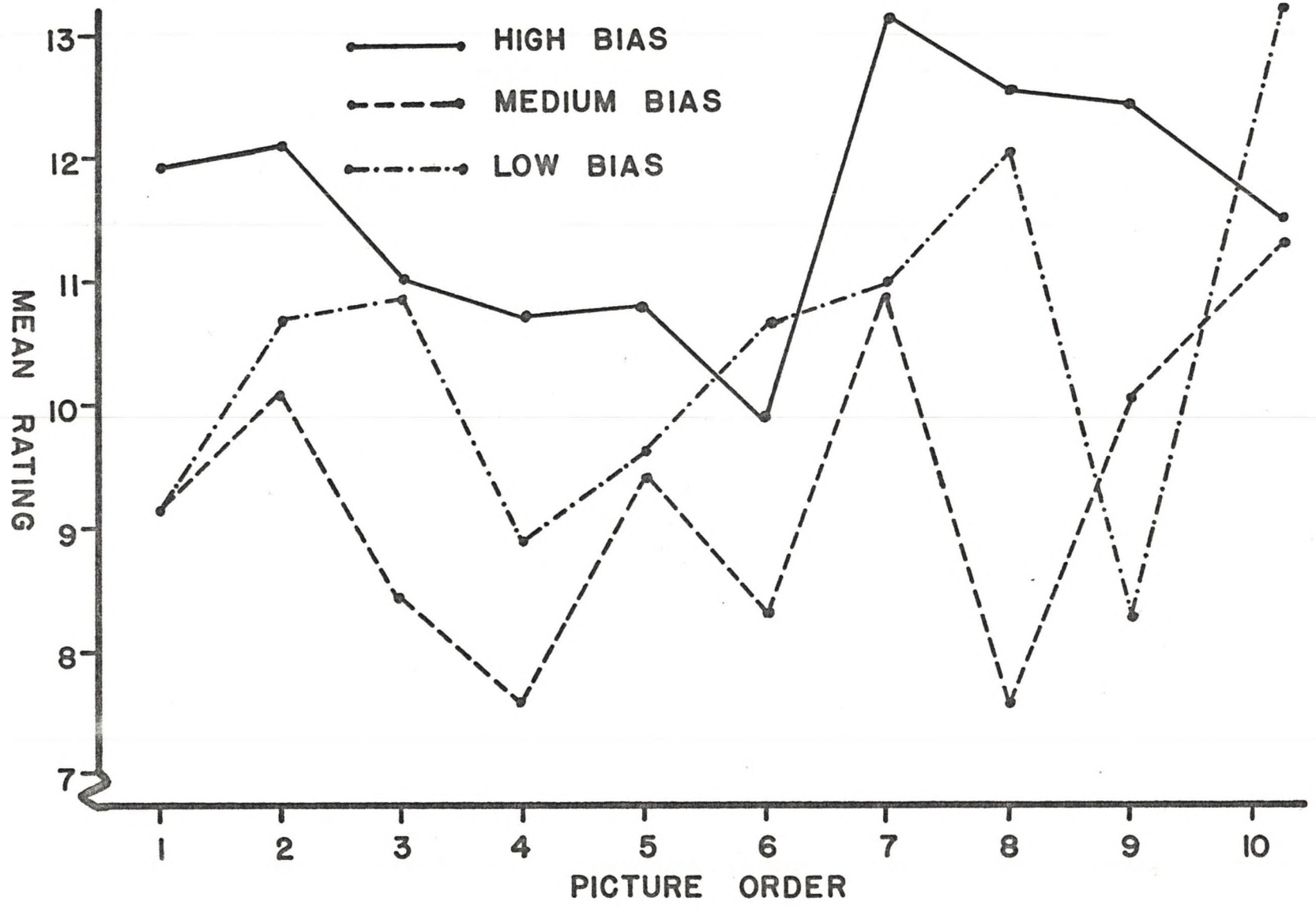
TABLE 10

ANALYSIS OF VARIANCE OF THE ORDERED INDIVIDUAL RATINGS OF THE AMBIGUOUS PHOTOS

Source	df	MS	F	P
Between				
Bias	2	204.17	5.14	<.025
Subjects within Biases	42	39.75		
Within				
Picture Order	9	37.51	1.68	<.10
Bias x Picture Order*	18	18.63	.83	NS
Bias x Ss within Biases*	378	22.51		
Pooled Error	396	22.32		

Note.--After a preliminary check at the .10 level of significance, the starred variance sources were pooled into a common error term.


HIGH BIAS
MEDIUM BIAS
LOW BIAS
MEAN RATING
17
16
15
14
13
12
11
1
2
3
4
5
6
7
8
9
10
PICTURE ORDER


HIGH BIAS
MEDIUM BIAS
LOW BIAS
MEAN RATING
13
12
11
10
9
8
7
1
2
3
4
5
6
7
8
9
10
PICTURE ORDER

three bias conditions, the general trend of the data is clearly toward lower ratings for later rated photos.

The analysis of the ordered ratings of the ambiguous photos, contained in Table 10, showed a significant experimenter bias effect ($\underline{P}$ less than .025) and a picture order effect that approaches significance ($\underline{P}$ less than .10). The bias x picture order interaction was not statistically significant although the graphic presentation of the data, Figure 4, shows some crossover of the lines representing the three bias conditions. Figure 4 also shows that the ratings elicited by the low bias experimenters tended to be intermediate between the ratings elicited by the high and medium bias experimenters. The ratings elicited by the high and medium bias experimenters tend to be in the predicted order.

CHAPTER V

DISCUSSION OF THE RESULTS

Inter-task Ambiguity

The first experimental hypothesis, that the experimenter bias effect is a function of inter-task ambiguity, is supported by the data. In both cases where the data from the ambiguous photos were analyzed separately from the unambiguous photo ratings, a significant experimenter bias effect was found with the ambiguous photos and not found with the unambiguous photos. However, this effect was small since the interaction reported in Table 2 was not significant. The magnitude of the F statistic for the ambiguity effect was extremely high, 87.41, indicating the great difference between the low ambiguity and the high ambiguity data.

While the data indicated a statistically significant experimenter bias effect, a closer examination of the data indicated that the ratings were not in the magnitude order predicted from the experimenters' expectancies. The ratings given by the subjects tested by the low expectancy experimenters were generally higher than the ratings given by the subjects tested by the medium expectancy experimenters. This is clearly shown in Figures 2 and 4. The internal analysis of the overall analysis of variance reported in Table 2 also showed this unpredicted order. The high expectancy and low expectancy means, while not significantly different from each other, were both significantly

larger than the intermediate expectancy mean. A possible explanation of this unexpected order is that the task may not have been believable to the low expectancy experimenters. The mean overall rating of the 20 photos was about +3 during the no-expectancy standardization sessions. It may be more believable to an experimenter to be told to expect a mean rating of +5 or 0 on these photos than to expect a mean rating of -5. An expected mean rating of -5 is just not plausible when all of the photos are rated without preconceptions. A second group of low ambiguity photos, selected from the extreme low success photos, might have had a balancing effect and made the task more plausible to the low expectancy experimenters.

There is some evidence against the above explanation of the unpredicted rating order. Figure 2, the graphic presentation of the mean ratings of the high ambiguity photos, shows that the unpredicted order of the mean ratings was evident with very early subjects as well as with later seen subjects. Generally, according to the observers stationed behind the one-way mirror, the experimenters did not look at the photos before testing the early subjects. Some experimenters did look at all of the photos, but only after testing some of their subjects. The anomalous order of the mean ratings, then, was probably a function of something brought to the testing situation by the experimenters or subjects and not acquired during the testing.

A second possible, but improbable, explanation for the rating order is a breakdown in the randomization procedures. It is possible that the experimenters randomly selected to be low expectancy experimenters were different in some way from the other six experimenters. For some unknown reason they did not communicate their expectancies. A

strictly random process (drawing names) was followed in assigning the experimenters to the bias conditions. It is very unlikely that there was any systematic difference between the experimenters introduced as a result of the random assignment of the experimenters.

Importance of the Bias Effect

While the experimenter bias effect was found to be statistically significant in this study, the amount of variability attributable to the bias effect was quite small. The variability attributable to the bias effect ranged from 1.2% in the data analysis summarized in Table 9 (the bias effect was not statistically significant in that analysis) to a high of 24.5% in Table 6, the analysis of the mean ratings of the ambiguous photos. According to Rosenthal (1969b), with a sample size of about 50 and a median P value of .10, expectancy effects, on the average, account for only about 5% of the total variance. The median variability accounted for by the experimenter bias effect in this study was 4.0%; this value came from the data summarized in Table 5. The amounts of variability attributable to bias were 3.8% for the data in Table 10 and 4.8% for Table 2. Rosenthal stated that these small effects are still worth worrying about, however, "because that (5%) may be about the average magnitude of variance accounted for in much or most behavioral prediction" (1969b, p. 14).

One can question the importance of an effect that typically accounts for only about 5% of the total variance. There are many experimenters, however, where 5% of the variance can make the difference between one conclusion and another. Psychologists currently seem to have a penchant for arbitrary points for statistical significance,

e.g., .01. When one holds strongly to some arbitrary level of significance for decision making, 5% of the variance can be very important.

Intra-task Ambiguity

The second experimental hypothesis was not supported by the data. Where a significant experimenter bias effect was found, it did not seem to be differentially related to the experimenters' or the subjects' familiarity with the task.

The prediction that the experimenters' increasing familiarity with the task would lead to more efficient bias communication with later seen subjects was not confirmed by the data. Figure 2, the ordered mean ratings of the ambiguous photos, indicates that early subject performance is about the same as late subject performance. The bias effect was not minimal with early subjects and maximal with late subjects but maximal with early and late subjects and minimal with the intermediate subjects. The experimenters' experience was not a relevant variable in this study.

The subject's performance over the course of the task (Figure 4) appeared to be relatively independent of their familiarity with the task. The prediction that the subjects' ratings would become more independent of the experimenters' biases with increasing familiarity with the task was not supported by the data. The findings of Weick (Rosenthal, 1966) and Rosenthal et al., (1964) were not corroborated by this study. The data pertaining to the subjects' performance over the course of the 10 photos each rated showed great variability; this made it difficult to state any clear conclusions.

The ratings of the unambiguous photos, while not showing a significant experimenter bias effect, showed a relationship between the subjects' and experimenters' familiarity with the task and the ratings given. Both analyses of variance that were based on the ratings of the unambiguous photos, summarized in Tables 5 and 9, showed a significant order effect. The analysis summarized in Table 9 also showed a significant bias x picture order interaction, however. The graphic presentation of the data, Figures 1 and 3, show clearly that the general trend of the data is toward lower ratings with increasing familiarity both from the experimenters' point of view, Figure 1, and the subjects' point of view, Figure 3. The early ratings depicted in Figures 1 and 3 were not much different from the mean ratings under the no-expectancy standardization condition. In Figure 1 the mean rating of the first subjects was 15.27; the mean rating of the last seen subjects was 12.88. In Figure 3 the mean rating of the first photo was 15.58, of the last photo, 13.58.

The trends were probably artifactual and not dependent upon familiarity or decreasing ambiguity. One possible reason why the later tested subjects rated lower on the unambiguous photos was that they had to wait longer before their testing sessions. When drawing for subject order in the presence of the subjects, the author observed that subjects who did not have to wait, the early drawn ones, were quite happy. They were getting experimental credit for a relatively short time investment. The task may have been more negative for subjects who had to wait for up to 45 minutes to be tested. This possible growing impatience with the whole situation could conceivably have lead to later lower ratings.

The ratings of the ambiguous photos in relation to subject order, Table 6, Figure 2, also showed some tendency for longer-wait subjects to rate lower. The first-subject mean was 11.68 while the fifth-subject mean was 9.68. The tendency toward lowered later subject ratings was not as strong as with the unambiguous photos. The order effect, Table 6, was not statistically significant (P less than .25) as it had been with the unambiguous photos (P less than .05), Table 5.

The trends of the ratings ordered by picture order, Figures 3 and 4, were quite different from the trends of the ratings ordered by subject order, Figures 1 and 2. Where the subject-ordered ratings show a roughly parallel decline, the picture-ordered ratings of the unambiguous and ambiguous photos show a tendency toward convergence. The picture-ordered ratings of the unambiguous photos had a mean of 15.58 for the first photo and 13.58 for the tenth photo so there was a tendency for later pictures to be rated lower than earlier pictures as stated earlier. The order effect was significant (F=6.00, P less than .001) but also the bias x picture order interaction (F=2.86, P less than .001) as Table 9 indicates. The picture ordered ratings of the ambiguous photos (Figure 4) had a mean of 10.04 for the first photo and 12.02 for the tenth photo; the general trend of the ratings seems to be toward higher ratings for later rated photos. The picture-order effect for the ambiguous photos approached but did not achieve statistical significance (F=1.70, P less than .10) as indicated by Table 10.

What appears to have happened is that the subjects over the course of the 10 photos each rated retreated to ratings of higher social desirability. For the unambiguous photos the subjects started rating

high, as expected, but retreated to lower ratings later. The ambiguous photos were initially rated as being not successful and not unsuccessful as in the no-expectancy situation; they tended to be rated as mildly successful later. While it is probably very acceptable to say someone is mildly successful, it is somewhat less acceptable to say someone is very successful or not successful and not unsuccessful. Objectively, the photos probably would have to be rated more as the subjects did on the first photos they rated.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Previous research and discussion has suggested that task ambiguity and experimenter-subject familiarity with the task are important factors affecting the generality and strength of the experimenter bias effect. These factors were conceptualized as inter-task ambiguity (inherent in the task) and two types of intra-task ambiguity (experimenter's familiarity with the task and subject's familiarity with the task). It was hypothesized that experimenter bias would: (1) be a significant factor in a more ambiguous task and not a significant factor in a less ambiguous task, (2) more likely be communicated to the subjects who were less familiar with the task, and (3) more likely be communicated to the subjects as the experimenters became more familiar with the task.

Nine upperclassmen experimenters were randomly assigned to one of three bias level, high, medium, and low (expect +5, 0, or -5 mean ratings on the person perception task). Each experimenter tested 10 subjects (female volunteers from introductory psychology) in a random order. Five subjects rated high ambiguity photos (mean ratings near 0 on the success-failure dimension under a no-expectancy condition) and five viewed low ambiguity photos (mean ratings high in the success direction under the no-expectancy condition).

Five analyses of variance were computed; a significant experimenter bias effect was found with the high ambiguity stimulus photos but not with the low ambiguity photos. However, the bias levels were not communicated in the predicted order. The high and low expectancy experimenters were both significantly higher than the medium bias experimenters. The hypotheses pertaining to intra-task ambiguity were not supported by the data.

From this study one can conclude that the experimenter bias effect is a function of ambiguity inherent in the task. The role of intra-task ambiguity is less clear and no conclusion can be stated based on the present data.

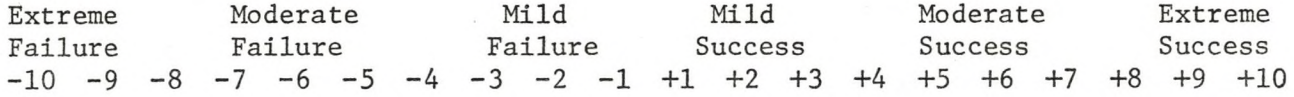

Extreme Failure
Moderate Failure
Mild Failure
Mild Success
Moderate Success
Extreme Success
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10

TABLE 11

MEAN RATINGS AND STANDARD DEVIATIONS OF THE PHOTOS

UNDER THE NO-EXPECTANCY CONDITION

High Ambiguity Photos		Low Ambiguity Photos	
M	SD	M	SD
+.4	4.6	+6.9	1.8
-.7	3.8	+6.1	4.0
-.6	3.2	+6.3	5.1
-.6	3.9	+7.2	2.5
-.9	4.8	+6.1	3.6
+.4	4.2	+7.5	2.0
+.8	4.5	+6.3	4.8
+.9	5.0	+6.1	3.9
-.8	4.3	+7.6	2.2
0	4.6	+8.2	1.7

TABLE 12

RAW SCORES FOR THE LOW EXPECTANCY EXPERIMENTERS

		Ambiguous Photos Photo Order											Unambiguous Photos Photo Order									
E	S	1	2	3	4	5	6	7	8	9	10	S	1	2	3	4	5	6	7	8	9	10
	1	6	9	14	11	11	19	11	16	5	8	2	17	15	19	13	20	20	17	12	20	20
	3	7	9	11	14	4	1	12	8	18	15	4	14	12	14	9	17	6	9	18	9	13
3	5	6	17	5	18	11	18	7	8	6	12	6	20	16	17	12	18	18	17	14	12	15
	7	14	5	3	0	15	12	9	16	3	18	8	18	16	19	20	18	15	12	9	16	15
	10	8	11	15	13	7	3	8	14	8	12	9	17	15	19	7	14	19	5	20	13	11
	5	11	12	13	6	5	8	12	16	12	12	1	12	19	13	14	18	16	11	11	14	11
	6	17	2	19	7	0	13	15	9	9	16	2	19	20	15	19	14	20	18	20	7	17
4	8	5	13	16	2	8	5	12	17	8	13	3	17	6	16	20	15	13	14	15	11	18
	9	11	14	9	0	13	7	9	9	15	13	4	13	18	7	8	14	13	14	16	14	14
	10	8	11	18	13	12	12	14	8	5	8	7	17	11	5	18	14	7	13	16	8	0
	3	9	19	14	6	16	15	16	20	3	18	1	20	18	18	19	16	20	18	15	16	19
	5	11	13	18	14	9	18	14	15	14	17	2	17	19	14	20	16	19	19	11	11	7
8	8	8	4	1	7	8	11	9	9	6	11	4	13	14	18	12	18	15	17	15	19	19
	9	7	11	3	15	9	5	9	13	7	14	6	18	15	15	12	12	7	14	11	15	15
	10	9	11	4	8	17	13	8	3	6	12	7	19	13	16	14	12	11	14	11	18	12

Note.--A constant of 10 was added to each rating.

TABLE 13

RAW SCORES FOR THE MEDIUM EXPECTANCY EXPERIMENTERS

		Ambiguous Photos Photo Order											Unambiguous Photos Photo Order									
E	S	1	2	3	4	5	6	7	8	9	10	S	1	2	3	4	5	6	7	8	9	10
	1	13	12	7	6	5	14	4	18	17	18	3	3	13	17	15	15	15	13	7	20	11
	2	7	5	13	17	13	19	11	8	12	5	4	11	8	12	15	16	20	12	12	11	12
2	6	13	18	4	6	20	3	8	13	15	0	5	19	18	16	12	7	12	16	17	16	15
	8	9	15	0	4	1	2	1	3	0	20	7	8	20	11	15	11	20	11	12	18	13
	10	9	0	5	8	5	12	9	8	2	9	9	12	16	7	9	3	0	5	17	2	19
	1	5	8	14	7	4	3	14	1	19	16	4	16	14	18	18	17	16	12	17	11	12
	2	8	9	4	11	17	1	18	13	5	17	5	20	15	15	9	16	16	11	14	11	11
6	3	14	8	6	3	5	2	2	2	11	4	6	19	20	18	14	17	20	19	16	16	16
	7	14	7	17	14	12	16	13	12	13	16	8	11	3	14	17	12	18	13	20	12	18
	9	4	14	19	1	8	3	19	2	16	18	10	18	17	19	15	12	18	18	14	11	11
	2	15	11	3	8	17	6	19	5	2	16	1	7	8	15	20	20	20	17	15	12	19
	4	6	2	8	7	13	15	5	8	11	9	3	11	8	8	13	11	14	16	13	7	12
7	5	6	17	12	12	19	5	16	2	14	15	7	17	13	16	13	15	14	16	14	2	14
	6	11	14	6	8	2	15	8	12	9	3	9	19	8	20	20	16	16	13	14	11	12
	8	3	12	8	2	1	9	17	7	5	3	10	18	15	13	18	12	15	12	13	11	7

Note.--A constant of 10 was added to each rating.

TABLE 14

RAW SCORES FOR THE HIGH EXPECTANCY EXPERIMENTERS

		Ambiguous Photos Photo Order											Unambiguous Photos Photo Order									
E	S	1	2	3	4	5	6	7	8	9	10	S	1	2	3	4	5	6	7	8	9	10
	1	12	13	9	8	12	16	14	14	12	18	3	13	13	14	18	15	8	6	20	12	13
	2	11	9	13	6	8	5	12	14	12	13	4	19	18	15	20	18	7	18	5	13	15
1	6	8	11	8	5	16	7	11	12	13	5	5	15	20	20	14	11	19	18	15	12	14
	7	13	11	17	12	8	13	11	13	14	7	8	14	18	15	18	18	13	11	15	15	13
	10	18	19	15	7	19	6	6	16	17	18	9	17	18	15	20	15	12	14	17	14	15
	1	13	15	12	18	6	12	18	15	9	18	2	20	17	9	15	17	12	9	17	17	14
	4	14	12	12	15	13	15	11	17	8	16	3	15	18	15	20	13	14	15	15	12	15
5	6	7	2	9	13	11	13	12	20	19	1	5	14	17	19	14	14	15	20	19	12	17
	9	20	15	11	11	20	15	17	15	12	15	7	20	12	11	17	12	9	11	11	7	8
	10	11	14	8	9	12	8	9	14	7	11	8	8	20	20	15	13	7	4	3	7	9
	1	18	12	16	9	13	7	17	14	16	6	2	18	18	18	18	13	20	12	16	20	14
	3	5	18	11	7	9	11	20	3	18	14	4	17	8	18	19	20	15	15	8	14	20
9	7	7	11	2	13	5	6	18	8	15	7	5	14	11	20	17	15	11	6	19	5	11
	8	14	12	9	11	9	7	11	8	8	11	6	20	16	15	13	11	5	7	8	20	16
	9	7	8	13	17	1	8	11	6	7	13	10	17	12	13	19	11	11	12	14	13	9

Note.--A constant of 10 was added to each rating.

REFERENCES

Adair, J. G., & Epstein, J. S. Verbal cues in the mediation of experimenter bias. Psychological Reports, 1968, 22, 1045-1053.

Barber, T. X., & Silver, M. J. Fact, fiction, and the experimenter bias effect. Psychological Bulletin, 1968a, 70 (6, Part 2), 1-29.

Barber, T. X., & Silver, M. J. Pitfalls in data analysis and interpretation: A reply to Rosenthal. Psychological Bulletin, 1968b, 70, (6, Part 2), 48-62.

Barber, T. X. Invalid arguments, postmortem analyses, and the experimenter bias effect. Journal of Consulting and Clinical Psychology, 1969, 33, 11-14.

Barber, T. X., Calverley, D. S., Forgione, A., McPeake, J. D., Chaves, J. F., & Bowen, B. Five attempts to replicate the experimenter bias effect. Journal of Consulting and Clinical Psychology, 1969, 33, 1-6.

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

Cordaro, L., & Ison, J. H. Psychology of the scientist: X. Observer bias in classical conditioning of the planarian. Psychological Reports, 1963, 13, 787-789.

Ekren, U. The effect of experimenter knowledge of a subject's scholastic standing on the performance of a reasoning task. Unpublished master's thesis, Marquette University, 1962.

Epstein, J. The role of verbal cues in the mediation of experimenter bias. Unpublished master's thesis, University of Manitoba, Winnipeg, Manitoba, Canada, 1966.

Ferber, R., & Wales, H. G. Detection and correction of interviewer bias. Public Opinion Quarterly, 1952, 16, 107-127.

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

Fode, K. L. The effects of experimenters' anxiety, and subjects' anxiety, social desirability and sex, on experimenter outcome-bias. Unpublished doctoral dissertation, University of North Dakota, 1967.

Friedman, N., Kurland, D., & Rosenthal, R. Experimenter behavior as an unintended determinant of experimental results. Journal of Projective Techniques and Personality Assessment, 1965, 29, 479-490.

Friedman, N. The social nature of psychological research. New York: Basic Books, 1967.

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

Glixman, A. E. Psychology of the scientist: XXII. Effects of examiner, examiner-sex, and subject-sex upon categorizing behavior. Perceptual and Motor Skills, 1967, 24, 107-117.

Goldstein, A. P. Therapist-patient expectancies in psychotherapy. New York: Macmillan, 1962.

Harrington, G. M. Experimenter bias: Occam's razor versus Pascal's wager. Psychological Reports, 1967, 21, 527-528.

Harrington, G. M., & Ingraham, L. H. Experimenter bias and tails of Pascal. Psychological Reports, 1967, 21, 513-516.

Ingraham, L. H., & Harrington, G. M. Psychology of the scientist: XVI. Experience of E as a variable in reducing experimenter bias. Psychological Reports, 1966, 19, 455-461.

Jacob, T. The experimenter bias effect: A failure to replicate. Psychonomic Science, 1968, 13, 239-240.

Kennedy, J. J. Experimenter outcome bias in verbal conditioning: A failure to detect the Rosenthal effect. Psychological Reports, 1969, 25, 495-500.

Kintz, B. L., Delprato, D. J., Mettee, D. R., Persons, C. E., & Schappe, R. H. The experimenter effect. Psychological Bulletin, 1965, 63, 223-232.

Laszlo, J. P., & Rosenthal, R. Subject dogmatism, experimenter status and experimenter expectancy effect. Cambridge, Mass.: Harvard University, Department of Social Relations, 1967. (Mimeographed.)

Lindzey, G. A note on interviewer bias. Journal of Applied Psychology, 1951, 35, 182-184.

McGuigan, F. J. The experimenter: A neglected stimulus object. Psychological Bulletin, 1963, 60, 421-428.

Merton, R. K. Self-fulfilling prophecy. Antioch Review, 1948, 8, 193-210.

Marwit, S. J., & Marcia, J. R. Tester-bias and response to projective instruments. Journal of Consulting Psychology, 1967, 31, 253-258.

Masling, J. Differential indoctrination of examiners and Rorschach responses. Journal of Consulting Psychology, 1965, 29, 198-201.

Masling, J. Role-related behavior of the subject and psychologist and its effects upon psychological data. Nebraska Symposium on Motivation, 1966, 14, 67-103.

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

Orne, M. T., & Scheibe, K. E. The contribution of nondeprivation factors in the production of sensory deprivation effects: The psychology of the "panic button." Journal of Abnormal and Social Psychology, 1964, 68, 3-12.

Pflugrath, J. Examiner influence in a group testing situation with particular reference to examiner bias. Unpublished master's thesis, University of North Dakota, Grand Forks, 1962.

Raffetto, A. M. Experimenter effects on subjects' reported hallucinatory experiences under visual and auditory deprivation. Unpublished master's thesis, San Francisco State College, 1967.

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

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

Rosenthal, R. On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of the experimenter results. American Scientist, 1963, 51, 268-283.

Rosenthal, R., & Fode, K. L. Psychology of the scientist: V. Three experiments in experimenter bias. Psychological Reports, 1963, 12, 491-511.

Rosenthal, R., Fode, K. L., Vikan-Kline, L. L., & Persinger, G. W. Verbal conditioning: Mediator of experimenter expectancy effects? Psychological Reports, 1964, 14, 71-74.

Rosenthal, R. Experimental outcome-orientation and the results of the psychological experiment. Psychological Bulletin, 1964a, 61, 405-412.

Rosenthal, R. The effect of the experimenter on the results of psychological research. In B. A. Maher (Ed.), Progress in experimenter personality research. Vol. 1. New York: Academic Press, 1964b.

Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.

Rosenthal, R. Covert communication in the psychological experiment. Psychological Bulletin, 1967a, 67, 356-367.

Rosenthal, R. Psychology of the scientist: XXIII. Experimenter expectancy, experimenter experience, and Pascal's Wager. Psychological Reports, 1967b, 20, 619-622.

Rosenthal, R. Psychology of the scientist: XXVI. Experimenter expectancy, one tale of Pascal, and the distribution of three tales. Psychological Reports, 1967c, 21, 517-520.

Rosenthal, R. Experimenter expectancy and the reassuring nature of the null hypothesis decision procedure. Psychological Bulletin, 1968, 70 (6, Part 2), 30-47.

Rosenthal, R., & Jacobson, L. Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. New York: Holt, Rinehart and Winston, 1968.

Rosenthal, R. On not so replicated experiments and not so null results. Journal of Consulting and Clinical Psychology, 1969a, 33, 7-10.

Rosenthal, R. Unintended effects of the clinician in clinical interaction: A taxonomy and a review of clinician expectancy effects. Australian Journal of Psychology, 1969b, 21, 1-20.

Shames, M. L., & Adair, J. G. Experimenter-bias as a function of the bype and structure of the task. Paper presented at the meeting of the Canadian Psychological Association, Ottawa, May, 1967.

Shinkman, P. G., & Kornblith, C. L. Comment on observer bias in classical condition of the planarian. Psychological Reports, 1965, 16, 56.

Silver, M. J. Experimenter modeling: A critique. Journal of Experimental Research in Personality, 1968, 3, 172-178.

Silverman, I. The effects of experimenter outcome expectancy on latency of word association. Journal of Clinical Psychology, 1968, 24, 60-63.

Stanton, F. Further contributions at the twentieth anniversary of the Psychological Corporation and to honor its founder, James McKeen Cattel. Journal of Applied Psychology, 1942, 26, 16-17.

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

Strauss, M. E. Examiner expectancy: Effects on Rorschach experience balance. Journal of Consulting and Clinical Psychology, 1968, 32, 125-129.

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

Vikan-Kline, L. L. The effect of an experimenter's perceived status on the mediation of experimenter bias. Unpublished master's thesis, University of North Dakota, Grand Forks, 1962.

Weiss, L. Experimenter bias as a function of stimulus ambiguity. Buffalo, New York: State University of New York, 1967. (Mimeographed.)

Weiss, L. Effects of subject, experimenter, and task variables on compliance with the experimenters expectation. Journal of Projective Techniques, 1969, 33, 247-256.

Wessler, R. L. Experimenter expectancy effects in psychomotor performance. Perceptual and Motor Skills, 1968, 26, 911-917.

Wessler, R. L., & Strauss, M. E. Experimenter expectance: A failure to replicate. Psychological Reports, 1968, 22, 687-688.

Wessler, R. L. Experimenter expectancy effects in three dissimilar tasks. Journal of Psychology, 1969, 71, 63-67.

Winer, B. J. Principles in experimental design. New York: McGraw-Hill, 1962.